Technische Universität München
TUM School School of Computation, Information and Technology

On Finite Gabor Frames and Data Transmission over Linear Time-Varying Channels

Alihan Kaplan

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitz: Prof. Dr.sc. Reinhard Heckel

Prüfende der Dissertation:

1. Priv.-Doz. Dr.-Ing. habil. Volker Pohl
2. Prof. Dr. Götz Pfander
3. Prof. Dr.-Ing. Dr.rer.nat Holger Boche

Die Dissertation wurde am 25.09.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 03.07.2024 angenommen.

Abstract

In communications, wireless channels are often modelled as time-invariant linear (LTI) systems. In such cases data transmission follows a two step scheme. First, the transmitter sends a known pilot signal through the channel and the receiver estimates the channel state information (CSI) from the received signal. In a second step, the actual message is transmitted and the receiver recovers the message using the previously obtained CSI.
However, real wireless channels have time-varying characteristics. Thus, the question arises whether it is possible to transmit data over a linear time-varying (LTV) channel without a prior estimate of CSI.
This thesis is devoted on developing methods to transmit data over LTV channels without any prior information on the CSI. In order to develop such a method a thorough understanding of LTV systems are essential.
In the formal analysis of LTV systems time-frequency representations, and in particular Gabor frames, play a key role. Thus, a significant part of this thesis is dedicated on the study of finite Gabor frames and their properties. Later on, these insights are then applied in order to develop data transmission schemes over LTV channels which do not rely on previously estimated CSI.

Kurzzusammenfassung

In der Kommunikationstheorie werden drahtlose Kanäle oft als zeitinvariante lineare (linear time-invariant, LTI) Systeme modelliert. In solchen Fällen folgt die Datenübertragung einem zweistufigen Schema. Zuerst sendet der Sender ein bekanntes Pilotsignal durch den Kanal und der Empfänger schätzt den Kanalzustand (channel state information, CSI) aus dem empfangenen Signal. In einem zweiten Schritt, wird die eigentliche Nachricht übertragen und der Empfänger rekonstruiert die Nachricht mithilfe von CSI, der im vorherigen Schritt geschätzt wurde.
Jedoch weisen reale drahtlose Kommunikationskanäle zeitvariante Eigenschaften auf. Daher stellt sich die Frage, ob es möglich ist, Daten über einen linearen zeitlich variierenden (linear time-varying, LTV) Kanal, ohne vorherige Schätzung des CSI zu, übertragen.
Diese Arbeit widmet sich der Entwicklung von Methoden zur Übertragung von Daten über LTV-Kanäle, die eine vorherige Schätzung des CSI nicht benötigen. Zum Entwurf einer solchen Methode ist ein tieferes Verständnis von LTV-Systemen unerlässlich.
Bei der formalen Untersuchung von LTV Systemen spielen Zeit-Frequenz Darstellungen, vorallem Gabor-Rahmen, eine Schlüsselrolle. Daher beschäftigt sich ein bedeutender Teil dieser Arbeit mit der Analyse von Eigenschaften der endlichen Gabor-Rahmen. In späteren Kapiteln werden die gewonnenen Erkenntnisse verwendet um Methoden zu entwicklen, die Datenübertragung über zeitvariante Kanäle, ohne eine vorherige Schätzung von CSI, ermöglichen.

Acknowledgments

First and foremost I want to thank my supervisors Dr. Volker Pohl and Prof. Dr. Holger Boche for their continuous support. I am especially grateful to Dr. Volker Pohl for his long time tutorage which by now encompasses almost seven years, the freedom to find my research interests and his advice in defining my own research agenda.
I am also grateful to Dr. Dae Gwan Lee for the insightful discussions, his useful comments and remarks which greatly shaped my research and contributed greatly to the learning process I went through.
I would like to thank my colleagues at the LTI for a very pleasant social atmosphere, who made my time there to a great experience. Each of them surely deserves a special mention but I will limit myself to the essential ones. I would like to thank Dr. Ezra Tampubolon for numerous discussions, Dr. Moritz Wiese and Jonathan Huffmann for amazing lunch breaks, to Dr. Anna Frank and Yannick Böck for making sure we had enough coffee supply at all times.
I am also grateful to my friends Anna Massalitin and Christoph Frisch for supporting me and being there for me all the years and for numerous adventures.
Last but not least, I am grateful to my mother Zarife Demiröz for her continuous support and for encouraging me to pursue my goals.

Contents

1. Introduction

In communications, wireless channels are often modelled as *linear time-invariant* (LTI) systems. Then, data transmission over communication channels follows a two step scheme. First, the transmitter feeds a known pilot signal into the channel and the receiver estimates the *channel state information* (CSI) from the received signal. In a second step, the actual message is transmitted and the receiver recovers the message using the previously estimated CSI [1]. Underlying this two step scheme is the block fading assumption [2] which means that the wireless communication channel does not change over a certain time interval. The block fading assumption enables to model the communication channel as a LTI system.
However, in certain applications, for instance, in underwater acoustics or high-frequency mobile radio channels, the time varying characteristics of the communication channel have significant influence on the output of the channel. Hence, the modelling of these communication channels as *linear time-varying* (LTV) systems is unavoidable [3–6]. Additionally, the rising application of *multiple-input multiple-output* (MIMO) systems severely magnified the effect of channel aging which describes the mismatch between the estimated CSI and the actual channel state while transmitting data. Up to some extend channel aging effects can be overcome by channel prediction [7]. Nonetheless, the time-varying nature of wireless communication channels increases in importance with advancing mobile communication technology.
In case of wireless channels the receiver observes the signal over multiple propagation paths. Hereby, each path acts as one (or multiple) delay-Doppler shifts on the original transmit signal creating a LTV channel. Hence, the observed signal at the receiver is described as a superposition of multiple delay-Doppler shifted versions of the input signal [3, 8–10]. For LTV channels the time-varying parameters of the channel are assumed to be constant over a certain time interval. Hence, the estimation of LTV channels has drawn considerable attention in the literature [3, 8, 11–13].
If we want to transmit a message over an LTV communication channel then the previously described two step scheme fails since the receiver does not has access to a reliable CSI. Therefore, the fundamental challenge in designing message transmission schemes over LTV channels is the condition that no CSI

is accessible at the receiver beforehand. Approaches involving channel estimations with training symbols would not be precise enough since the estimated channel does not stay constant during data transmission stage [14]. Therefore, a transmission scheme for a LTV channel needs to directly estimate both the data and the CSI from the received signal simultaneously.
In 1946 Dennis Gabor published his paper on communication theory [15] whose underlying ideas significantly shaped the development of modern communication theory [11, 16–18]. Additionally, Gabor's work has led to the development of Gabor frames which are generated by time-frequency shifts of a single seed signal [19]. Therefore, in order to design and analyse message transmission schemes over LTV channels we need to study the properties of Gabor frames, since the signal obtained at the receiver over a LTV channel is the superposition of delay-Doppler shifted copies of the input signal. Hence, the superposition of atoms of a Gabor frame generated by the input signal.

1.1. Motivational example

To get a first flavour of the ideas involved in Sections 9, 10 and 11, devoted to message transmission over LTV channels without prior CSI, let us consider the time-invariant case and ask the question whether it is possible to transmit data over an LTI channel without prior knowledge of the CSI at the receiver. In order to emphasise the core idea the following discussion will be restricted to the noiseless cases. Let $L^2(0,1)$ be the set of periodic square integrable functions, i.e. for $f \in L^2(0,1)$ we have $\|f\|^2 = \int_0^1 |f(t)|dt < \infty$ and the inner product with $f, g \in L^2(0,1)$ given by $\langle f, g\rangle = \int_0^1 f(t)\overline{g(t)}dt$. A fading multipath channel can be described by an LTI system [10] given by

$$y(t) = \sum_{\ell=0}^{P-1} \alpha_\ell s(t-\tau_\ell)\,, \tag{1.1}$$

where $y(t)$ is the received signal, $s(t)$ is the signal fed into the channel, P is the number of paths and α_ℓ and τ_ℓ are the path attenuation factors and the path delays, respectively. Let $\mathrm{T}_\tau : L^2(0,1) \to L^2(0,1)$ denote the time-delay operator that is defined by $\mathrm{T}_\tau s\,(t) = s\,(t-\tau)$, then (1.1) can be equivalently written by

$$y(t) = \sum_{\ell=0}^{P-1} \alpha_\ell \mathrm{T}_{\tau_\ell} s\,(t)\ .$$

Consider the function $s(t) = \mathrm{e}^{j2\pi at}$ with $a \in \mathbb{R}$ for $t \in [0, 1)$ then we have

$$\mathrm{T}_\tau s\,(t) = \mathrm{e}^{j2\pi a(t-\tau)} = \mathrm{e}^{j2\pi at}\mathrm{e}^{-j2\pi a\tau} = s(t)\mathrm{e}^{-j2\pi a\tau}\,, \tag{1.2}$$

hence the functions $\mathrm{e}^{j2\pi at}$ are common eigenfunctions of the time-delay operators T_τ with eigenvalues $\mathrm{e}^{-j2\pi a\tau}$. Further, if $a \in \mathbb{Z}$ we have the orthogonality property

$$\left\langle \mathrm{e}^{j2\pi at}, \mathrm{e}^{j2\pi a't} \right\rangle = \begin{cases} 1, & \text{if } a' = a\,, \\ 0, & \text{if } a' \neq a\,. \end{cases} \tag{1.3}$$

Now say we have K message symbols $\{m_0, \dots m_{K-1}\}$ with $m_k \in \mathbb{C}$ for $k = 0, \dots, K-1$ which we would like to transmit via the channel described in (1.1) but we don't have CSI, i.e. we don't know the channel parameters τ_ℓ and α_ℓ. To this end consider the signal

$$s(t) = p(t) + m(t) \quad \text{for } t \in [0, 1)\ , \tag{1.4}$$

where $p\,(t)$ and $m\,(t)$ are given by

$$p(t) = \sum_{\ell=0}^{N-1} \mathrm{e}^{j2\pi\ell t}\,, \quad m(t) = \sum_{\ell=0}^{K-1} m_\ell \mathrm{e}^{j2\pi(N+\ell)t}\,, \tag{1.5}$$

The signal $s\,(t)$ is a superposition of two separate signals, the pilot signal $p\,(t)$ and the data signal $m\,(t)$ that contains the message symbols. The signals $p\,(t)$ and $m\,(t)$ occupy disjoint frequencies. Plugging (1.4) into (1.1) we get

$$\begin{aligned} y(t) &= \sum_{\ell=0}^{P-1} \alpha_\ell \mathrm{T}_{\tau_\ell} s(t) = \sum_{\ell=0}^{P-1} \alpha_\ell \mathrm{T}_{\tau_\ell} p(t) + \sum_{\ell=0}^{P-1} \alpha_\ell \mathrm{T}_{\tau_\ell} m(t) \\ &= \sum_{\ell=0}^{P-1} \alpha_\ell \sum_{n=0}^{N-1} \mathrm{e}^{-j2\pi n\tau_\ell} \mathrm{e}^{j2\pi nt} + \sum_{\ell=0}^{P-1} \alpha_\ell \sum_{k=0}^{K-1} m_k \mathrm{e}^{-j2\pi(N+k)\tau_\ell} \mathrm{e}^{j2\pi(N+k)t}\,, \end{aligned}$$

where we used (1.2). Using the orthogonality relation of harmonics in (1.3) we obtain

$$h_n = \left\langle y(t), \mathrm{e}^{j2\pi nt} \right\rangle = \sum_{\ell=0}^{P-1} \alpha_\ell \mathrm{e}^{-j2\pi n\tau_\ell}\,, \quad \text{for } n = 0, \dots, N-1\,, \tag{1.6}$$

$$d_k = \left\langle y(t), \mathrm{e}^{j2\pi(N+k)t} \right\rangle = m_k \sum_{\ell=0}^{P-1} \alpha_\ell \mathrm{e}^{-j2\pi(N+k)\tau_\ell}\,, \quad \text{for } k = 0, \dots, K-1\,. \tag{1.7}$$

Thus, we have split the received signal $y(t)$ into two separate set of coefficients, $\{h_n\}_{n=0}^{N-1}$ given in (1.6) which describes the channel and $\{d_k\}_{k=0}^{K-1}$ given in (1.7) which describe the received data.
The problem of recovering the parameters α_ℓ and τ_ℓ from measurements h_n is well known in the literature as line spectral estimation [20]. The methods to solve the problem in (1.6) range from Prony's method (or stable variants thereof) [21, 22] to atomic norm minimization [20]. Thus, the recovery problem in (1.6) can be solved by standard methods found in the literature (provided that the sample size N is large enough), see for instance [20–23].
After obtaining the parameters $\{\tau_\ell\}_{\ell=0}^{P-1}$ and $\{\alpha_\ell\}_{\ell=0}^{P-1}$ recovering the message symbols m_k from d_k in (1.7) is straight forward.
The idea of transmitting different message symbols over different frequencies that are orthogonal to each other, as we did with the data carrying signal $m(t)$, is not new and is known in the literature as frequency-division multiplexing [24]. As promised we avoided a two step scheme of first estimating the channel and then transmitting data over it. The main idea we have employed is to split the data carrying signal $m(t)$ and the pilot signal $p(t)$ into orthogonal bases and to this end we mainly used the idea from frequency-division multiplexing of choosing basis functions that have disjoint support in the frequency domain. In two-step schemes the pilot signal and the message carrying signals are also supported by functions that are orthogonal to each other and the orthogonality arises from the fact that the pilot signal and message carrying signal have disjoint supports in the time domain [1]. However, the main difference between the standard two-step scheme and the one discussed in this section is, that the orthogonality of the pilot signal and message carrying signal is preserved after passing through the communication channel. Hence, the received signal can be split into two orthogonal parts wherein the coefficients of one part contain the CSI (1.6) and the other part contains the data (1.7). A similar orthogonality condition at the received signal can be obtained for the two-step scheme by employing guard intervals.

1.2. Generalization of the example

Since our aim is to transmit data over LTV channels without prior CSI let us explore whether we are able to extend the method discussed in Section 1.1 to more channel types then the one given in (1.1).
Let $\mathrm{M}_\nu : L^2(0,1) \to L^2(0,1)$ denote the Doppler-shift operator that is defined by $\mathrm{M}_\nu s\,(t) = \mathrm{e}^{j2\pi\nu t} s\,(t)$. A time-varying communication channel can be described

as an LTV system [3, 25] given by

$$y\left(t\right)=\sum_{\ell=0}^{P-1}\alpha_{\ell}\mathrm{M}_{\nu_{\ell}}\mathrm{T}_{\tau_{\ell}}s(t)\,,\quad\text{with }(\nu_{\ell},\tau_{\ell})\in\mathcal{S}\,,\tag{1.8}$$

where $y(t)$ is the received signal, $s(t)$ is the input signal, P the number of paths, $\mathcal{S}$ the support of the channel in the delay-Doppler domain and α_ℓ and (ν_ℓ,τ_ℓ) are the path attenuation and delay-Doppler shift corresponding to the ℓ-th path.
Now the question is whether we can find sets $\mathcal{S}$ with $(\nu_\ell,\tau_\ell)\in\mathcal{S}$ such that we can guarantee an orthogonal splitting condition for the received signal $y(t)$ in (1.8) similar to (1.6) and (1.7).
To this end let us consider the chirp signals, which are commonly used in radar [26]. The chirp signal with base frequency $b\in\mathbb{R}$ and chirp rate $r\in\mathbb{R}$ is given by $c_{b,r}(t)=\mathrm{e}^{j2\pi\left(bt^2+rt\right)}$. If $r,r'\in\mathbb{Z}$ then for two chirp signals with the same base frequency we have the orthogonality relation

$$\langle c_{b,r}(t),c_{b,r'}(t)\rangle=\begin{cases}1, & \text{if }\ r'=r\,,\\ 0, & \text{if }\ r'\neq r\,.\end{cases}\tag{1.9}$$

Now let $\nu,\tau,b,r\in\mathbb{R}$ and observe

$$\begin{aligned}\mathrm{M}_{\nu}\mathrm{T}_{\tau}c_{b,r}\left(t\right)&=\mathrm{e}^{j2\pi\nu t}\mathrm{e}^{j2\pi\left(b(t-\tau)^2+r(t-\tau)\right)}\\&=\mathrm{e}^{j2\pi\left(b\tau^2-r\tau\right)}\mathrm{e}^{j2\pi\left(bt^2+(\nu-2b\tau+r)t\right)}\\&=\mathrm{e}^{j2\pi\left(b\tau^2-r\tau\right)}c_{b,\nu-2b\tau+r}\,.\end{aligned}\tag{1.10}$$

From (1.10) we see that the delay-Doppler shift operator $\mathrm{M}_\nu\mathrm{T}_\tau$ does not change the base frequency of the chirp signal and that whenever $\nu-2b\tau=0$ holds the chirp signal $c_{b,r}$ is an eigenfunction of $\mathrm{M}_\nu\mathrm{T}_\tau$. Clearly, the set $\mathcal{L}_{b,0}=\{(\nu,\tau)\ :\ \nu-2b\tau=0\ \text{ for }\ \tau\in\mathbb{R}\}$ for $b\in\mathbb{R}$ defines a line through the origin in the delay-Doppler domain.[1] For $b=0$ the set $\mathcal{L}_{0,0}$ is simply the line defining the delays and we have the case from our original example. All operators $\{\mathrm{M}_\nu\mathrm{T}_\tau\ :\ (\nu,\tau)\in\mathcal{L}_{b,0}\}$ have common eigenfunctions and we can use (1.9) to apply the same method as discussed in Section 1.1.
In the next step let us consider all lines in the delay-Doppler domain. For

[1] In fact, the set of delay-Doppler shift operators $\{\mathrm{M}_\nu\mathrm{T}_\tau\}_{(\nu,\tau)\in\mathcal{L}_{b,0}}$ for $b\in\mathbb{R}$ defines a commutative subgroup of delay-Doppler shift operators, see Lemma 4.2.6 and (8.6) .

$a, b \in \mathbb{R}$, define the set $\mathcal{L}_{b,a}$ as

$$\mathcal{L}_{b,a} = \{(\nu, \tau) \ : \ \nu = 2b\tau + a \ \text{ for } \ \tau \in \mathbb{R}\} \ .$$

For a $(\nu, \tau) \in \mathcal{L}_{b,a}$ we have

$$\mathrm{M}_\nu \mathrm{T}_\tau c_{b,r} = \mathrm{e}^{j2\pi\left(b\tau^2 - r\tau\right)} \mathrm{e}^{j2\pi\left(bt^2 + (r+a)t\right)} = c_{b,r'} \ , \tag{1.11}$$

where we set $r' = r + a$. Note that the delay-Doppler shift operators indexed by the set $\mathcal{L}_{b,a}$ applied on a chirp signal with base frequency b simply changes the chirp rate by shifting it to $r' = r + a$. Hence, if the delay-Doppler shift operators from (1.8) are supported on a line in the delay-Doppler domain and we know the slope b and the intercept a of this line we can apply the idea developed in Section 1.1.
Let the channel be given as in (1.8) where the delay-Doppler operators are supported on the line $\mathcal{L}_{b,a}$, i.e $\mathcal{S} = \mathcal{L}_{b,a}$. We consider a transmit signal $s(t)$ of the form (1.4) where $p\,(t)$ and $m\,(t)$ are given by

$$p(t) = \sum_{n=0}^{N-1} c_{b,n}(t) \, , \quad m(t) = \sum_{k=0}^{K-1} m_k c_{b,N+k}(t) \, ,$$

with m_k for $k = 0, \dots K-1$ being the message symbols. Let $a = z + q$ such that $z \in \mathbb{Z}$ and $q \in [0, 1)$. Then plugging $s(t)$ in (1.8) we get

$$\begin{aligned} y(t) &= \sum_{\ell=0}^{P-1} \alpha_\ell \mathrm{M}_{\nu_\ell} \mathrm{T}_{\tau_\ell} s(t) \\ &= \sum_{\ell=0}^{P-1} \alpha_\ell \sum_{n=0}^{N-1} \mathrm{e}^{j2\pi\left(\tau_\ell^2 - n\tau_\ell\right)} c_{b,n+z+q}(t) + \sum_{\ell=0}^{P-1} \alpha_\ell \sum_{k=0}^{K-1} m_k \mathrm{e}^{j2\pi\left(\tau_\ell^2 - (N+k)\tau_\ell\right)} c_{b,N+k+z+q}(t), \end{aligned}$$

where we used (1.11). Note that $n + z \in \mathbb{Z}$ and $N + k + z \in \mathbb{Z}$, hence we can use the orthogonality property of the chirp signals (1.9) and get the samples

$$h_n = \langle y(t), c_{b,n+z+q}(t) \rangle = \sum_{\ell=0}^{P-1} \alpha_\ell \mathrm{e}^{j2\pi\tau_\ell^2} \mathrm{e}^{-j2\pi\tau_\ell n} = \sum_{\ell=0}^{P-1} \widetilde{\alpha_\ell} \mathrm{e}^{-j2\pi\tau_\ell n} \, , \tag{1.12}$$

$$d_k = \langle y(t), c_{b,N+k+z+q}(t) \rangle = m_k \sum_{\ell=0}^{P-1} \widetilde{\alpha_\ell} \mathrm{e}^{-j2\pi\tau_\ell (N+k)} \, , \tag{1.13}$$

for $n = 0, \dots, N-1$, $k = 0, \dots, K-1$ and $\widetilde{\alpha_\ell} = \alpha_\ell \mathrm{e}^{j2\pi\tau_\ell^2}$. Similar to Section 1.1 we have split the received signal into coefficients only depending on the channel

parameters (1.12) and into coefficients that depend on the message symbols (1.13). Again by applying methods from line spectral estimation [20–23] we can recover from the coefficients h_n the parameters $\widetilde{\alpha_\ell}$ and τ_ℓ, provided N is sufficiently large. Then, in a second step recovering the message symbols from the coefficients d_k is straight forward.
Again we have a splitting into orthogonal bases of the pilot signal and message carrying signal. In this case the splitting is neither in time domain nor in the frequency domain but on some basis of chirp signals.

The idea of splitting the received signal into a channel and data component will accompany us throughout the chapters devoted to data transmission over LTV channels. Contrary to the discussion here we won't be using orthogonal bases but Gabor frames with low correlation properties. We will apply methods from sparse signal recovery to split the received signal into its components and recover the channel parameters as well as the transmitted message symbols.

The aim of this thesis is to develop message transmission schemes for LTV communication channels. To this end, first some preliminaries are discussed in Sections 2 and 3. Then properties of finite Gabor frames are recapped and studied in Sections 4, 5 and 6. Wherein some relevant properties of finite Gabor frames are derived in Sections 5 and 6. Finally, based on these insights message transmission schemes for LTV systems are developed and analysed in Sections 9, 10 and 11.
Parts of this thesis have already been published. Sections 5 - 11 have been partially published in [27–36].

Part I.

Preliminaries

2. Compressed Sensing

Compressed sensing (CS) is a method for recovery of sparse signals [37–40], which has inspired substantial research in signal processing in the past decade [8, 41–46]. The aim of this section is to give a short introduction on the idea of compressed sensing and develop some intuition for its methods and conditions which will be relevant in the later chapters. An indepth discussion and overview of compressed sensing is provided in [37].

2.1. Compressed sensing problem

The standard problem in CS is to solve an underdetermined system of linear equations $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$ with $\mathbf{y} \in \mathbb{C}^L$, $\mathbf{A} \in \mathbb{C}^{L\times N}$ with $L \ll N$, $\mathbf{x} \in \mathbb{C}^N$ and $\mathbf{n} \in \mathbb{C}^L$ some noise vector with $\|\mathbf{n}\|_2 \leq \epsilon$. Note that $\mathbf{A}$ is referred to as the measurement matrix. Since we consider an underdetermined linear system there exits infinitely many solutions. However, the core observation lying at the heart of CS is that under certain conditions the sparsest solution to the underdetermined linear equation is the unique sparse solution to that equation. To this end we define the support set of a vector as follows.

Definition 2.1.1. *The support of a vector* $\mathbf{x} \in \mathbb{C}^N$ *is the index set of its nonzero entries*

$$\mathrm{supp}(\mathbf{x}) = \{j \in \{0, 1, ..., N-1\} : \mathbf{x}[j] \neq 0\}$$

where the j-th entry of $\mathbf{x}$ *is denoted by* $\mathbf{x}[j]$*. The vector* $\mathbf{x}$ *is called k-sparse if* $|\mathrm{supp}(\mathbf{x})| \leqslant k$ *holds.*

In order to derive a criteria for the uniqueness of a sparse solution to an underdetermined linear system we first need to define the spark of a matrix.

Definition 2.1.2. *The number* $\mathrm{spark}(\mathbf{A})$ *of a matrix* $\mathbf{A} \in \mathbb{C}^{L\times N}$ *is the smallest number of linear dependent columns(vectors), i.e.*

$$\mathrm{spark}(\mathbf{A}) = \min\{|\mathrm{supp}\,(\mathbf{x})| : \mathbf{x} \in \mathbb{C}^N \text{ with } \mathbf{x} \in \ker(\mathbf{A}) \setminus \{0\}\}.$$

Essentially, the spark is the minimum number of columns in a matrix that are linearly dependent. Equivalently, spark is the support size of the sparsest vector in the nontrivial kernel of the matrix, as defined in Definition 2.1.2. Note that the kernel of a matrix $\mathbf{A} \in \mathbb{C}^{L\times N}$ is defined by $\ker(\mathbf{A}) = \{\mathbf{x} \in \mathbb{C}^N : \mathbf{A}\mathbf{x} = 0\}$. Let $\mathbf{A} \in \mathbb{C}^{L\times N}$, following [37] we formulate the sparse recovery problem as follows,

$$\underset{\mathbf{z}\in\mathbb{C}^N}{\text{minimize}} \ |\text{supp}(\mathbf{z})| \quad \text{subject to} \quad \mathbf{A}\mathbf{z} = \mathbf{y}. \tag{P_0}$$

Finally, we can formulate a uniqueness criteria for sparse solutions of (P_0)-problem.

Theorem 2.1.1. *The (P_0)-problem has a unique sparse solution $\mathbf{x} \in \mathbb{C}^N$ with $\mathbf{A}\mathbf{x} = \mathbf{y}$ if*

$$|\text{supp}(\mathbf{x})| < \frac{\text{spark}(\mathbf{A})}{2}.$$

Proof. For a contradiction assume there exist a vector $\mathbf{x}_* \neq \mathbf{x}$ with $|\text{supp}(\mathbf{x}_*)| < \text{spark}(\mathbf{A})/2$ and $\mathbf{y} = \mathbf{A}\mathbf{x}_*$, then we have $\mathbf{A}(\mathbf{x} - \mathbf{x}_*) = 0$ and $\mathbf{x} - \mathbf{x}_* \in \ker(\mathbf{A})$ by using the definition of spark (Definition 2.1.2) we obtain

$$\text{spark}(\mathbf{A}) \leq |\text{supp}(\mathbf{x} - \mathbf{x}_*)| \leq |\text{supp}(\mathbf{x})| + |\text{supp}(\mathbf{x}_*)| < \text{sprak}(\mathbf{A})$$

which is a contradiction. Thus, $\mathbf{x}$ is the sparsest solution to (P_0). □

Theorem 2.1.1 guarantees that if a solution $\mathbf{x}$ to (P_0) fulfils $|\text{supp}(\mathbf{x})| < \text{spark}(\mathbf{A})/2$ then $\mathbf{x}$ is guaranteed to be the unique sparsest solution.
However, the (P_0)-problem is unfeasible for practical applications since its an NP-hard problem [37, 47].

2.2. Null space property

The previously discussed (P_0)-problem being an NP-hard problem motivated the search for possible feasible solution approaches to the sparse recovery problem. Indeed, already in [47] a greedy heuristic method is suggested to solve the (P_0)-problem. Later in several other papers convex relaxation of the (P_0)-problem was successfully analysed [37, 39, 48–51].
In this subsection we will present the *Null Space Property* (NSP) which is a condition on the measurement matrix that is sufficient and necessary to replace the (P_0)-problem by a convex relaxation and obtain the same unique sparse solution. To present the convex relaxation we first define the ℓ_p-norms.

Definition 2.2.1. *The ℓ_p-norm of a vector $\mathbf{x} \in \mathbb{C}^n$ is given by the positive functional $\|\cdot\|_p : \mathbb{C}^n \to \mathbb{R}_+$ with the formula:*

$$\|\mathbf{x}\|_p := \begin{cases} (\sum_{i=1}^{n} |\mathbf{x}[i]|^p)^{\frac{1}{p}} & \text{for } 0 < p < \infty \\ \max_{i \in \{1,\dots,n\}} |\mathbf{x}[i]| & \text{for } p = \infty \end{cases}$$

Next we state the basis pursuit or the ℓ_1-minimization problem.

$$\underset{\mathbf{z} \in \mathbb{C}^N}{\text{minimize}} \|\mathbf{z}\|_1 \quad \text{subject to } \mathbf{A}\mathbf{z} = \mathbf{y} \tag{P_1}$$

But for the main part of our analysis we will consider a more general case of (P_1) which also considers noisy measurements.

$$\underset{\mathbf{z} \in \mathbb{C}^N}{\text{minimize}} \|\mathbf{z}\|_1 \quad \text{subject to } \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leqslant \epsilon \tag{$P_{1,\epsilon}$}$$

Where ϵ is the noise level. Note that setting $\epsilon = 0$ in $(P_{1,\epsilon})$ would lead to (P_1). Next, following [37], we define the NSP which is a necessary and sufficient condition for (P_0) and (P_1) to share the same solution.

Definition 2.2.2. *A matrix $\mathbf{A} \in \mathbb{C}^{L \times N}$ is said to satisfy the NSP of order k if for any set $\mathcal{S} \subset \{0, 1, ..., N-1\}$ with $|\mathcal{S}| \leqslant k$,*

$$\|\mathbf{v}(\mathcal{S})\|_1 < \|\mathbf{v}(\mathcal{S}^c)\|_1 \quad \forall \mathbf{v} \in \ker \mathbf{A} \setminus \{0\}$$

holds. Where $\mathcal{S}^c = \{0, 1, ..., N-1\} \setminus \mathcal{S}$ and $\mathbf{v}(\mathcal{S}) = \hat{\mathbf{v}}$ such that

$$\hat{\mathbf{v}}[j] = \begin{cases} \mathbf{v}[j] & \text{for } j \in \mathcal{S} \\ 0 & \text{for } j \in \mathcal{S}^c \end{cases}$$

obviously it holds $\mathbf{v}(\mathcal{S}) + \mathbf{v}(\mathcal{S}^c) = \mathbf{v}$.

The following theorem from [37] gives the connection between NSP of the specific sensing matrix, (P_1) and (P_0).

Theorem 2.2.1. *Given a matrix $\mathbf{A} \in \mathbb{C}^{L \times N}$, every k-sparse vector $\mathbf{x} \in \mathbb{C}^N$ is the unique solution of (P_1) and (P_0) with $\mathbf{y} = \mathbf{A}\mathbf{x}$ if and only if $\mathbf{A}$ satisfies the NSP of order k.*

Proof. See [37, Chapter 4] proof of Theorem 4.4 and Remark 4.6 (a). □

Although Theorem 2.2.1 delivers a nice connection between NSP and the unique solution of problems (P_1) and (P_0) it is not convincing for practical applications,

since it is not possible to measure a signal $\mathbf{x} \in \mathbb{C}^N$ with infinite precision in real world settings. Therefore an additional strengthening of the NSP is needed that enables statements on the solution of the ℓ_1-minimization problem as defined in $(P_{1,\epsilon})$. For this reason we consider *nearly* k-sparse vectors, measured by the error of best k-term approximation. We begin with the following definition.

Definition 2.2.3. *For $p > 0$, the ℓ_p-error of best k-term approximation to a vector $\mathbf{x} \in \mathbb{C}^N$ is defined by*

$$\sigma_k(\mathbf{x})_p = \inf\{\|\mathbf{x} - \mathbf{z}\|_p, \mathbf{z} \in \mathbb{C}^N \text{ is } k\text{-sparse } \}$$

Next we define a stronger NSP, the so called ℓ_2-robust NSP.

Definition 2.2.4. *The sensing matrix $\mathbf{A} \in \mathbb{C}^{L \times N}$ is said to satisfy the ℓ_2-robust NSP of order k (with respect to $\|\cdot\|_2$) with constants $0 < \rho < 1$ and $\tau > 0$ if, for any set $\mathcal{S} \subset \{0, 1, ..., N-1\}$ with $|\mathcal{S}| \leqslant k$,*

$$\|\mathbf{v}(\mathcal{S})\|_2 \leq \frac{\rho}{\sqrt{k}}\|\mathbf{v}(\mathcal{S}^c)\|_1 + \tau\|\mathbf{A}\mathbf{v}\|_2, \quad \forall \mathbf{v} \in \mathbb{C}^N$$

holds.

Equipped with the two definitions we can now state the next theorem from [37, Thm. 4.22].

Theorem 2.2.2. *Suppose that the sensing matrix $\mathbf{A} \in \mathbb{C}^{L \times N}$ satisfies the ℓ_2-robust NSP of order k with constants $0 < \rho < 1$ and $\tau > 0$. Then, for any $\mathbf{x} \in \mathbb{C}^N$, a solution $\mathbf{x}_*$ of the problem $(P_{1,\epsilon})$ with $\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leqslant \epsilon$ approximates the vector $\mathbf{x}$ with the ℓ_2-error*

$$\|\mathbf{x} - \mathbf{x}_*\|_2 \leq 2\frac{1+\rho}{1-\rho}\sigma_k(\mathbf{x})_1 + \frac{4\tau}{1-\rho}\epsilon$$

for some constants, $C, D > 0$ depending only on ρ and τ.

Proof. See [37, Chapter 4] proof of Theorem 4.25. □

Note that the Theorem 2.2.2 gives an ℓ_2-error approximation of (P_1), since (P_1) and (P_0) have the same solution, if the sensing matrix fulfills the NSP - as stated in Theorem 2.2.1 - follows that $(P_{1,\epsilon})$ delivers an approximation for the ℓ_0-minimization problem (P_0) as well, if the sensing matrix fulfills the ℓ_2-robust NSP.

2.3. Restricted isometry property

Another important property in CS is the *restricted isometry property* (RIP) which implies the robust NSP [37, 39, 40, 50–52]. In fact random matrices exhibit this property with high probability if the dimension is large enough [37, 50, 51, 53, 54]. We start by giving a formal definition of RIP.

Definition 2.3.1. *The k-th restricted isometry constant δ_k of a sensing matrix $\mathbf{A} \in \mathbb{C}^{L\times N}$ is the smallest $\delta \geq 0$ such that*

$$(1-\delta)\,\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1+\delta)\,\|\mathbf{x}\|_2^2$$

for all k-sparse vectors $\mathbf{x} \in \mathbb{C}^N$. Then we say that $\mathbf{A}$ fulfills the k-th RIP.

In [52] it is proven that if a matrix fulfils the RIP with $\delta \leq 1/\sqrt{2}$ then that matrix satisfies the robust NSP. The following theorem can be found with its proof in [40].

Theorem 2.3.1. *If a matrix $\mathbf{A} \in \mathbb{C}^{L\times N}$ fulfils the restricted isometry property with the restricted isometry constant δ such that δ satisfies*

$$\delta_{2k} \leq \frac{1}{\sqrt{2}}$$

then the matrix $\mathbf{A}$ satisfies the robust NSP of order k with constants

$$\rho = \frac{\delta}{\sqrt{1-\delta^2}} \quad \textit{and} \quad \tau = \frac{2\sqrt{k}}{(1-\delta)\sqrt{1+\delta}}\,.$$

As we see from Theorem 2.3.1 the RIP via the NSP provides us with sparse recovery guarantees. More precisely if a measurement matrix has $\delta_{2k} \leq 1/\sqrt{2}$ then recovery of k-sparse vectors are guaranteed by the NSP property, i.e. Theorem 2.2.1 and Theorem 2.2.2. For a random measurement matrix, where every entry is for instance a Gaussian random variable, it is very likely that such a matrix is well conditioned, i.e. has a low restricted isometry constant. In fact, in that case the sparsity of the signal scales proportional (up to logarithmic factors) to the dimension of the measurement space L, i.e. $k \sim \frac{L}{\log(N)}$ [37, 53, 54]. Although random matrices are well conditioned and exhibit the RIP [53, 54], checking whether a given matrix has RIP or NSP is NP-hard and therefore not feasible [55, 56].

2.4. Coherence

So far in the previous sections we only discussed criteria for measurement matrices that enables sparse signal recovery. From a practical point of view, deterministic constructions of measurement matrices for CS is a very appealing problem and has attracted quite some attention [29, 57–62]. Most of these approaches (for instance [29, 57–60]) rely on the frame work of coherence which has been introduced to the sparse recovery framework in [49, 63]. The coherence of a matrix delivers a simple quantity to check if a given matrix is suitable as a measurement matrix for CS. In general, the smaller the coherence the better the recovery algorithm performs.
However, unlike the RIP the analysis of recovery algorithms based on the coherence delivers worse results. This is due to the Welch bound [64, 65], which lower bounds the coherence of a matrix. We can upper bound the restricted isometry constant of a matrix by using the coherence and the Gershgorin circle theorem [37, 61, 66].
We start by defining the coherence property as follows.

Definition 2.4.1. *Let $\mathbf{A} \in \mathbb{C}^{L\times N}$ be a matrix with ℓ_2-normalized columns and let each column of $\mathbf{A}$ be denoted by $\mathbf{a}_1, ..., \mathbf{a}_N$, i.e. $\|\mathbf{a}_i\|_2 = 1$ for all $i \in \{1, ..., N\}$. The coherence of the matrix $\mathbf{A}$ is defined as*

$$\mu := \max_{1\leqslant i\neq j\leqslant N} |\langle \mathbf{a}_i, \mathbf{a}_j\rangle|$$

Before continuing we are going to need the Gershgorin circle theorem which will be stated in the following lemma, a proof of this result can be found in [66].

Lemma 2.4.1. *Let λ be an eigenvalue of a square matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$. There exists an index $j \in \{1, \dots, n\}$ such that*

$$|\lambda - \mathbf{A}_{j,j}| \leq \sum_{l\in\{1,\dots,n\}\setminus\{j\}} |\mathbf{A}_{j,l}| \,,$$

where $\mathbf{A}_{j,l}$ denotes the entry in $\mathbf{A}$ on j-th row and l-th column.

Next we give a well known result in CS [37, 61].

Theorem 2.4.2. *Let $\mathbf{A} \in \mathbb{C}^{L\times N}$ be a matrix with ℓ_2-normalized columns and coherence μ. Then the k-th restricted isometry constant δ_k of $\mathbf{A}$ can be upperbounded by*

$$\delta_k \leq (k-1)\mu \,.$$

Proof. For a set $\mathcal{S} = \{s_1, \dots, s_{|\mathcal{S}|}\} \subset \{1, \dots, N\}$ with $|\mathcal{S}| \leq k$, we define the matrix $\mathbf{G} = \mathbf{A}_{\mathcal{S}}^* \mathbf{A}_{\mathcal{S}}$ with $\mathbf{A}_{\mathcal{S}} = \left[\mathbf{a}_{s_1}, \dots, \mathbf{a}_{|\mathcal{S}|}\right]$. $\mathbf{G}$ is a positive semidefinite matrix, hence it has an orthonormal basis of eigenvectors associated with real, positive eigenvalues. Denote the minimal eigenvalue by $\lambda_{\min}$ and the maximal eigenvalue by $\lambda_{\max}$. Then, for a $|\mathcal{S}|$-sparse vector $\mathbf{x} \in \mathbb{C}^N$ with $\operatorname{supp}(\mathbf{x}) = \mathcal{S}$ we have $\mathbf{A}\mathbf{x} = \mathbf{A}_{\mathcal{S}}\mathbf{x}$. Hence, for $\lambda_{\min}$ and $\lambda_{\max}$ we have

$$\lambda_{\max} = \max_{\substack{\operatorname{supp}(\mathbf{x}) \subset \mathcal{S} \\ \|\mathbf{x}\|_2 = 1}} \|\mathbf{A}\mathbf{x}\|_2^2 \quad \text{and} \quad \lambda_{\min} = \min_{\substack{\operatorname{supp}(\mathbf{x}) \subset \mathcal{S} \\ \|\mathbf{x}\|_2 = 1}} \|\mathbf{A}\mathbf{x}\|_2^2 \,.$$

Now, due to the ℓ_2-normalization and Gershgorin's circle theorem we have that the eigenvalues of $\mathbf{G}$ must be contained in the union of disks centered at 1 with radii

$$r_j = \sum_{\substack{l=1 \\ l \neq j}}^{|\mathcal{S}|} |\mathbf{G}[j,l]| = \sum_{\substack{l \in \mathcal{S} \\ l \neq j}} |\langle \mathbf{a}_l, \mathbf{a}_j \rangle| \leq (k-1)\mu \,,\; j \in \mathcal{S} \,,$$

where $\mathbf{G}[j,l]$ denotes the entry of $\mathbf{G}$ in j-th row and l-th column.
Since all the eigenvalues of $\mathbf{G}$ are real, they all must lie within the interval of $[1-(k-1)\mu, 1+(k-1)\mu]$, hence we obtain the estimation in the theorem. □

From Theorem 2.4.2 we see that the smaller the coherence is the bigger the sparsity of our vectors can be. However, the coherence cannot be arbitrary small, for instance if the coherence would be equal to zero then that would correspond the columns of the measurement matrix being mutually orthogonal. In fact there is a limitation on how small the coherence can be, note that we consider measurement matrices which have more columns then rows. In [64] a lower bound for the coherence is established this bound is called the Welch bound [37, 64, 65].

Theorem 2.4.3. *The coherence of a matrix $\mathbf{A} \in \mathbb{C}^{L \times N}$ with ℓ_2-normalized columns satisfies*

$$\mu \geq \sqrt{\frac{N-L}{L(N-1)}} \tag{2.1}$$

Equality holds if and only if the columns $\mathbf{a}_1, ..., \mathbf{a}_N$ of the matrix $\mathbf{A}$ are equiangular. Equiangularity of the columns is defined by

$$|\langle \mathbf{a}_i, \mathbf{a}_j \rangle| = c, i \neq j, \forall i, j \in \{1, ..., N\}$$

with some constant $c > 0$.

The Welch bound (Theorem 2.4.3) limits the sparsity level one can reach applying Theorem 2.4.2 to a quadratic factor of the measurement dimension, i.e. if $\mathbf{A}$ is a $L \times N$ measurement matrix then the highest sparsity k a signal can have is $k \sim \sqrt{L}$. This relation is known in the literature as the quadratic bottleneck. So far no deterministic constructions are known which are significantly better. To this date the only known deterministic construction that overcomes the quadratic bottleneck is described in the breakthrough paper [62]. There the sparsity level is $k \sim L^{0,5+\epsilon}$ for some small value of ϵ.

3. Statistical Sparse Signal Recovery

As discussed previously in Section 2.4, Theorem 2.4.2 and Theorem 2.4.3 severely limit the sparsity levels of signals to a square-root factor of the ambient dimension. However, in the literature there are plenty reports of deterministic sensing matrices that perform comparable to random sensing matrices in numerical simulations [27, 29, 30, 43, 59, 60, 67–70]. In these numerical simulations a random signal is generated and the reconstruction performance is analysed based on which measurement matrix is used. Deterministic matrices being able to keep up with the random matrices in these simulations led to the analysis of these simulation setups. The main idea is to evaluate how well a deterministic matrix is suited to be used as a measurement matrix in CS by analysing how well it performs when recovering a sparse signal with random support and random entries [67, 70–74]. In this section we are going to discuss two criteria the *statistical restricted isometry property* [70] and the *strong coherence property* [67].

3.1. Statistical restricted isometry property

The notion of *statistical restricted isometry property* (StRIP) that we are going to use was introduced in [70]. The StRIP is a statistical version of the RIP which aims to analyse how well certain deterministic sensing matrices fit into the compressed sensing framework. Since the sensing matrices are deterministic, the probability enters the signal model. In this subsection our aim is to give a short overview of the StRIP as discussed in [70]. For this section we denote the set of all k sparse vectors in $\mathbb{C}^N$ by $\mathcal{K}_k^N$, i.e. $\mathcal{K}_k^N = \left\{\mathbf{z} \in \mathbb{C}^N \ : \ |\mathrm{supp}\,(\mathbf{z})| \leq k\right\}$.

Definition 3.1.1. *A matrix* $\mathbf{A} = \frac{1}{\sqrt{m}}\mathbf{\Phi} \in \mathbb{C}^{L\times N}$ *with* ℓ_2*-normalized columns is said to have* (k, δ, ϵ)*-StRIP if*

$$(1-\delta)\left\|\mathbf{x}\right\|_2^2 \leq \left\|\mathbf{A}\mathbf{x}\right\|_2^2 \leq (1+\delta)\left\|\mathbf{x}\right\|_2^2$$

holds with probability exceeding $1-\epsilon$ *for a random vector* $\mathbf{x} \in \mathcal{K}_k^N$ *drawn from a uniform distribution over all* $\left\{\mathbf{z} \in \mathcal{K}_k^N \ : \ \|\mathbf{z}\|_2 = 1\right\}$. *Further we say that* $\mathbf{A}$ *has* (k,δ,ϵ)*-uniqueness guaranteed StRIP (abbr.* (s,δ,ϵ)*-UStRIP) if*

$$\left\{\mathbf{z} \in \mathcal{K}_k^N \ : \ \mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}\right\} = \{\mathbf{x}\}$$

is also satisfied with probability exceeding $1-\epsilon$.

Note that measurement matrices having (s,δ,ϵ)-StRIP satisfy RIP with high probability but they do not guarantee unique recovery by standard CS recovery algorithms. Unique recovery is only guaranteed (with high probability) for UStRIP-matrices.
In [70] the following conditions are introduced to identify matrices satisfying UStRIP.

Definition 3.1.2. *A matrix* $\mathbf{A} = \frac{1}{\sqrt{m}}\mathbf{\Phi} \in \mathbb{C}^{L\times N}$ *with all entries of* $\mathbf{\Phi}$ *having absolute value* 1, *is said to be* η*-StRIP-able if the following conditions are satisfied with* $\eta > 0$.

(S1) *The rows of* $\mathbf{\Phi}$ *are mutually orthogonal, and the sum of all entries in each row is zero, i.e.,*

$$\sum_{j=1}^{N} \boldsymbol{\phi}_j[k]\,\overline{\boldsymbol{\phi}_j[\ell]} = 0 \qquad \text{if } k \neq \ell,$$
$$\sum_{j=1}^{N} \boldsymbol{\phi}_j[k] = 0 \qquad \text{for all } k = 1,\ldots,L\,,$$

where $\boldsymbol{\phi}_j$ *is the* j*-th column of* $\mathbf{\Phi}$ *and* $\boldsymbol{\phi}_j[k]$ *is the* k*-th entry in* $\boldsymbol{\phi}_j$.

(S2) *The columns of* $\mathbf{\Phi}$ *form a group under pointwise multiplication, i.e.*

For all $j,\ell \in \{1,\ldots,N\}$ *there exists* $r \in \{1,\ldots,N\}$ *such that* $\boldsymbol{\phi}_j[k]\boldsymbol{\phi}_\ell[k] = \boldsymbol{\phi}_r[k]$ *for all* $k \in \{1,\ldots,L\}$.

In particular, there is one column of $\mathbf{\Phi}$ *with all its entries equal to* 1. *Without loss of generality, we assume that* $\boldsymbol{\phi}_1$ *is this identity vector.*

(S3) *There exists* $\eta > 0$ *such that*

$$\left|\sum_{k=1}^{L} \boldsymbol{\phi}_j(k)\right|^2 \leq L^{2-\eta} \quad \text{for all } j = 2,3,\ldots,N\,.$$

Note that there is a direct relation between the coherence μ and the parameter η defined in (S3). We simply consider the inner product of two columns of $\mathbf{\Phi}$

$$\begin{aligned}\mu \geq \left|\left\langle \boldsymbol{\phi}_j, \boldsymbol{\phi}_\ell \right\rangle\right| &= \left|\sum_{k=1}^{L} \boldsymbol{\phi}_j[k]\overline{\boldsymbol{\phi}_\ell[k]}\right| \\ &= \left|\sum_{k=1}^{L} \boldsymbol{\phi}_j[k]\boldsymbol{\phi}_{-\ell}[k]\right| = \left|\sum_{k=1}^{L} \boldsymbol{\phi}_r[k]\right| \leq \sqrt{L^{2-\eta}}\,,\end{aligned}$$

where we have used the group properties of the columns of $\mathbf{\Phi}$ and $\boldsymbol{\phi}_{-\ell}$ is the pointwise multiplicative inverse of $\boldsymbol{\phi}_\ell$, i.e. $\boldsymbol{\phi}_{-\ell}[k]\boldsymbol{\phi}_\ell[k] = \boldsymbol{\phi}_1[k]$. Note that $\boldsymbol{\phi}_{-\ell}$ must exist because of the group property (S2). Therefore, large η implies small coherence μ.
The following theorem is proved in [70].

Theorem 3.1.1 (Theorem 8 in [70]). *Let $\mathbf{A} = \frac{1}{\sqrt{L}}\mathbf{\Phi} \in \mathbb{C}^{L\times N}$ be an η-StRIP-able matrix with $\eta > 1/2$, and assume that $k < 1 + (N-1)\delta$. Then*

(a) *$\mathbf{A}$ has (k, δ, ϵ) with $\epsilon = 2\exp\left(-\left(\delta - \frac{k-1}{N-1}\right)^2 \frac{L^\eta}{8k}\right)$*

(b) *if additionally $L \geq \frac{ck\log N}{\delta^{-2}}$ for some $c > 0$, then $\mathbf{A}$ has $(k, \delta, 2\epsilon)$-*UStRIP *with $\epsilon = 2\exp\left(-\left(\delta - \frac{k-1}{N-1}\right)^2 \frac{L^\eta}{8k}\right)$.*

Theorem 3.1.1 reduces the search of good deterministic sensing matrices to the problem of finding matrices that satisfy Conditions (S1) - (S3) with $\eta > 1/2$. Whereas it is basically impossible to calculate the restricted isometry constant for a given matrix $\mathbf{A}$, it is fairly easy to check whether a matrix $\mathbf{A}$ satisfied (S1) - (S3), i.e. whether $\mathbf{A}$ is η-StRIP-able.

3.2. Strong coherence property

In [71] the *the One-Step Thresholding* (OST) algorithm is introduced and discussed with respect to its probabilistic success in sparse signal recovery. This analysis is continued in [67] with probabilistic results on the ℓ_2-error of the recovered signal. In both discussions the *strong coherence property* plays a fundamental role. Given a linear system of equations $\mathbf{y} = \mathbf{A} + \mathbf{n}$ with measurement matrix $\mathbf{A} \in \mathbb{C}^{L\times N}$, a sparse signal $\mathbf{x} \in \mathbb{C}^N$ and a noise vector $\mathbf{n} \in \mathbb{C}^L$ where the entries are i.i.d. complex-Gaussian random variables with mean zero and variance σ^2, the OST-algorithm returns an approximation $\hat{\mathbf{x}}$ of $\mathbf{x}$. The OST algorithm is given in Algorithm 1.

Algorithm 1 One-Step Thresholding (OST)

Input: A measurement matrix $\mathbf{A} \in \mathbb{C}^{L \times N}$, a vector $\mathbf{y} \in \mathbb{C}^L$ with $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$ and a threshold $\lambda > 0$

1) $\hat{\mathbf{x}} = \mathbf{0}$

2) $\mathbf{z} = \mathbf{A}^* y$

3) $\hat{\mathcal{S}} = \{i \in \{1, \ldots, L\} \, : \, \mathbf{z}[i] > \lambda\}$

4) $\hat{\mathbf{x}}_{\hat{\mathcal{S}}} = \mathbf{A}_{\hat{\mathcal{S}}}^{+} \mathbf{y}$

Output: An estimate $\hat{\mathbf{x}}$ of the signal $\mathbf{x}$

When $\mathbf{A}$ satisfies the strong coherence property, [67] establishes bounds on the error $\|\mathbf{x} - \hat{\mathbf{x}}\|_2$ with high probability. *The average coherence* Δ of a matrix $\mathbf{A} \in \mathbb{C}^{L \times N}$ is defined by

$$\Delta = \frac{1}{N-1} \max_{i \in \{1,\ldots,N\}} \left| \sum_{j \neq i} \langle \mathbf{a}_i, \mathbf{a}_j \rangle \right| , \tag{3.1}$$

where $\mathbf{a}_j$ is the j-th column of $\mathbf{A}$. Now we can define the strong coherence property.

Definition 3.2.1. *A matrix $\mathbf{A} \in \mathbb{C}^{L \times N}$ with coherence μ, average coherence Δ and unit norm columns is said to satisfy the strong coherence property if*

$$1) \mu \leq \frac{1}{164 \log(N)} \qquad \textit{and} \qquad 2) \; \Delta \leq \frac{\mu}{\sqrt{L}} \, . \tag{3.2}$$

Before we can state the probabilistic recovery guarantee from [67] we need to define some notations. Thus, following [67] we denote the signal-to-noise ratio by

$$\mathrm{SNR} = \frac{\|\mathbf{x}\|_2^2}{\mathbb{E}\left[\|\mathbf{n}\|_2^2\right]} ,$$

where $\mathbb{E}[\cdot]$ denotes the expected value. Furthermore, denote the sets of entries

of $\mathbf{x}$ that lie above a a certain noise floor σ by

$$\mathcal{T}_\sigma(t) = \left\{ k \,:\, |\mathbf{x}\,[k]| > \frac{2\sqrt{2}}{1-t}\sqrt{2\sigma \log N} \right\} \quad \text{for } t \in (0,1)\ ,$$

and the locations of entries of $\mathbf{x}$ that lie above a certain interference floor by

$$\mathcal{T}_\mu = \left\{ k \,:\, |\mathbf{x}\,[k]| > \frac{20}{t}\mu \left\|\mathbf{x}\right\|_2 \sqrt{2 \log N} \right\} \quad \text{for } t \in (0,1)\ .$$

Equipped with these definitions we can finally state the probabilistic recovery guarantee from [67].

Theorem 3.2.1 (Theorem 4 in [67])**.** *Let $\mathbf{A} \in \mathbb{C}^{L\times N}$ be a measurement matrix satisfying the strong coherence property, pick $t \in (0,1)$, and choose $\lambda = \sqrt{2\sigma^2 \log N} \cdot \max\left\{ \frac{10}{t}\mu\sqrt{L\mathrm{SNR}}, \frac{\sqrt{2}}{1-t} \right\}$. Further, suppose $\mathbf{x} \in \mathbb{C}^N$ has support $\mathcal{S}$ with $|\mathcal{S}| \leq k$ drawn uniformly at random from $\{1,\dots,N\}$. Then provided*

$$k \leq \frac{N}{c_1^2 \left\|\mathbf{A}\right\|_2^2 \log N}\ ,$$

where $\left\|\mathbf{A}\right\|_2$ denotes the spectral norm of $\mathbf{A}$, Algorithm 1 produces $\hat{\mathcal{S}}$ such that $\mathcal{T}_\sigma\,(t) \cap \mathcal{T}_\mu\,(t) \subseteq \hat{\mathcal{S}} \subset \mathcal{S}$ and $\hat{\mathbf{x}}$ such that

$$\left\|\mathbf{x} - \hat{\mathbf{x}}\right\|_2 \leq c_2\sqrt{\sigma^2 \left|\hat{\mathcal{S}}\right| \log N} + c_3 \left\|\mathbf{x}_{\mathcal{S}\setminus\hat{\mathcal{S}}}\right\|_2$$

with probability exceeding $1 - \frac{10}{N}$. Finally, defining $\mathcal{T} = \mathcal{T}_\sigma\,(t) \cap \mathcal{T}_\mu\,(t)$, we further have

$$\left\|\mathbf{x} - \hat{\mathbf{x}}\right\|_2 \leq c_2\sqrt{\sigma^2 k \log N} + c_3 \left\|\mathbf{x} - \mathbf{x}_{\mathcal{T}}\right\|_2$$

in the same probability event. Here, $c_1 = 37\mathrm{e}$, $c_2 = \frac{2}{1-\mathrm{e}^{-1/2}}$, and $c_3 = 1 + \frac{\mathrm{e}^{-1/2}}{1-\mathrm{e}^{-1/2}}$ are numerical constants.

The results from Sections 3.1 and 3.2 enables us to use deterministic matrices for Compressed Sensing. Especially, we will be interested in deterministic Gabor matrices since those will be essential in later Sections 9 and 10. Using Gabor matrices as sensing matrices we will use methods from Compressed Sensing to separate channel parameters from messages. Thus, Sections 5 and 6.6 are devoted on proving sparse recovery guarantees for specific deterministic Gabor matrices.

4. Finite Gabor Systems

In 1946 Dennis Gabor suggested to use time-frequency shifts of a single prototype signal to represent communication signals [15]. Since then his work has been very influential on communication theory [11, 16–18]. The idea of Gabor was to assign a time-frequency box of size $(\Delta t, \Delta f)$ to an information-carrying symbol, where Δt and Δf are effective time duration and effective frequency bandwith of the signal, respectively. Using the uncertainty principle [75] he derives that the smallest boxes are achieved by Gaussians. Gabor further divides the time-frequency plane into cells where each cell corresponds to a box and associates each cell with an information-carrying symbol. Basically, he proposes to transmit a signal of the form

$$\Psi(t) = \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_{nk}\, \mathrm{e}^{-\pi \frac{(t-n\Delta t)^2}{2(\Delta t)^2}}\, \mathrm{e}^{2\pi j \frac{kt}{\Delta t}}$$

where $\{c_{nk}\}$ are complex valued information-carrying symbols. In the language of frame theory [76–78] Gabor suggested to transmit using elements of the Gabor frame [79] generated by the window function $g(t) = \mathrm{e}^{-\pi \frac{t^2}{2(\Delta t)^2}}$.
In order to discuss finite Gabor systems and frames we need first some basics from finite frame theory.

4.1. Finite frames

In this subsection we define finite frames and discuss their basic properties. Frames on Hilbert spaces were originally introduced in [77], a general introduction and discussion of frames is given in [76]. This subsection is mostly taken from [78], here we focus on finite dimensional frames and we will consider elements from $\mathbb{C}^L$. We begin with the definition of frames.

Definition 4.1.1. *A family of vectors $\{\mathbf{h}_\ell\}_{\ell=0}^{N-1}$ in $\mathbb{C}^L$ is called a frame, if there exist constants $0 < A \leq B < \infty$ such that*

$$A \left\|\mathbf{x}\right\|_2^2 \leq \sum_{\ell=0}^{N-1} \left|\langle \mathbf{x}, \mathbf{h}_\ell \rangle\right|^2 \leq B \left\|\mathbf{x}\right\|_2^2 \quad \textit{for all } \mathbf{x} \in \mathbb{C}^L\,. \tag{4.1}$$

In the following we will always assume that the largest possible A and the smallest possible B are chosen such that (4.1) is fulfilled. Next we give some notation and name some important classes of frames.

Remark 4.1.2. *The following notations are related to a frame* $\{\mathbf{h}_\ell\}_{\ell=0}^{N-1}$.

1) *The constants* A *and* B *in* (4.1) *are called lower and upper frame bound respectively.*

2) *If* $A = B$ *then the frame is said to be tight.*

3) *If* $A = B = 1$ *then the frame is called a Parseval frame.*

4) *If a* c *exist such that* $|\langle \mathbf{h}_\ell, \mathbf{h}_k \rangle| = c$ *for all* $k \neq \ell$ *then the frame is called to be an equiangular frame.*

5) *The values* $\{\langle \mathbf{x}, \mathbf{h}_\ell \rangle\}_{\ell=0}^{N-1}$ *are called the frame coefficients of the vector* $\mathbf{x}$ *with respect to frame* $\{\mathbf{h}_\ell\}$.

Next we will discuss the analysis, synthesis and the Gramian operator of frames. The analysis operator computes the frame coefficients of a signal, formally we define the synthesis operator as follows.

Definition 4.1.3. *Let* $\{\mathbf{h}_\ell\}_{\ell=0}^{N-1}$ *be a family of vectors in* $\mathbb{C}^L$. *Then the associated analysis operator* $\mathbf{H} : \mathbb{C}^L \to \mathbb{C}^N$ *is defined by*

$$\mathbf{H}\mathbf{x} = \{\langle \mathbf{x}, \mathbf{h}_\ell \rangle\}_{\ell=0}^{N-1} , \quad \mathbf{x} \in \mathbb{C}^L .$$

Next we define the corresponding synthesis operator.

Definition 4.1.4. *Let* $\{\mathbf{h}_\ell\}_{\ell=0}^{N-1}$ *be a family of vectors in* $\mathbb{C}^L$ *with associated analysis operator* $\mathbf{H}$. *Then the associated synthesis operator is defined as the adjoint operator* $\mathbf{H}^* \mathbb{C}^N \to \mathbb{C}^L$ *and is given by*

$$\mathbf{H}^* \mathbf{y} = \sum_{\ell=0}^{N-1} \mathbf{y}[\ell] \mathbf{h}_\ell , \quad \mathbf{y} \in \mathbb{C}^N ,$$

where $\mathbf{y}[\ell]$ *denotes the* ℓ*-th entry in* $\mathbf{y}$.

Note that when we consider vectors in $\mathbb{C}^L$ we consider column vectors, therefore $\mathbf{h}_\ell$ is a column vector. The matrix representation of the synthesis operator $\mathbf{H}^*$ of the frame $\{\mathbf{h}_\ell\}_{\ell=0}^{N-1}$ is straight forward and is given as follows.

$$\mathbf{H}^* = \begin{bmatrix} | & | & \dots & | \\ \mathbf{h}_0 & \mathbf{h}_1 & \dots & \mathbf{h}_{N-1} \\ | & | & \dots & | \end{bmatrix} \in \mathbb{C}^{L\times N}\,.$$

Note that later we will consider finite frames for sparse signal recovery. In these cases when we talk about properties that are introduced in Section 2 for measurement matrices, we will use these properties for the finite frames in question. Essentially, we will use the synthesis operator of a frame $\mathbf{H}^*$ as a measurement matrix for compressed sensing. For instance, when we talk about the coherence of a finite frame we mean the coherence as defined in Definition 2.4.1 of the matrix representation of the synthesis operator $\mathbf{H}^*$ of the finite frame.
Next we define the frame operator which is a fundamental operator and contains crucial properties of the frame. Additionally, the frame operator plays an important role in the reconstruction of a signal from the frame coefficients. The frame operator is defined as follows.

Definition 4.1.5. *Let $\{\mathbf{h}_\ell\}_{\ell=0}^{N-1}$ be a family of vectors in $\mathbb{C}^L$ with associated analysis operator $\mathbf{H}$. Then the associated frame operator $\mathbf{S} : \mathbb{C}^L \to \mathbb{C}^L$ is defined by*

$$\mathbf{S}\mathbf{x} = \mathbf{H}^*\mathbf{H}\mathbf{x} = \sum_{k=0}^{N-1} \langle \mathbf{x}, \mathbf{h}_k\rangle\, \mathbf{h}_k\,, \quad \mathbf{x} \in \mathbb{C}^L\,.$$

Before we continue we need to define some notation on Hermitian matrices. A matrix $\mathbf{M} \in \mathbb{C}^{L\times L}$ is said to be Hermitian if it coincides with its conjugate transpose, i.e. if $\mathbf{M}^* = \mathbf{M}$. The eigenvalues of the Hermitian matrices are all real valued [66]. Next we define the following notation.

- $\mathbf{M}$ is called positive-definite if and only if $\langle \mathbf{M}\mathbf{x}, \mathbf{x}\rangle > 0$ for all $\mathbf{x} \in \mathbb{C}^L \setminus \{\mathbf{0}\}$. In other words, the matrix $\mathbf{M}$ has only positive eigenvalues.
- $\mathbf{M}$ is called positive semi-definite if and only if $\langle \mathbf{M}\mathbf{x}, \mathbf{x}\rangle \geq 0$ for all $\mathbf{x} \in \mathbb{C}^L \setminus \{\mathbf{0}\}$. In other words, the eigenvalues of the matrix $\mathbf{M}$ are all non-negative.
- $\mathbf{M}$ is called negative-definite if and only if $\langle \mathbf{M}\mathbf{x}, \mathbf{x}\rangle < 0$ for all $\mathbf{x} \in \mathbb{C}^L \setminus \{\mathbf{0}\}$. In other words, the matrix $\mathbf{M}$ has only negative eigenvalues.
- $\mathbf{M}$ is called negative semi-definite if and only if $\langle \mathbf{M}\mathbf{x}, \mathbf{x}\rangle \leq 0$ for all $\mathbf{x} \in \mathbb{C}^L \setminus \{\mathbf{0}\}$. In other words, the eigenvalues of the matrix $\mathbf{M}$ are all non-positive.

Furthermore, we define the following notation for two Hermitian matrices $\mathbf{M}$ and $\mathbf{A}$. We write $\mathbf{A} > \mathbf{M}$ if and only if the matrix $\mathbf{A} - \mathbf{M}$ is positive-definite. The same notation is used analogously for the positive semi-definite, negative-definite and negative semi-definite cases.

Now continuing with the frame operator, we have that the frame operator $\mathbf{S} = \mathbf{H}^*\mathbf{H}$ is positive definite and self-adjoint, i.e. Hermitian. In fact, if the underlying set of vectors $\{\mathbf{h}_\ell\}_{\ell=0}^{N-1}$ forms a frame then the frame operator $\mathbf{S}$ is invertible. These facts are now proven in the following Theorem which can be found in [78].

Theorem 4.1.1 (Theorem 1.4 in [78])**.** *The frame operator* $\mathbf{S}$ *of a frame* $\{\mathbf{h}_\ell\}_{\ell=0}^{N-1}$ *with* $\mathbf{h}_\ell \in \mathbb{C}^L$ *for all* $\ell = \{0, \dots, N-1\}$*, with frame bounds* A *and* B *is a positive, self-adjoint invertible operator satisfying*

$$A \cdot \mathrm{I}_L \leq S \leq B \cdot \mathrm{I}_L \,,$$

where I_L *denotes the* L*-dimensional identity matrix.*

Proof. First, we show

$$\langle \mathbf{S}\mathbf{x}, \mathbf{x} \rangle = \|\mathbf{H}\mathbf{x}\|_2^2 = \sum_{k=0}^{N-1} |\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 \,.$$

From Definition 4.1.3 we have $\mathbf{H} : \mathbb{C}^L \to \mathbb{C}^N$ thus $\mathbf{H}\mathbf{x} \in \mathbb{C}^N$ and since we have $\mathbf{H}\mathbf{x} = \{\langle \mathbf{x}, \mathbf{h}_\ell \rangle\}_{\ell=0}^{N-1}$, we get the ℓ_2-norm of $\mathbf{H}\mathbf{x}$ by simply applying the definition and arrive at,

$$\|\mathbf{H}\mathbf{x}\|_2 = \sqrt{\sum_{k=0}^{N-1} |\langle \mathbf{x}, \mathbf{h}_k \rangle|^2} \,.$$

The left-hand side follows from the definition of the frame operator

$$\langle \mathbf{S}\mathbf{x}, \mathbf{x} \rangle = \langle \mathbf{H}^*\mathbf{H}\mathbf{x}, \mathbf{x} \rangle = \langle \mathbf{H}\mathbf{x}, \mathbf{H}\mathbf{x} \rangle = \|\mathbf{H}\mathbf{x}\|_2^2 \,.$$

Next observe that,

$$\langle A\mathbf{x}, \mathbf{x} \rangle = A \|\mathbf{x}\|_2^2 \leq \sum_{k=0}^{N-1} |\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 = \langle \mathbf{S}\mathbf{x}, \mathbf{x} \rangle \leq B \|\mathbf{x}\|_2^2 = \langle B\mathbf{x}, \mathbf{x} \rangle$$

holds for all $\mathbf{x} \in \mathbb{C}^L$. Note that by definition of the frame we have $0 < A, B < \infty$ and with the inequalities above we see that all eigenvalues of $\mathbf{S}$ must lie in the

interval $[A, B]$. Which implies the assertion of the theorem. □

We are going to need the following result later, for a proof see Propositions 1.9 and 1.10 in [78].

Proposition 4.1.2 (Proposition 1.10 in [78]). *Let $\{\mathbf{h}_\ell\}_{\ell=0}^{N-1}$ be a frame for $\mathbb{C}^L$ with frame operator $\mathbf{S}$, and let $\mathbf{M}$ be an invertible operator on $\mathbb{C}^L$. Then $\{\mathbf{M}\mathbf{h}_\ell\}_{\ell=0}^{N-1}$ is also a frame for $\mathbb{C}^L$ with frame operator $\mathbf{M}\mathbf{S}\mathbf{M}^*$.*

As already mentioned the frame operator contains crucial properties of the frame. One of which is that the largest and smallest eigenvalues of the frame operator coincide with the optimal lower and upper frame bound. This observation is formulated in the following theorem which can be found with its proof in [78].

Theorem 4.1.3 (Theorem 1.5 in [78]). *Let $\{\mathbf{h}_\ell\}_{\ell=0}^{N-1}$ be a frame for $\mathbb{C}^L$ with frame operator $\mathbf{S}$ having eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$. Then λ_1 coincides with the optimal upper frame bound and λ_L is the optimal lower frame bound.*

There is also an intimate connection between the vectors of the frame and the eigenvalues and eigenvectors of the corresponding frame operator. This will be stated in the following theorem and can be found with its proof in [78].

Theorem 4.1.4 (Theorem 1.6 in [78]). *Let $\{\mathbf{h}_\ell\}_{\ell=0}^{N-1}$ be a frame for $\mathbb{C}^L$ with frame operator $\mathbf{S}$ having ℓ_2-normalized eigenvectors $\{\mathbf{v}_k\}_{k=0}^{L-1}$ and respective eigenvalues $\{\lambda_k\}_{k=0}^{L-1}$. Then for all $k = 0, \ldots L-1$ we have*

$$\lambda_k = \sum_{j=0}^{L-1} |\langle \mathbf{v}_k, \mathbf{h}_j \rangle|^2 \ .$$

In particular,

$$\mathrm{Tr}\,(\mathbf{S}) = \sum_{j=0}^{L-1} \lambda_j = \sum_{j=0}^{L-1} \|\mathbf{h}_j\|_2^2 \ .$$

Next we discuss the class of tight frames which will be especially important since finite Gabor frames (which will be discussed later) are a class of tight frames. In its simplest form a tight frame can be characterized as those frames whose frame operator equals a positive multiple of the identity. Subsequently we give some alternative classifications.

Proposition 4.1.5. *Let $\{\mathbf{h}_\ell\}_{\ell=0}^{N-1}$ be a frame for $\mathbb{C}^L$ with frame operator $\mathbf{S}$. Then the following conditions are equivalent.*

(a) *$\{\mathbf{h}_\ell\}_{\ell=0}^{N-1}$ is a tight frame with frame bounds $A = B$ for $\mathbb{C}^L$.*

(b) $\mathbf{S} = A \cdot \mathrm{I}_L$.

(c) *For every* $\mathbf{x} \in \mathbb{C}^L$, *we have*

$$\mathbf{x} = \frac{1}{A} \sum_{k=0}^{N-1} \langle \mathbf{x}, \mathbf{h}_k \rangle \, \mathbf{h}_k \, .$$

(d) *For every* $\mathbf{x} \in \mathbb{C}^L$, *we have*

$$A \, \|\mathbf{x}\|_2^2 = \sum_{k=0}^{N-1} |\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 \, .$$

Proof. First note that the equivalence of (a) and (d) directly follows from the definition of the frame in Definition 4.1.1 and the definition of tight frames with $A = B$.
Next we show the equivalence of (a) and (b), i.e. (a) $\Leftrightarrow$ (b). As mentioned before we have

$$\langle \mathbf{S}\mathbf{x}, \mathbf{x} \rangle = \|\mathbf{H}\mathbf{x}\|_2^2 = \sum_{k=0}^{N-1} |\langle \mathbf{x}, \mathbf{h}_k \rangle|^2 = A \, \|\mathbf{x}\|_2^2 \, ,$$

where the last equivalence is the consequence of the frame being tight, i.e. property (d). Now simply observing

$$\langle \mathbf{S}\mathbf{x}, \mathbf{x} \rangle = A \, \|\mathbf{x}\|_2^2 = A \, \langle \mathbf{x}, \mathbf{x} \rangle = \langle A \cdot \mathrm{I}_L \mathbf{x}, \mathbf{x} \rangle$$

we see that $\mathbf{S} = A\mathrm{I}_L$ holds. For the final equivalence of (b) and (c) we observe

$$A\mathbf{x} = A \cdot \mathrm{I}_L \mathbf{x} = \mathbf{S}\mathbf{x} = \sum_{k=0}^{N-1} \langle \mathbf{x}, \mathbf{h}_k \rangle \, \mathbf{h}_k \, ,$$

Simply multiplying both sides with $1/A$ we obtain the claimed assertion. □

Next we discuss the Gramian operator (or Gramian matrix) of the frame. The Gramian is obtained by first applying the synthesis operator and then the analysis operator, i.e. $\mathbf{G} = \mathbf{H}\mathbf{H}^*$ with $\mathbf{G}$ being the Gramian operator. We start with its formal definition.

Definition 4.1.6. *Let* $\{\mathbf{h}_\ell\}_{\ell=0}^{N-1}$ *be a frame for* $\mathbb{C}^L$ *with analysis operator* $\mathbf{H}$. *Then the operator* $\mathbf{G} : \mathbb{C}^N \to \mathbb{C}^N$, *defined by*

$$\mathbf{G}\mathbf{y} = \mathbf{H}\mathbf{H}^*\mathbf{y} \, ,$$

for $\mathbf{y} \in \mathbb{C}^N$*, is called the Gramian operator of the frame* $\{\mathbf{h}_\ell\}_{\ell=0}^{N-1}$*.*

The matrix representation of the Gramian of a frame $\{\mathbf{h}_\ell\}_{\ell=0}^{N-1}$ is given as follows.

$$\mathbf{G} = \begin{bmatrix} \|\mathbf{h}_0\|^2 & |\langle \mathbf{h}_1, \mathbf{h}_0 \rangle| & |\langle \mathbf{h}_2, \mathbf{h}_0 \rangle| & \dots & \dots & |\langle \mathbf{h}_{N-1}, \mathbf{h}_0 \rangle| \\ |\langle \mathbf{h}_0, \mathbf{h}_1 \rangle| & \|\mathbf{h}_1\|^2 & |\langle \mathbf{h}_2, \mathbf{h}_1 \rangle| & \dots & \dots & |\langle \mathbf{h}_{N-1}, \mathbf{h}_1 \rangle| \\ |\langle \mathbf{h}_0, \mathbf{h}_2 \rangle| & |\langle \mathbf{h}_1, \mathbf{h}_2 \rangle| & \|\mathbf{h}_2\|^2 & \dots & \dots & |\langle \mathbf{h}_{N-1}, \mathbf{h}_2 \rangle| \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ |\langle \mathbf{h}_0, \mathbf{h}_{N-1} \rangle| & |\langle \mathbf{h}_1, \mathbf{h}_{N-1} \rangle| & |\langle \mathbf{h}_2, \mathbf{h}_{N-1} \rangle| & \dots & \dots & \|\mathbf{h}_{N-1}\|^2 \end{bmatrix}$$

The entries in the Gramian are the angles between the frame vectors. For instance if a frame is equiangular then all the off diagonal elements of the Gramian have the same value. In fact we already used the Gramian representation without mentioning in the proof of Theorem 2.4.2. There the columns of the measurement matrix can be regarded as a frame for $\mathbf{C}^L$ and we were concerned with a subset of the frame vectors. Then estimating the restricted isometry constant is equivalent to estimating the biggest eigenvalue of a symmetric sub-matrix of the Gramian of the frame. Note that a symmetric sub-matrix of the Gramian will also be positive semi-definite and Hermitian. We collect some of the results on the Gramian in the following proposition.

Proposition 4.1.6. *Let* $\{\mathbf{h}_k\}_{k=0}^{N-1}$ *be a frame for* $\mathbb{C}^L$ *with analysis operator* $\mathbf{H}$*, frame operator* $\mathbf{S}$ *and Gramian* $\mathbf{G}$*. Then the following statements hold.*

(a) *An operator* $\mathbf{U}$ *on* $\mathbb{C}^L$ *is unitary if and only if the Gramian of* $\{\mathbf{U}\mathbf{h}_k\}_{k=0}^{N-1}$ *is the same as the Gramian of* $\{\mathbf{h}_k\}_{k=0}^{N-1}$*.*

(b) *The non-zero eigenvalues of* $\mathbf{G}$ *and* $\mathbf{S}$ *are equal.*

(c) $\mathbf{G}$ *is invertible if and only if* $L = N$*.*

Proof. First, note that Property (c) is obvious from Property (b).
The property (a) is directly seen from the matrix representation of the Gramian, i.e. from the definition that the j,k-th entry of the Gramian is given by

$$\langle \mathbf{U}\mathbf{h}_k, \mathbf{U}\mathbf{h}_j \rangle = \langle \mathbf{h}_k, \mathbf{h}_j \rangle$$

for all $j, k = 0, \dots, N-1$.
The property (b) follows from the fact that $\mathbf{H}\mathbf{H}^*$ and $\mathbf{H}^*\mathbf{H}$ have the same non-zero eigenvalues, for an in-depth discussion see [66]. □

Signal recovery from frame coefficients

A discussion on frames wouldn't be complete without discussing methods to signal recovery from frame coefficients, i.e. from

$$\{\mathbf{y}[\ell]\}_{\ell=0}^{N-1} = \{\langle \mathbf{x}, \mathbf{h}_\ell \rangle\}_{\ell=0}^{N-1} \ .$$

Although we won't really need the contents of this section later in this work, we will discuss dual-frames and reconstruction algorithms for the sake of completeness.
In order to exactly recover a signal from its frame coefficients a so called dual frame for a frame is essential. We begin by giving the definition of the canonical dual frame.

Definition 4.1.7. *Let $\{\mathbf{h}_k\}_{k=0}^{N-1}$ be a frame for $\mathbb{C}^L$ with frame operator $\mathbf{S}$. Then $\{\mathbf{S}^{-1}\mathbf{h}_k\}_{k=0}^{N-1}$ is called the canonical dual frame for $\{\mathbf{h}_k\}_{k=0}^{N-1}$.*

Remark 4.1.8. *Note that since $\mathbf{S}$ is invertible, i.e. $\mathbf{S}^{-1}$ exists, it follows from Proposition 4.1.2 that $\{\mathbf{S}^{-1}\mathbf{h}_k\}_{k=0}^{N-1}$ is indeed a frame. Furthermore, from Theorem 4.1.3 and basic linear algebra we have that A^{-1} is the upper frame bound and B^{-1} is the lower frame bound of the canonical dual frame, $\{\mathbf{S}^{-1}\mathbf{h}_k\}_{k=0}^{N-1}$.*

Next we formulate an exact reconstruction formula using the canonical dual frame.

Theorem 4.1.7 (Theorem 1.8 in [78])**.** *Let $\{\mathbf{h}_k\}_{k=0}^{N-1}$ be a frame and denote its canonical dual frame by $\{\mathbf{S}^{-1}\mathbf{h}_k\}_{k=0}^{N-1}$. Then for every signal $\mathbf{x} \in \mathbb{C}^L$, we have*

$$\mathbf{x} = \sum_{k=0}^{N-1} \langle \mathbf{x}, \mathbf{h}_k \rangle \, \mathbf{S}^{-1}\mathbf{h}_k = \sum_{k=0}^{N-1} \left\langle \mathbf{x}, \mathbf{S}^{-1}\mathbf{h}_k \right\rangle \mathbf{h}_k \, .$$

We see in Theorem 4.1.7 that the frame $\{\mathbf{h}_k\}_{k=0}^{N-1}$ is the canonical dual frame to $\{\mathbf{S}^{-1}\mathbf{h}_k\}_{k=0}^{N-1}$ which is also apparent from the definition of the canonical dual frame.
Interestingly we do not necessarily need the canonical dual frame to exactly recover a signal from its frame coefficients, a so called dual frame is enough. We now give the definition for a dual frame.

Definition 4.1.9. *Let $\{\mathbf{h}_k\}_{k=0}^{N-1}$ be a frame for $\mathbb{C}^L$. Then a frame $\{\widetilde{\mathbf{h}}_k\}_{k=0}^{N-1}$ is called a dual frame for $\{\mathbf{h}_k\}_{k=0}^{N-1}$ if the following holds*

$$\mathbf{x} = \sum_{k=0}^{N-1} \langle \mathbf{x}, \mathbf{h}_k \rangle \, \widetilde{\mathbf{h}}_k \, ,$$

for all $\mathbf{x} \in \mathbb{C}^L$.

Now an interesting question is what is the distinguishing feature of the canonical dual frame from the alternate dual frames apart from that the canonical dual frame has an explicit formulation depending on the initial frame. Indeed, the sequence generated by the canonical dual, i.e. $\{\langle \mathbf{x}, \mathbf{S}^{-1}\mathbf{h}_\ell \rangle\}_{\ell=0}^{N-1}$ has the minimal ℓ_2-norm among all sequences that represent the signal $\mathbf{x}$ using the frame elements $\{\mathbf{h}_k\}_{k=0}^{N-1}$ as atoms. The following result can be found as Proposition 1.16 and Corollary 1.8 in [78].

Proposition 4.1.8. *Let* $\{\mathbf{h}_k\}_{k=0}^{N-1}$ *be a frame for* $\mathbb{C}^L$*, and let* $\{\widetilde{\mathbf{h}}_k\}_{k=0}^{N-1}$ *be an associated dual frame. Then, for all* $\mathbf{x} \in \mathbb{C}^L$ *we have*

$$\left\| \left\{ \left\langle \mathbf{x}, \mathbf{S}^{-1}\mathbf{h}_k \right\rangle \right\}_{k=0}^{N-1} \right\|_2 \leq \left\| \left\{ \left\langle \mathbf{x}, \widetilde{\mathbf{h}}_k \right\rangle \right\}_{k=0}^{N-1} \right\|_2 .$$

Proof. We begin by fixing an arbitrary $\mathbf{x} \in \mathbb{C}^L \setminus \{0\}$ and assume a sequence $\{a_i\}_{i=0}^{N-1}$ such that

$$\mathbf{x} = \sum_{i=0}^{N-1} a_i \mathbf{h}_i .$$

Our aim is to prove that for all sequences $\{a_i\}_{i=0}^{N-1}$ we have

$$\left\| \left\{ \left\langle \mathbf{x}, \mathbf{S}^{-1}\mathbf{h}_k \right\rangle \right\}_{k=0}^{N-1} \right\|_2 \leq \left\| \{a_i\}_{i=0}^{N-1} \right\|_2 .$$

First we denote by $\mathbf{H}$ the analysis operator of the frame $\{\mathbf{h}_k\}_{k=0}^{N-1}$ and observe

$$\left\{ \left\langle \mathbf{x}, \mathbf{S}^{-1}\mathbf{h}_k \right\rangle \right\}_{k=0}^{N-1} = \left\{ \left\langle \mathbf{S}^{-1}\mathbf{x}, \mathbf{h}_k \right\rangle \right\}_{k=0}^{N-1} \in \operatorname{imag}(\mathbf{H}) ,$$

where $\operatorname{imag}(\mathbf{H})$ denotes the image of the analysis operator $\mathbf{H}$. Since we have

$$0 = \sum_{i=0}^{N-1} \left(a_i - \left\langle \mathbf{x}, \mathbf{S}^{-1}\mathbf{h}_k \right\rangle \right) \mathbf{h}_i ,$$

follows

$$\left\{ a_k - \left\langle \mathbf{x}, \mathbf{S}^{-1}\mathbf{h}_k \right\rangle \right\}_{k=0}^{N-1} \in \ker(\mathbf{H}^*) .$$

Furthermore, we have that $\ker(\mathbf{H}^*)$ is orthogonal to $\operatorname{imag}(\mathbf{H})$. Now considering

$$\{a_i\}_{i=0}^{N-1} = \left\{ \left\langle \mathbf{x}, \mathbf{S}^{-1}\mathbf{h}_k \right\rangle \right\}_{k=0}^{N-1} + \left\{ a_k - \left\langle \mathbf{x}, \mathbf{S}^{-1}\mathbf{h}_k \right\rangle \right\}_{k=0}^{N-1} ,$$

and considering $\ker(\mathbf{H}^*) = \operatorname{imag}(\mathbf{H})^\perp$ we get

$$\sum_{i=0}^{N-1} |a_i|^2 = \sum_{i=0}^{N-1} \left|\left\langle \mathbf{x}, \mathbf{S}^{-1}\mathbf{h}_k \right\rangle\right|^2 + \sum_{i=0}^{N-1} \left|a_k - \left\langle \mathbf{x}, \mathbf{S}^{-1}\mathbf{h}_k \right\rangle\right|^2 .$$

Since $\sum_{i=0}^{N-1} |a_k - \langle \mathbf{x}, \mathbf{S}^{-1}\mathbf{h}_k\rangle|^2 \geq 0$ we see that the ℓ_2-norm of any sequence including one generated from a dual frame, i.e. $\left\{\left\langle \mathbf{x}, \widetilde{\mathbf{h}}_k \right\rangle\right\}_{k=0}^{N-1}$, is lower bounded by the sequence generated using the canonical dual frame. □

Note that we are considering finite frames hence the objects we are studying are all finite dimensional. It is also possible to give a linear algebraic proof for Proposition 4.1.8. The main idea there would be to consider the ℓ_2-minimization problem

$$\underset{\mathbf{z}\in\mathbb{C}^N}{\text{minimize}} \|\mathbf{z}\|_2 \quad \text{subject to } \mathbf{Hz} = \mathbf{x} .$$

The optimal solution for the ℓ_2 minimization is given by the pseudo-inverse of $\mathbf{H}$ [37, 80]. Then considering the sequence $\{\langle \mathbf{x}, \mathbf{S}^{-1}\mathbf{h}_k\rangle\}_{k=0}^{N-1}$ as an N-dimensional vector and looking at the definition of the pseudo-inverse we would see that the vector representation of the sequence defined by the canonical dual frame would be the solution to the ℓ_2-minimization problem.

So far we have discussed the exact signal recovery, however computing the (canonical) dual frame to a given frame is computationally expensive. Furthermore, the inversion formulas are not necessarily robust against corrupted measurements. Therefore, we are going to discuss the Chebyshev algorithm introduced in [81].

Theorem 4.1.9 (Theorem 1 in [81])**.** *Let $\{\mathbf{h}_k\}_{k=0}^{N-1}$ be a frame for $\mathbb{C}^L$ with frame bounds A, B and frame operator $\mathbf{S}$, and set*

$$\rho = \frac{B-A}{B+A} \quad \textit{and} \quad \sigma = \frac{\sqrt{B}-\sqrt{A}}{\sqrt{B}+\sqrt{A}} .$$

A signal $\mathbf{x} \in \mathbb{C}^L$ can be approximated and reconstructed from the coefficients $\{\langle \mathbf{x}, \mathbf{h}_i\rangle\}_{i=0}^{N-1}$ by the following recursion formula:
Define a sequence $\{\mathbf{z}\}_{j=0}^{\infty}$ in $\mathbb{C}^L$ and corresponding scalars $\{a_j\}_{j=1}^{\infty}$ by

$$\mathbf{z}_0 = 0 , \quad \mathbf{z}_1 = \frac{2}{B+A}\mathbf{S}\mathbf{x} , \quad \textit{and} \quad a_1 = 2 ,$$

and for $j > 2$ *set*

$$a_j = \frac{1}{1 - \frac{\rho^2}{4} a_{j-1}} \quad \textit{and} \quad \mathbf{z}_j = a_j \left(\mathbf{z}_{j-1} - \mathbf{z}_{j-2} + \frac{2}{A+B} \mathbf{S} \left(\mathbf{x} - \mathbf{z}_{j-1} \right) \right) + \mathbf{z}_{j-2} \, .$$

Then $\{\mathbf{z}\}_{j=0}^{\infty}$ *converges to* $\mathbf{x} \in \mathbb{C}^L$*, and the rate of convergence is*

$$\|\mathbf{x} - \mathbf{z}_j\| \leq \frac{2\sigma^j}{1 + \sigma^{2j}} \|\mathbf{x}\| \, .$$

Note that the algorithm indeed depends on the frame coefficients and the frame itself since we have from Definition 4.1.5 $\mathbf{S}\mathbf{x} = \sum_{i=0}^{N-1} \langle \mathbf{x}, \mathbf{h}_i \rangle \, \mathbf{h}_i$.
It should be mentioned that in [81] another method - Conjugate Gradient Method - is also introduced, this one does not need to know the frame bounds. Nevertheless, the convergence rate still depends on the frame bounds.

In the next subsection we continue with our original discussion on finite Gabor frames and systems.

4.2. Finite Gabor frames and systems

Finite Gabor frames have been extensively studied in the literature and there are already some good overview articles [19, 82, 83], the linear independence properties of vectors of finite Gabor frames were studied in [84, 85]. Finite Gabor frames have also been analysed in the context of sparse signal recovery in [71, 86–89]. Furthermore, it is well known that under a suitable selection of a window vector finite Gabor frames deliver finite frames with low coherence values [29, 90, 91]. Thus, finite Gabor frames have established themself as time-frequency analysis tools in digital signal processing applications [19, 92, 93].
In this section we will present some basic results on finite Gabor frames which will be relevant in later sections.
We begin this section by some conventional notation.
For the L-element cyclic group we write $\mathbb{Z}_L = \mathbb{Z}/L\mathbb{Z} = \{0, \ldots, L-1\}$. We index the entries of a vector in $\mathbb{C}^L$ by elements of $\mathbb{Z}_L$, i.e. $\mathbf{g} = (\mathbf{g}[0], \ldots, \mathbf{g}[L-1])^T$ where $\mathbf{g}^T$ denotes the transpose of a vector. Note that arithmetic operators on $\mathbb{Z}_L$ are computed $\mod L$ and that if L is a prime number then $\mathbb{Z}_L$ is equivalent to the L element finite field $\mathbb{F}_L$ [94]. Analysing finite Gabor frames will use the discrete Fourier transform which is defined for an L-dimensional vector $\mathbf{g} \in \mathbb{C}^L$

by

$$(\mathbf{Fg})[k] = \hat{\mathbf{g}}[k] = \sum_{n=0}^{L-1} \mathbf{g}[n]\, \mathrm{e}^{-\frac{j2\pi}{L}kn} \quad \text{for } k = 0, 1, \ldots, L-1\,,$$

with $j^2 = -1$. In order to give the matrix representation of the discrete Fourier transform in $\mathbb{C}^L$ [95], we denote by ω the L-th root of unity, i.e. $\omega = \mathrm{e}^{\frac{j2\pi}{L}}$. Then the matrix representation of the discrete Fourier transform is given by $\mathbf{F}[k,\ell] = \omega^{-k\cdot\ell}$ for $\ell, k = 0, 1, \ldots, L-1$ where $\mathbf{F}[k,\ell]$ denotes the entry of $\mathbf{F}$ in k-th row and ℓ-th column, i.e. the full matrix can be written as

$$\mathbf{F} = \begin{bmatrix} \omega^{(-1)\cdot 0\cdot 0} & \omega^{(-1)\cdot 1\cdot 0} & \omega^{(-1)\cdot 2\cdot 0} & \ldots & \ldots & \omega^{(-1)\cdot (L-1)\cdot 0} \\ \omega^{(-1)\cdot 0\cdot 1} & \omega^{(-1)\cdot 1\cdot 1} & \omega^{(-1)\cdot 2\cdot 1} & \ldots & \ldots & \omega^{(-1)\cdot (L-1)\cdot 1} \\ \omega^{(-1)\cdot 0\cdot 3} & \omega^{(-1)\cdot 1\cdot 3} & \omega^{(-1)\cdot 2\cdot 3} & \ldots & \ldots & \omega^{(-1)\cdot (L-1)\cdot 3} \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ \omega^{(-1)\cdot 0\cdot (L-1)} & \omega^{(-1)\cdot 1\cdot (L-1)} & \omega^{(-1)\cdot 2\cdot (L-1)} & \ldots & \ldots & \omega^{(-1)\cdot (L-1)\cdot (L-1)} \end{bmatrix}.$$

4.2.1. Time-frequency shifts and their properties

To define finite Gabor frames we first need to define the time-frequency shift operators. To this end, we define the translation and modulation operators. The translation operator $\mathbf{T} : \mathbb{C}^L \to \mathbb{C}^L$ is given by

$$(\mathbf{Tg})[k] = \mathbf{g}[k-1]\,, \quad \text{for all } k \in \mathbb{Z}_L\,.$$

It acts as a cyclic shift on a vector in $\mathbb{C}^L$. Using concatenations of the translation operators we can represent the τ-times cyclic shift of a vector $\mathbf{g} \in \mathbb{C}^L$ by

$$(\mathbf{T}^\tau \mathbf{g})[k] = \mathbf{g}[k-\tau]\,, \quad \text{for all } k \in \mathbb{Z}_L\,.$$

The matrix representation of the translation operator $\mathbf{T}$ using the identity matrix is given as

$$\mathbf{T} = \begin{bmatrix} 0 & 1 \\ \mathrm{I}_{L-1} & 0 \end{bmatrix},$$

I_{L-1} is the $L-1 \times L-1$ identity matrix and 0 corresponds to zero entries. In a similar fashion we can write $\mathbf{T}^\tau$ as

$$\mathbf{T}^\tau = \begin{bmatrix} 0 & \mathrm{I}_\tau \\ \mathrm{I}_{L-\tau} & 0 \end{bmatrix} .$$

Furthermore, it is noteworthy to observe that we have $(\mathbf{T}^\tau)^{-1} = (\mathbf{T}^\tau)^* = \mathbf{T}^{-\tau}$.

Analogous to the translation operator we define the modulation operator $\mathbf{M} : \mathbb{C}^L \to \mathbb{C}^L$ as

$$(\mathbf{M}\mathbf{g})[k] = \mathbf{g}[k] \cdot \omega^k , \quad \text{for all } k \in \mathbb{Z}_L ,$$

where $\omega = \mathrm{e}^{j\frac{2\pi}{L}}$. Again analogous to the translation operator concatenating the modulation operator ν times (with $\nu \in \{0, 1, \dots L-1\}$) and applying on a vector $\mathbf{g} \in \mathbb{C}^L$ we have

$$(\mathbf{M}^\nu \mathbf{g})[k] = \mathbf{g}[k]\, \omega^{\nu k} , \quad \text{for all } k \in \mathbb{Z}_L .$$

The modulation operator has the following matrix representation

$$\mathbf{M} = \mathrm{diag}\left(\left[\omega^0 , \omega^1 , \dots , \omega^{L-1} \right]^T \right) ,$$

where $\mathrm{diag}(\mathbf{g})$ with $\mathbf{g} \in \mathbb{C}^L$ denotes diagonal matrix with the elements of $\mathbf{g}$ as diagonal entries. In explicit matrix for $\mathbf{M}$ is written as

$$\mathbf{M} = \begin{bmatrix} \omega^0 & & & & & \\ & \omega^1 & & & & \\ & & \omega^2 & & & \\ & & & \ddots & & \\ & & & & \omega^{L-1} \end{bmatrix} .$$

Accordingly the matrix form of $\mathbf{M}^\nu$, $\nu \in \{0, 1, \dots, L-1\}$, i.e. ν-times concatenation of the modulation operator, has the following matrix form

$$\mathbf{M}^\nu = \mathrm{diag}\left(\left[\omega^{\nu\cdot 0} , \omega^{\nu\cdot 1} , \dots , \omega^{\nu\cdot(L-1)} \right]^T \right) .$$

Similar to the translation operator we have $(\mathbf{M}^\nu)^{-1} = (\mathbf{M}^\nu)^* = \mathbf{M}^{-\nu}$.

Let $\mathbf{f}_k$ be the k-th column of the discrete Fourier matrix, i.e. $\mathbf{F} = [\mathbf{f}_0, \mathbf{f}_1, \dots, \mathbf{f}_{L-1}]$,

then the modulation operator can also be denoted by $\mathbf{M}^\nu = \operatorname{diag}\left(\overline{\mathbf{f}_\nu}\right)$, where $\overline{\mathbf{f}_\nu}$ denotes the complex conjugate of the vector $\mathbf{f}_\nu$. The modulation operator is sometimes refereed as frequency shift, because applying the modulation operator onto a signal is the same as shifting the signal in the frequency, i.e. cyclic shift of the Fourier transform of the signal. This can be seen by straight forward computation,

$$\begin{aligned}\widehat{(\mathbf{M}^\nu \mathbf{g})}[k] &= (\mathbf{F}\mathbf{M}^\nu \mathbf{g})[k] \\ &= \sum_{n=0}^{L-1} (\omega^{\nu n}\mathbf{g}[n])\, \mathrm{e}^{-j\frac{2\pi}{L}kn} = \sum_{n=0}^{L-1} \mathbf{g}[n]\, \mathrm{e}^{-j\frac{2\pi}{L}(k-\nu)n} \\ &= \hat{\mathbf{g}}[k-\nu] \quad \text{for } k \in \mathbb{Z}_L\,,\end{aligned} \tag{4.2}$$

where we used $\omega = \mathrm{e}^{j\frac{2\pi}{L}}$. From (4.2), we immediately get the following relation

$$\mathbf{F}\mathbf{M}^k = \mathbf{T}^k\mathbf{F} \quad \text{for } k \in \mathbb{Z}_L\,. \tag{4.3}$$

Furthermore, a simple reformulation of (4.3) delivers us

$$\mathbf{F}\mathbf{M}^k\mathbf{F}^* = \mathbf{T}^k\,,$$

and we see that we can write the eigendecomposition of $\mathbf{T}^k$ in terms of the discrete Fourier transform matrix and frequency-shifts which are diagonal matrices. Indeed, this is not surprising another possibility to see it is to notice that $\mathbf{T}^k$ with $k \in \mathbb{Z}_L$ are circulant matrices and that the columns of the Fourier matrix are eigenvectors of circulant matrices [96].
Analogous to (4.3) we can also derive the following relation

$$\mathbf{F}^*\mathbf{M}^{-k}\mathbf{F} = \mathbf{T}^k \quad \text{for } k \in \mathbb{Z}_L\,.$$

The following lemma describing the commutation relation between the translation and modulation operator plays a key role in the analysis of Gabor frames.

Lemma 4.2.1. *Set $\omega = \mathrm{e}^{j\frac{2\pi}{L}}$ then for the translation and modulation operators in $\mathbb{C}^L$ the following commutation relation*

$$\mathbf{T}^\tau\mathbf{M}^\nu = \omega^{-\nu\tau}\mathbf{M}^\nu\mathbf{T}^\tau\,,$$

holds.

Proof. Let $\mathbf{g} \in \mathbb{C}^L$ then for any $(\tau, \nu) \in \mathbb{Z}_L \times \mathbb{Z}_L$ we obtain by straight forward

computation

$$\begin{aligned}\mathbf{T}^{\tau}\mathbf{M}^{\nu}\mathbf{g}\left[k\right] &= \left(\mathbf{M}\mathbf{g}\right)\left[k-\tau\right] \\ &= \omega^{\nu(k-\tau)}\mathbf{g}\left[k-\tau\right] \\ &= \omega^{-\nu\tau}\omega^{\nu k}\mathbf{g}\left[k-\tau\right] \\ &= \omega^{-\nu\tau}\mathbf{M}^{\nu}\mathbf{T}^{\tau}\mathbf{g}\left[k\right] ,\end{aligned}$$

for $k = \{0, 1, \dots L-1\}$. □

Next, following the notation in [19, 82] we define the time-frequency shift operator by

$$\pi\left(\tau,\nu\right) = \mathbf{M}^{\nu}\mathbf{T}^{\tau} \quad \text{for } \tau,\nu \in \mathbb{Z}_L.$$

The set of time-frequency shifts have the interesting property that they form an orthogonal basis for the space of linear operators on $\mathbb{C}^L$ under the Hilbert-Schmidt inner product, we denote this space of all $L \times L$ square matrices by $\mathcal{HS}^L$. This is a Hilbert space with the Hilbert-Schmidt inner product.
The Hilbert-Schmidt inner product is given by

$$\left\langle \mathbf{A},\mathbf{B}\right\rangle_{\mathrm{HS}} = \mathrm{Tr}\left(B^{*}A\right) = \sum_{n=0}^{L-1}\sum_{m=0}^{L-1}\left\langle \mathbf{A}\tilde{\mathbf{e}}_n,\tilde{\mathbf{e}}_m\right\rangle\overline{\left\langle \mathbf{B}\tilde{\mathbf{e}}_n,\tilde{\mathbf{e}}_m\right\rangle},$$

where $\mathrm{Tr}\left(\cdot\right)$ denotes the trace operator and $\tilde{\mathbf{e}}_n \in \mathbb{C}^L$ for $n \in \{0, 1, \dots, L-1\}$ denotes an orthogonal basis.
For later use let us define the canonical basis on $\mathbb{C}^L$ by

$$\mathbf{e}_n\left[m\right] = \begin{cases} 1 & \text{if } m = n, \\ 0 & \text{if } m \neq n\,. \end{cases}$$

We will denote the norm induced by the Hilbert-Schmidt inner product by $\left\|\cdot\right\|_{\mathrm{HS}}$. In order to prove that the set of all normalized time-frequency shifts $\left\{\frac{1}{\sqrt{L}}\pi\left(\tau,\nu\right)\right\}_{\tau,\nu\in\mathbb{Z}_L}$ form an orthogonal basis under the Hilbert-Schmidt inner product we need the following well known result.

Lemma 4.2.2. *For any $L \in \mathbb{N}$ we have*

$$\sum_{k=0}^{L-1}\mathrm{e}^{j\frac{2\pi}{L}Nk} = \begin{cases} L, & \textit{if } L \textit{ divides } N, \\ 0, & \textit{otherwise.} \end{cases}$$

In particular, we have

$$\sum_{k=0}^{L-1} \mathrm{e}^{j\frac{2\pi}{L}Nk} = \begin{cases} L, & \text{for } N = 0, \\ 0, & \text{for } N = 1, \dots, L-1. \end{cases}$$

Proof. Set $x = \mathrm{e}^{j\frac{2\pi}{L}N}$.
If L divides N then we have $\frac{N}{L} \in \mathbb{N}$, hence $x = 1$.
In case L does not divide N we can then write the sum as

$$\sum_{k=0}^{L-1} x^k = \frac{\left(1 + x + x^2 + \cdots + x^{L-1}\right)(1-x)}{1-x} = \frac{1-x^L}{1-x} = \frac{1-\mathrm{e}^{j2\pi N}}{1-x} = 0\,.$$

□

Now we can prove the following theorem which can also be found in [19, 82, 84, 86].

Theorem 4.2.3. *The set of all normalized time-frequency shift matrices* $\left\{\frac{1}{\sqrt{L}}\pi(\tau,\nu)\right\}_{\tau,\nu\in\mathbb{Z}_L}$ *forms an orthonormal basis for* $\mathcal{HS}^L$.

Proof. We need to show two properties. First, that the normalized time-frequency shifts are indeed orthogonal under the Hilbert-Schmidt inner product and secondly that they form a basis for square matrices in $\mathbb{C}^L$.
We verify first hand

$$\mathrm{Tr}\left(\mathbf{M}^\nu \mathbf{T}^\tau\right) = 0 \quad \text{for } (\tau,\nu) \neq (0,0)\ .$$

The cases of $(0,\nu)$ for $\nu \neq 0$ follows immediately from Lemma 4.2.2. Indeed,

$$\begin{aligned} \left\langle \frac{1}{\sqrt{L}}\pi(0,\nu), \frac{1}{\sqrt{L}}\pi(0,\mu) \right\rangle_{\mathrm{HS}} &= \frac{1}{L}\mathrm{Tr}\left(\mathbf{M}^{\nu-\mu}\right) = \frac{1}{L}\sum_{\ell=0}^{L-1}\left\langle \mathbf{M}^{\nu-\mu}\mathbf{e}_\ell, \mathbf{e}_\ell \right\rangle \\ &= \frac{1}{L}\sum_{\ell=0}^{L-1} \mathrm{e}^{j\frac{2\pi}{L}(\nu-\mu)\ell} = \begin{cases} 1, & \text{if } \nu = \mu, \\ 0, & \text{if } \nu \neq \mu. \end{cases} \end{aligned}$$

Next observe

$$\begin{aligned} \left\langle \frac{1}{\sqrt{L}}\pi(\tau,\nu), \frac{1}{\sqrt{L}}\pi(k,\ell) \right\rangle_{\mathrm{HS}} &= \frac{1}{L}\mathrm{Tr}\left(\pi(k,\ell)^*\,\pi(\tau,\nu)\right) \\ &= \frac{1}{L}\mathrm{Tr}\left(\mathbf{T}^{-k}\mathbf{M}^{\nu-\ell}\mathbf{T}^\tau\right) \end{aligned}$$

$$= \omega^{k(\nu-\ell)} \frac{1}{L} \mathrm{Tr}\left(\mathbf{M}^{\nu-\ell}\mathbf{T}^{\tau-k}\right)$$
$$= \begin{cases} 1, & \text{if } (\tau,\nu) = (k,\ell)\,, \\ 0, & \text{if } (\tau,\nu) \neq (k,\ell)\,. \end{cases}$$

Hence, we have shown that each mutually distinct time-frequency shifts are orthogonal to each other under the Hilbert-Schmidt inner product.
Next, we show that the set of all time-frequency shifts form a basis for square matrices on $\mathbb{C}^L$. We define the matrix $\delta_{n,m} \in \mathbb{C}^{L\times L}$ as

$$\langle \delta_{n,m}\mathbf{e}_c, \mathbf{e}_r\rangle = \begin{cases} 1, & \text{for } (r,c) = (n,m)\,, \\ 0, & \text{else.} \end{cases}$$

Note that $\delta_{n,m}$ is a matrix with a 1 in the entry corresponding to the n-th row and m-th column and zeros every where else. Obviously, the set of matrices $\{\delta_{n,m}\}_{n,m=0}^{L-1}$ forms a basis for the square matrices on $\mathbb{C}^L$. Our aim is to show that we can write all matrices in the set $\{\delta_{n,m}\}_{n,m=0}^{L-1}$ as linear combinations of time-frequency shift matrices. First, observe that from Lemma 4.2.2 we have

$$\delta_{0,0} = \frac{1}{L}\sum_{k=0}^{L-1}\mathbf{M}^k = \sum_{k=0}^{L-1}\frac{1}{\sqrt{L}}\cdot\frac{1}{\sqrt{L}}\pi(0,k)\,.$$

Now noticing that

$$\delta_{n,m} = \mathbf{T}^n\delta_{0,0}\mathbf{T}^{-m} = \frac{1}{L}\sum_{k=0}^{L-1}\omega^{-nk}\mathbf{M}^k\mathbf{T}^{n-m} = \sum_{k=0}^{L-1}\frac{\omega^{-nk}}{\sqrt{L}}\cdot\frac{1}{\sqrt{L}}\pi(n-m,k)$$

holds, concludes the proof. □

4.2.2. Finite Gabor systems

Now we continue with defining finite Gabor frames and systems.
The finite Gabor frame with respect to a window vector $\mathbf{g} \in \mathbb{C}^L$ is defined as the collection of all time-frequency shifts of $\mathbf{g}$, i.e. the set of vectors

$$\{\pi(\tau,\nu)\,\mathbf{g}\}_{\tau,\nu\in\mathbb{Z}_L\times\mathbb{Z}_L}\,,$$

and a finite Gabor system refers to a subset of the full Gabor frame indexed by $\Lambda \subset \mathbb{Z}_L\times\mathbb{Z}_L$, formally the collection $\{\pi(\tau,\nu)\,\mathbf{g}\}_{\tau,\nu\in\Lambda}$. Moreover, if $\Lambda = \mathbb{Z}_L\times\mathbb{Z}_L$ we refer to the Gabor system as the full Gabor system. We denote by $\mathbf{G}_\mathbf{g}$ the

$L \times L^2$ matrix whose columns consists of elements of the full Gabor system, i.e.

$$\mathbf{G_g} = \left[\pi(0,0)\,\mathbf{g} \,\middle|\, \pi(0,1)\,\mathbf{g} \,\middle|\, \ldots \,\middle|\, \pi(1,0)\,\mathbf{g} \,\middle|\, \ldots \,\middle|\, \pi(L-1,L-1)\,\mathbf{g}\right].$$

We will also refer to $\mathbf{G_g}$ as the Gabor matrix generated by the window $\mathbf{g}$. The matrix $\mathbf{G_g}$ can also be written using the discrete Fourier transform matrix. To this end we define the diagonal matrices

$$\mathbf{D}_k = \operatorname{diag}\left(\mathbf{T}^k\mathbf{g}\right),$$

with $k = \{0, 1, \ldots L-1\}$ and where $\operatorname{diag}(\cdot)$ generates a diagonal matrix with entries from its argument. Now using the matrices $\mathbf{D}_k$ and the Fourier matrix we can write the matrix $\mathbf{G_g}$ as

$$\mathbf{G_g} = \left[\mathbf{D}_0\mathbf{F} \,\middle|\, \mathbf{D}_1\mathbf{F} \,\middle|\, \ldots \,\middle|\, \mathbf{D}_{L-1}\mathbf{F}\right].$$

Additionally, note that the matrix $\mathbf{G_g}$ is the matrix representation of the synthesis operator corresponding to the full Gabor frame generated by the window $\mathbf{g}$ and that $\mathbf{G_g^*}$ corresponds to the analysis operator.
An interesting property of the full Gabor frame is that it always forms a tight frame for any $\mathbf{g} \in \mathbb{C}^L \setminus \{0\}$. This property is proven in the following proposition which follows the same proof as in [84].

Proposition 4.2.4 (Proposition 2 in [84])**.** *For any $\mathbf{g} \neq 0$, the full Gabor system $\{\pi(\tau,\nu)\,\mathbf{g}\}_{\tau,\nu\in\mathbb{Z}_L\times\mathbb{Z}_L}$ is a tight finite frame for $\mathbb{C}^L$ with frame bounds $A = B = L^2\,\|\mathbf{g}\|_2^2$.*

Proof. We start by considering the synthesis operator of the Gabor frame by $\mathbf{H}^* = \mathbf{G_g}$. Then the rows of the analysis operator $\mathbf{H}$ are complex conjugates of the frame elements $\{\pi(\tau,\nu)\,\mathbf{g}\}_{\tau,\nu\in\mathbb{Z}_L\times\mathbb{Z}_L}$. Using the notation with the discrete Fourier matrix we can write the analysis operator as

$$\mathbf{H} = \begin{bmatrix} \mathbf{F}^*\mathbf{D}_0^* \\ \mathbf{F}^*\mathbf{D}_1^* \\ \mathbf{F}^*\mathbf{D}_2^* \\ \vdots \\ \mathbf{F}^*\mathbf{D}_{L-1}^* \end{bmatrix}.$$

Now using the definition of the frame operator $\mathbf{S}$ from Definition 4.1.5 we get

$$\begin{aligned}\mathbf{S} = \mathbf{H}^*\mathbf{H} &= \left[\mathbf{D}_0\mathbf{F} \,\middle|\, \mathbf{D}_1\mathbf{F} \,\middle|\, \dots \,\middle|\, \mathbf{D}_{L-1}\mathbf{F}\right] \cdot \begin{bmatrix} \mathbf{F}^*\mathbf{D}_0^* \\ \mathbf{F}^*\mathbf{D}_1^* \\ \vdots \\ \mathbf{F}^*\mathbf{D}_{L-1}^* \end{bmatrix} \\ &= \mathbf{D}_0\mathbf{F}\mathbf{F}^*\mathbf{D}_0^* + \mathbf{D}_1\mathbf{F}\mathbf{F}^*\mathbf{D}_1^* + \cdots + \mathbf{D}_{L-1}\mathbf{F}\mathbf{F}^*\mathbf{D}_{L-1}^* \\ &= L\cdot\mathbf{D}_0\mathrm{I}_L\mathbf{D}_0^* + L\cdot\mathbf{D}_1\mathrm{I}_L\mathbf{D}_1^* + \cdots + L\cdot\mathbf{D}_{L-1}\mathrm{I}_L\mathbf{D}_{L-1}^* \\ &= L\left(\mathbf{D}_0\mathbf{D}_0^* + \mathbf{D}_1\mathbf{D}_1^* + \cdots + \mathbf{D}_{L-1}\mathbf{D}_{L-1}^*\right) \\ &= \left(L^2\sum_{k=0}^{L-1}|\mathbf{g}[k]|^2\right)\mathrm{I}_L = L^2\|\mathbf{g}\|_2^2\,\mathrm{I}_L\,.\end{aligned}$$

Hence we see that the frame operator $\mathbf{S}$ is a multiple of identity, and therefore the full Gabor frame is a tight frame as claimed. □

Since the full finite Gabor frame is a tight frame the canonical dual frame of the full Gabor frame is again a Gabor frame.

Subsets of the full Gabor system, i.e. Gabor systems are generally not a frame for $\mathbb{C}^L$, but they may form a frame in some cases. Later we will see that for any L there exists $\mathbf{g}\in\mathbb{C}^L$ such that for any subset $\Lambda\subset\mathbb{Z}_L\times\mathbb{Z}_L$ with $|\Lambda|\geq L$ the collection $\{\pi(\lambda)\,\mathbf{g}\}_{\lambda\in\Lambda}$ forms a finite frame.
However, the collection $\{\pi(\lambda)\,\mathbf{g}\}_{\lambda\in\Lambda}$ does not necessarily need to end up forming a finite frame. For instance choosing $\mathbf{g}=\mathbf{e}_0$ and setting $\Lambda = \{(0,0),(0,1),\dots,(0,L-1)\}$ it is easy to see that the dimension of the span of the set $\{\pi(\lambda)\,\mathbf{e}_0\}_{\lambda\in\Lambda}$ will be one.
If however the collection $\{\pi(\lambda)\,\mathbf{g}\}_{\lambda\in\Lambda}$ forms a frame then for certain subsets $\Lambda\subset\mathbb{Z}_L\times\mathbb{Z}_L$ the Gabor frame has the remarkable property that its corresponding canonical dual frame is again a Gabor frame on the same set Λ. The condition for the set Λ is that $\Lambda\subset\mathbb{Z}_L\times\mathbb{Z}_L$ needs to be a subgroup of $\mathbb{Z}_L\times\mathbb{Z}_L$. For the frame operator of the finite Gabor frame $\{\pi(\lambda)\,\mathbf{g}\}_{\lambda\in\Lambda}$ where Λ is a subgroup of $\mathbb{Z}_L\times\mathbb{Z}_L$, we have the following result from [82].

Proposition 4.2.5 (Lemma 4.12 in [82])**.** *Let $\{\pi(\lambda)\,\mathbf{g}\}_{\lambda\in\Lambda}$ be a finite Gabor frame for $\mathbb{C}^L$ and denote by $\mathbf{S}$ the corresponding frame operator and additionally, if the set $\Lambda\subset\mathbb{Z}_L\times\mathbb{Z}_L$ is a subgroup of $\mathbb{Z}_L\times\mathbb{Z}_L$. Then we have the commutation relation*

$$\pi(\lambda)\,\mathbf{S} = \mathbf{S}\pi(\lambda)\ ,$$

for all $\lambda\in\Lambda$.

For the proof of Proposition 4.2.5 we need the following lemma.

Lemma 4.2.6. *Let $(k,l),(\tau,\nu) \in \mathbb{Z}_L \times \mathbb{Z}_L$ and set $\omega = \mathrm{e}^{j\frac{2\pi}{L}}$, for the time-frequency shift matrices $\pi(k,\ell)$ and $\pi(\tau,\nu)$ the properties*

$$\pi(k,\ell)\,\pi(\tau,\nu) = \omega^{-\nu k}\pi(\tau+k,\nu+\ell)\;, \tag{4.4}$$

$$\pi(k,\ell)\,\pi(\tau,\nu) = \omega^{\ell\tau-\nu k}\pi(\tau,\nu)\,\pi(k,\ell) \tag{4.5}$$

$$\pi(\tau,\nu)^* = \omega^{-\nu\tau}\pi(-\tau,-\nu)\;, \tag{4.6}$$

hold.

Proof. The proof heavily relies on Lemma 4.2.1. Applying Lemma 4.2.1, we get

$$\begin{aligned}\pi(k,\ell)\,\pi(\tau,\nu) &= \mathbf{M}^{\ell}\mathbf{T}^{k}\mathbf{M}^{\nu}\mathbf{T}^{\tau}\\ &= \omega^{-\nu k}\mathbf{M}^{\ell+\nu}\mathbf{T}^{\tau+k}\\ &= \omega^{\ell\tau-\nu k}\mathbf{M}^{\nu}\mathbf{T}^{\tau}\mathbf{M}^{\ell}\mathbf{T}^{k}\,,\end{aligned}$$

which proves (4.4) and (4.5). To derive (4.6), we observe

$$\begin{aligned}\pi(\tau,\nu)^* &= (\mathbf{M}^{\nu}\mathbf{T}^{\tau})^*\\ &= \mathbf{T}^{-\tau}\mathbf{M}^{-\nu}\\ &= \omega^{-\nu\tau}\mathbf{M}^{-\nu}\mathbf{T}^{\tau}\\ &= \omega^{-\nu\tau}\pi(-\tau,-\nu)\;,\end{aligned}$$

where we used $(\mathbf{M}^{\nu})^* = \mathbf{M}^{-\nu}$ and $(\mathbf{T}^{\tau})^* = \mathbf{T}^{-\tau}$. □

Now equipped with Lemma 4.2.6 we can prove Proposition 4.2.5.

Proof of Proposition 4.2.5. We are going to show $\pi(k,\ell)\,\mathbf{S}\pi(k,\ell)^* = \mathbf{S}$ for any $(k,\ell) \in \Lambda$ which implies the claimed assertion in the proposition. For an arbitrary $\mathbf{x} \in \mathbb{C}^L$ we have

$$\begin{aligned}\pi(k,\ell)\,\mathbf{S}\pi(k,\ell)^*\mathbf{x} &= \pi(k,\ell)\sum_{(\tau,\nu)\in\Lambda}\langle\pi(k,\ell)^*\mathbf{x},\pi(\tau,\nu)\,\mathbf{g}\rangle\,\pi(\tau,\nu)\,\mathbf{g}\\ &= \sum_{(\tau,\nu)\in\Lambda}\langle\pi(k,\ell)^*\mathbf{x},\pi(\tau,\nu)\,\mathbf{g}\rangle\,\pi(k,\ell)\,\pi(\tau,\nu)\,\mathbf{g}\\ &= \omega^{-\nu k}\sum_{(\tau,\nu)\in\Lambda}\langle\pi(k,\ell)^*\mathbf{x},\pi(\tau,\nu)\,\mathbf{g}\rangle\,\pi(\tau+k,\nu+\ell)\,\mathbf{g}\\ &= \omega^{-\nu k}\sum_{(\tau,\nu)\in\Lambda}\langle\mathbf{x},\pi(k,\ell)\,\pi(\tau,\nu)\,\mathbf{g}\rangle\,\pi(\tau+k,\nu+\ell)\,\mathbf{g}\end{aligned}$$

$$
\begin{aligned}
&= \underbrace{\omega^{-\nu k}\omega^{\nu k}}_{=1} \sum_{(\tau,\nu)\in\Lambda} \langle \mathbf{x}, \pi(\tau+k,\nu+\ell)\,\mathbf{g}\rangle\, \pi(\tau+k,\nu+\ell)\,\mathbf{g} \\
&= \sum_{(\tau,\nu)\in\Lambda} \langle \mathbf{x}, \pi(\tau,\nu)\,\mathbf{g}\rangle\, \pi(\tau,\nu)\,\mathbf{g} \\
&= \mathbf{S}\mathbf{x}\,,
\end{aligned}
$$

where we have used the definition of the frame operator $\mathbf{S}$ (Definition 4.1.5), Lemma 4.2.6 and in the final step that Λ is a subgroup of $\mathbb{Z}_L \times \mathbb{Z}_L$. □

An immediate consequence of Proposition 4.2.5 is the remarkable property of finite Gabor frames that the canonical dual frame of a finite Gabor frame is again a finite Gabor frame generated by the same time-frequency shifts. We formalize it in the following corollary.

Corollary 4.2.7. *Let $\{\pi(\lambda)\,\mathbf{g}\}_{\lambda\in\Lambda}$ be a finite Gabor frame for $\mathbb{C}^L$ with frame operator $\mathbf{S}$, if $\Lambda \subset \mathbb{Z}_L \times \mathbb{Z}_L$ is a subgroup of $\mathbb{Z}_L \times \mathbb{Z}_L$, then the canonical dual frame is again a finite Gabor frame and it is given by $\{\pi(\lambda)\,\mathbf{S}^{-1}\mathbf{g}\}_{\lambda\in\Lambda}$.*

Proof. Following Definition 4.1.7, the canonical dual frame of the frame $\{\pi(\lambda)\,\mathbf{g}\}_{\lambda\in\Lambda}$ is given by $\{\mathbf{S}^{-1}\pi(\lambda)\,\mathbf{g}\}_{\lambda\in\Lambda}$.
From Proposition 4.2.5 we have the commutation relation

$$\pi(\lambda)\,\mathbf{S} = \mathbf{S}\pi(\lambda)\ ,$$

for $\lambda \in \Lambda$. Applying $\mathbf{S}^{-1}$ from the left and from the right, to both sides of the equation we obtain

$$\mathbf{S}^{-1}\pi(\lambda) = \pi(\lambda)\,\mathbf{S}^{-1}\,.$$

Thus, we see that the commutation relation from Proposition 4.2.5 also holds for the inverse of the frame operator, $\mathbf{S}^{-1}$. The assertion in the corollary is now immediate. □

4.2.3. Linear independence of finite Gabor frames and matrix identification problem

Given $\Lambda \subset \mathbb{Z}_L \times \mathbb{Z}_L$ with $|\Lambda| \geq L$, another interesting question is whether it is possible to always find a vector $\mathbf{g} \in \mathbb{C}^L$ such that the set $\{\pi(\lambda)\,\mathbf{g}\}_{\lambda\in\Lambda}$ forms a frame. Indeed this is possible and [84] proves that such a window $\mathbf{g} \in \mathbb{C}^L$ always exists if L is a prime number and [85] extends the result to all dimensions, $L \in \mathbb{N}$.
Before we present the main results from [84] and [85] lets first turn our attention to a motivation which will also be of relevance to us. For now lets consider a

purely mathematical motivation and assume we have a linear equation of the form

$$\mathbf{b} = \mathbf{A} \cdot \mathbf{g}\,,$$

with $\mathbf{b}, \mathbf{g} \in \mathbb{C}^L$ and a $L \times L$ square matrix $\mathbf{A}$. The question of interest is whether we can obtain $\mathbf{A}$ from the observation $\mathbf{b}$ and the pilot signal $\mathbf{g}$. Recall that in Theorem 4.2.3 we have proven that the set of normalized time-frequency shift matrices forms an orthonormal basis for the set of square matrices on $\mathbb{C}^L$. Thus, there is a representation of $\mathbf{A}$ as a linear combination of normalized time-frequency shifts, i.e.

$$\mathbf{A} = \frac{1}{\sqrt{L}} \sum_{\lambda \in \mathbb{Z}_L \times \mathbb{Z}_L} \eta\left(\lambda\right) \pi\left(\lambda\right)\,,$$

where $\eta\left(\lambda\right)$ is called the spreading function of the matrix $\mathbf{A}$. Note that the values of the spreading function, $\eta\left(\lambda\right)$, are simply the coefficients of $\mathbf{A}$ in the basis of the normalized time-frequency matrices. Therefore, the values of the spreading function are simply given by the Hilbert-Schmidt inner products,

$$\eta\left(\lambda\right) = \frac{1}{\sqrt{L}} \left\langle \mathbf{A}, \pi\left(\lambda\right) \right\rangle_{\mathrm{HS}}\,,$$

for all $\lambda \in \mathbb{Z}_L \times \mathbb{Z}_L$. Substituting the expansion of $\mathbf{A}$ in the basis of time-frequency matrices we obtain for the observed vector $\mathbf{b}$,

$$\mathbf{b} = \mathbf{A}\mathbf{g} = \left(\frac{1}{\sqrt{L}} \sum_{\lambda \in \mathbb{Z}_L \times \mathbb{Z}_L} \eta\left(\lambda\right) \pi\left(\lambda\right) \right) \mathbf{g} = \frac{1}{\sqrt{L}} \sum_{\lambda \in \mathbb{Z}_L \times \mathbb{Z}_L} \eta\left(\lambda\right) \pi\left(\lambda\right) \mathbf{g}$$

$$= \frac{1}{\sqrt{L}} \left[\pi\left(\lambda_0\right) \mathbf{g} \,\middle|\, \pi\left(\lambda_1\right) \mathbf{g} \,\middle|\, \ldots \,\middle|\, \pi\left(\lambda_{L^2-1}\right) \mathbf{g} \right] \overrightarrow{\eta} = \mathbf{G}_{\mathbf{g}} \overrightarrow{\eta}\,, \quad (4.7)$$

where $\overrightarrow{\eta} \in \mathbb{C}^{L^2}$ is simply the values of the spreading function stacked in a vector, i.e. the coefficients of $\mathbf{A}$ in the basis of time-frequency shift matrices. Therefore, identifying $\mathbf{A}$ from the observation $\mathbf{b}$ reduces back to solving a linear equation. However, it is immediately obvious that if the support of $\mathbf{A}$ in the basis of time-frequency shift matrices is bigger then L the linear equation is an under determined system and has no unique solution. But in case where we have $\left|\mathrm{supp}\left(\mathbf{A}\right)\right| \leq L$ it is possible to recover $\mathbf{A}$ from $\mathbf{b}$, where the support of $\mathbf{A}$ is taken with respect to the basis of time-frequency shift matrices. We already saw that such a recovery is only possible if the submatrix of the matrix in (4.7)

indexed by $\lambda \in \operatorname{supp}(\mathbf{A})$ has full rank which is equivalent of its columns being linearly independent. Note that in fact the matrix in (4.7) is the Gabor matrix $\mathbf{G_g}$ generated with the window $\mathbf{g}$.
Now this discussion brings us back to the article [84], which poses the question whether it is possible to find a window $\mathbf{g} \in \mathbb{C}^L$ such that for any set $\Lambda \subset \mathbb{Z}_L \times \mathbb{Z}_L$ with $|\Lambda| \leq L$ the matrix

$$\left[\pi(\lambda_0)\,\mathbf{g} \,\middle|\, \pi(\lambda_1)\,\mathbf{g} \,\middle|\, \ldots \,\middle|\, \pi\left(\lambda_{|\Lambda|}\right)\mathbf{g}\right],$$

has full column rank or equivalently whether its columns are linearly independent. Note that the above matrix is a submatrix of the full Gabor matrix $\mathbf{G_g}$ in (4.7) composed of those columns that are indexed by Λ. In fact, this formulation is the same as to ask whether we are able to find a window $\mathbf{g}$ such that the Gabor matrix $\mathbf{G_g}$ generated by the window $\mathbf{g}$ has spark $L+1$, $\operatorname{spark}(\mathbf{G_g}) = L+1$.
This question is answered in [84] for dimensions L where L is a prime number and extended to dimensions where L is not a prime number in the article [85]. We state the main result of [84] in the following theorem.

Theorem 4.2.8 ([84])**.** *If L is a prime number then there is a dense open set E of full measure in $\mathbb{C}^L$ such that for every $\mathbf{g} \in E$, the Gabor matrix $\mathbf{G_g}$ generated by the window vector $\mathbf{g}$ has* $\operatorname{spark}(\mathbf{G_g}) = L+1$.

The following is the main result of [85].

Theorem 4.2.9 ([85])**.** *For every positive integer L, there is some window $\mathbf{g} \in \mathbb{C}^L$, such that the Gabor matrix $\mathbf{G_g}$ generated by $\mathbf{g}$ has* $\operatorname{spark}(\mathbf{G_g}) = L+1$. *Moreover, the set of all such windows $\mathbf{g} \in \mathbb{C}^L$ is of full measure, i.e. its complement in $\mathbb{C}^L$ has Lebesgue measure zero.*

Furthermore, in [85] an explicit construction of such a window $\mathbf{g} \in \mathbb{C}^L$ is given in a corollary which we state here.

Proposition 4.2.10 (Corollary 5.3 in [85])**.** *Let $\xi = \mathrm{e}^{j\frac{2\pi}{(L-1)^4}}$, or any other primitive root of unity of order $(L-1)^4$, where $L \geq 4$. Then the vector*

$$\mathbf{g} = \left[\; 1\,,\;\; \xi\,,\;\; \xi^4\,,\;\; \xi^9\,, \ldots,\;\; \xi^{(N-1)^2}\right],$$

generates a Gabor matrix $\mathbf{G_g}$ with $\operatorname{spark}(\mathbf{G_g}) = L+1$.

Theorems 4.2.8 and 4.2.9 imply that in every dimension $L \in \mathbb{N}$ it is possible to chose a window vector $\mathbf{g} \in \mathbb{C}^L$ such that the generated finite Gabor frame is

robust to erasures, that is the finite Gabor frame remains a frame for $\mathbb{C}^L$ even if elements are removed. This is a consequence of the fact that for finite Gabor frames generated with a suitable window vector $\mathbf{g}$ every L-element subset of the finite Gabor frame are linearly independent vectors, i.e. span $\mathbb{C}^L$.

4.2.4. Short-time Fourier transform and uncertainty principles in finite time-frequency representations

The short-time Fourier transform (STFT) is a time-frequency analysis tool that has by now a long standing history in signal processing [92, 97–99]. Here we are interested in its finite version and in an uncertainty principle of the finite short-time Fourier transform discussed in [87].
We first begin by formally defining the short-time Fourier transform.

Definition 4.2.1. *The short-time Fourier transform* $\mathrm{V}_{\mathbf{g}} : \mathbb{C}^L \to \mathbb{C}^L \times \mathbb{C}^L$ *with respect to the window vector* $\mathbf{g} \in \mathbb{C}^L \setminus \{0\}$ *of a signal* $\mathbf{x} \in \mathbb{C}^L$ *is given by*

$$\left(\mathrm{V}_{\mathbf{g}}\mathbf{x}\right)[\tau,\nu] = \langle \mathbf{x}, \pi(\tau,\nu)\,\mathbf{g}\rangle \ ,$$

where $(\tau,\nu) \in \mathbb{Z}_L \times \mathbb{Z}_L$.

The inversion formula for the short-time Fourier transform is given as

$$\begin{aligned}\mathbf{x}[n] &= \frac{1}{L\,\|\mathbf{g}\|_2^2} \sum_{\tau=0}^{L-1}\sum_{\nu=0}^{L-1} \left(\mathrm{V}_{\mathbf{g}}\mathbf{x}\right)[\tau,\nu]\,\mathbf{g}[n-\tau]\,\mathrm{e}^{-j\frac{2\pi}{L}n\nu} \\ &= \frac{1}{L\,\|\mathbf{g}\|_2^2} \sum_{\tau=0}^{L-1}\sum_{\nu=0}^{L-1} \langle \mathbf{x}, \pi(\tau,\nu)\,\mathbf{g}\rangle\,\pi(\tau,\nu)\,\mathbf{g}[n] \ ,\end{aligned}$$

for $n = 0, 1, \ldots, L-1$ and for any $\mathbf{x} \in \mathbb{C}^L$.
Next we want to introduce some uncertainty principles for finite dimensions. We will present some uncertainty principles concerning finite discrete Fourier transforms although they will not be explicitly used in later sections. Nonetheless, those uncertainty principles give insights into the discrete Fourier transform and belong to an introductory discussion for the actual uncertainty principle on the finite short-time Fourier transform which will play a central role in a later section.
We begin with a fundamental uncertainty principle that is defined for functions which are defined on $\mathbb{Z}_L$ [100].

Theorem 4.2.11. *Let $\mathbf{x} \in \mathbb{C}^L \setminus \{0\}$ be a vector and denote by $\widehat{\mathbf{x}} \in \mathbb{C}^L$ the discrete Fourier transform of $\mathbf{x}$, i.e.*

$$\widehat{\mathbf{x}} = \frac{1}{\sqrt{L}} \cdot \mathbf{F}\mathbf{x}\,,$$

where $\mathbf{F}$ is the discrete Fourier transform matrix. Then for the support of $\mathbf{x}$ and $\widehat{\mathbf{x}}$ the relation

$$|\mathrm{supp}\,(\mathbf{x})| \cdot |\mathrm{supp}\,(\widehat{\mathbf{x}})| \geq L\,,$$

holds.

Theorem 4.2.11 is proven in [100]. Furthermore, the bound in Theorem 4.2.11 is sharp. That means, if for instance an integer n divides L, i.e. $\frac{L}{n} \in \mathbb{N}$ then there exists a vector $\mathbf{x} \in \mathbb{C}^L$ with $|\mathrm{supp}\,(\mathbf{x})| = n$ and $|\mathrm{supp}\,(\widehat{\mathbf{x}})| = \frac{L}{n}$. Note that Theorem 4.2.11 provides a finite analogous to the standard continuous uncertainty principle in Fourier analysis [101].
If the dimension L is a prime number then the bound in Theorem 4.2.11 can be improved significantly. Indeed, in [102] a bound on the addition of the support sizes of the original signal and its discrete Fourier transform is derived.

Theorem 4.2.12. *Let L be a prime number and $\mathbf{x} \in \mathbb{C}^L \setminus \{0\}$ be a vector and denote by $\widehat{\mathbf{x}} \in \mathbb{C}^L$ the discrete Fourier transform of $\mathbf{x}$. Then*

$$|\mathrm{supp}\,(\mathbf{x})| + |\mathrm{supp}\,(\widehat{\mathbf{x}})| \geq L\,,$$

holds.

The Theorem 4.2.12 is proven in [102]. The proof uses that every minor of the discrete Fourier transform matrix, $\mathbf{F} \in \mathbb{C}^{L \times L}$ where L is a prime number, is non-singular.
The next theorem on the uncertainty principle on the finite short-time Fourier transform will play a key role in a later section.

Theorem 4.2.13 (Theorem 4.4 in [87])**.** *Let L be a prime number. Then for almost every window vector $\mathbf{g} \in \mathbb{C}^L$, we have*

$$|\mathrm{supp}\,(\mathbf{x})| + |\mathrm{supp}\,(\mathrm{V}_{\mathbf{g}}\mathbf{x})| \geq L^2 + 1\,,$$

for all $\mathbf{x} \in \mathbb{C}^L \setminus \{0\}$. Moreover, for two integers $k, \ell \in \mathbb{N}$ assume $1 \leq k \leq L$ and $1 \leq \ell \leq L^2$ with $k + \ell \geq L^2 + 1$, then there exists a vector $\mathbf{x} \in \mathbb{C}^L$ with $|\mathrm{supp}\,(\mathbf{x})| = k$ and $|\mathrm{supp}\,(\mathrm{V}_{\mathbf{g}}\mathbf{x})| = \ell$.

Furthermore, regarding the window vector $\mathbf{g} \in \mathbb{C}^L$ in Theorem 4.2.13 one can make the following statement.

Proposition 4.2.14 (Proposition 4.6 in [87]). *There exists a unimodular window vector $\mathbf{g} \in \mathbb{C}^L$, with L a prime number, that satisfies the conclusions of Theorem 4.2.13. A unimodular vector $\mathbf{g} \in \mathbb{C}^L$ is defined as a vector where the absolute value of all its entries is equal to 1, i.e. $|\mathbf{g}[n]| = 1$ for $n \in \mathbb{Z}_L$.*

The following theorem summarizes the equivalences of properties of the finite Gabor frames and the finite short-time Fourier transform discussed so far.

Theorem 4.2.15 (Theorem 6.9 in [19]). *Let L be a prime number. The following are equivalent for a window vector $\mathbf{g} \in \mathbb{C}^L \setminus \{0\}$:*

1. *The Gabor matrix $\mathbf{G_g}$ satisfies $\operatorname{spark}(\mathbf{G_g}) = L + 1$.*
2. *The finite Gabor frame $\{\pi(\lambda)\mathbf{g}\}_{\lambda \in \mathbb{Z}_L \times \mathbb{Z}_L}$ forms an equal norm tight frame which is maximally robust to erasures.*
3. *For all $\mathbf{x} \in \mathbb{C}^L \setminus \{0\}$, $|\operatorname{supp}(\mathrm{V_g x})| \geq L^2 - L + 1$.*
4. *For all $\mathbf{x} \in \mathbb{C}^L \setminus \{0\}$, $\mathrm{V_g x}$ and, therefore, $\mathbf{x}$ is completely determined by its values on any set $\Lambda \subset \mathbb{Z}_L \times \mathbb{Z}_L$ with $|\Lambda| = L$.*
5. *A matrix $\mathbf{A} \in \mathbb{C}^{L \times L}$ is identifiable by $\mathbf{g}$ if and only if $|\operatorname{supp}(\mathbf{A})| \leq L$, where the support of $\mathbf{A}$ is taken with respect to the basis of time-frequency shift matrices.*

The proof of Theorem 4.2.15 can be found in [19]. Further, it should be mentioned that Theorem 4.2.15 actually holds in the more general setup of Gabor analysis on finite Abelian groups [82] and is in fact formulated in that setup in [19]. Since we index the vectors in $\mathbb{C}^L$ with elements from $\mathbb{Z}_L$, whenever L is a prime number, we considered in fact simply Abelian groups of order L. But for our discussion, simply considering the vector space $\mathbb{C}^L$ where L is a prime number is enough for now.

Part II.

Properties of Finite Gabor Frames

5. On the Statistical Restricted Isometry Property for Finite Gabor Frames

Gabor matrices are important in many different areas that utilize tools from time-frequency analysis like communications or radar [43, 83, 103, 104]. However, for applications with sparse data the question arises whether these matrices satisfy some recovery guarantees for compressed sensing and furthermore, which generating window vectors yield a matrix with restricted isometry property.
In the literature a well known fact is that finite Gabor frames that are generated by random window vectors, for instance where each entry of the window vector is independent and identically distributed (i.i.d) random variables (for example a subguassian random variable), deliver suitable sensing matrices [88, 89, 105]. Here the matrix representation of the synthesis operator of the finite Gabor frame is used as the sensing matrix or, as introduced in Section 4, the Gabor matrix $\mathbf{G_g}$ generated with the window vector $\mathbf{g} \in \mathbb{C}^L$.
Although random constructions of Gabor sensing matrices deliver very good theoretical results, they do, however, have major drawbacks in terms of reconstruction complexity and efficiency, see for instance [57, 70]. Moreover, in applications, the measurement matrices are often not random but have a certain structure that is predefined by the measurement setup. But since proving the restricted isometry property is difficult, or rather there is almost no known deterministic constructions of measurement matrices for compressed sensing fulfilling the restricted isometry property [61]. As well as, the verification of the restricted isometry property for a given measurement matrix is computationally unfeasible [56], there has been attempts to analyse the statistical recovery guarantees of deterministic sensing matrices, where the probability enters the signal model (the measured signals are assumed to be random) rather than the measurement matrix. We have already discussed in Section 3 the strong coherence property [67, 71] and the statistical restricted isometry property [70]. In this section we will pay special attention to the statistical restricted isometry property of finite Gabor frames (Gabor sensing matrices). Despite the impor-

tance of finite Gabor frames in various applications [19, 43, 83, 86, 103, 104, 106, 107], there aren't much reports on the sparse signal recovery performance of deterministic constructions of finite Gabor frames. Analysis regarding sparse recovery performance of deterministic finite Gabor frames are provided in [27, 43, 67].
The results presented in this section have already been partially published in [27].

5.1. Construction of low coherence finite Gabor frames with Alltop window vector

In [91] Alltop constructed a window vector $\boldsymbol{\alpha} \in \mathbb{C}^L$ that gives rise to a low coherence finite Gabor frame. The aim of [91] is to construct discrete periodic sequences that have low correlation values, since these sequences have applications in radar and communication systems [83, 90, 108].
Let $\omega = \mathrm{e}^{j\frac{2\pi}{L}}$ and $L \geq 5$ a prime number, then the window vector $\boldsymbol{\alpha} \in \mathbb{C}^L$ suggested in [91] is given by

$$\boldsymbol{\alpha}[k] = \frac{1}{\sqrt{L}}\omega^{k^3} \tag{5.1}$$

for $k = 0, 1, \ldots, L-1$, we will refer to $\boldsymbol{\alpha}$ as the Alltop window vector. Note that by design we have $\|\boldsymbol{\alpha}\|_2 = 1$. Our aim in this section is to show that the Gabor matrix generated by the Alltop window vector, $\mathbf{G}_{\boldsymbol{\alpha}}$, has low coherence (Definition 2.4.1), more precisely we have $\mu(\mathbf{G}_{\boldsymbol{\alpha}}) = \frac{1}{\sqrt{L}}$.

Proposition 5.1.1. *Set $\omega = \mathrm{e}^{j\frac{2\pi}{L}}$ and let $L \in \mathbb{N}$ be a prime number with $L \geq 5$, define the vector $\boldsymbol{\alpha} \in \mathbb{C}^L$ as in* (5.1) *Then the Gabor frame generated with the window vector $\boldsymbol{\alpha}$ has coherence $\mu(\mathbf{G}_{\boldsymbol{\alpha}}) = \frac{1}{\sqrt{L}}$.*
Further, the absolute values of the inner products between the elements of the finite Gabor frame are given by

$$|\langle \pi(\tau,\nu)\boldsymbol{\alpha}, \boldsymbol{\alpha}\rangle| = \begin{cases} 1, & \text{if } (\tau,\nu) = (0,0) \\ 0, & \text{if } \tau = 0 \text{ and } \nu \neq 0 \\ \frac{1}{\sqrt{L}}, & \text{if } \tau \neq 0 \end{cases}.$$

The following proof follows the proof of Proposition 5.13 in [37].

Proof. First, note that we only need to consider the inner products $|\langle \pi(\tau,\nu)\boldsymbol{\alpha}, \boldsymbol{\alpha}\rangle|$ with $(\tau,\nu) \in \mathbb{Z}_L \times \mathbb{Z}_L$. That can easily be seen as follows, let $(\tau,\nu), (k,\ell) \in$

$\mathbb{Z}_L \times \mathbb{Z}_L$ and $(\tau, \nu) \neq (k, \ell)$ and consider the absolute value of the inner product $|\langle \pi(\tau,\nu)\boldsymbol{\alpha}, \pi(k,\ell)\boldsymbol{\alpha}\rangle|$ as follows

$$\begin{aligned}|\langle \pi(\tau,\nu)\boldsymbol{\alpha}, \pi(k,\ell)\boldsymbol{\alpha}\rangle| &= |\langle \pi(k,\ell)^* \pi(\tau,\nu)\boldsymbol{\alpha}, \boldsymbol{\alpha}\rangle| \\ &= |\langle \pi(-k,-\ell)\pi(\tau,\nu)\boldsymbol{\alpha}, \boldsymbol{\alpha}\rangle| \\ &= |\langle \pi(\tau-k,\nu-\ell)\boldsymbol{\alpha}, \boldsymbol{\alpha}\rangle| \ ,\end{aligned}$$

where we applied Lemma 4.2.6. Since we have $(\tau - k, \nu - \ell) \in \mathbb{Z}_L \times \mathbb{Z}_L$ it is enough if we only consider $|\langle \pi(\tau,\nu)\boldsymbol{\alpha}, \boldsymbol{\alpha}\rangle|$ with $(\tau,\nu) \in \mathbb{Z}_L \times \mathbb{Z}_L$.
The absolute value of the inner product $|\langle \pi(\tau,\nu)\boldsymbol{\alpha}, \boldsymbol{\alpha}\rangle|$ can explicitly be written as

$$\begin{aligned}|\langle \pi(\tau,\nu)\boldsymbol{\alpha}, \boldsymbol{\alpha}\rangle| &= \frac{1}{L}\left|\sum_{\ell=0}^{L-1} \omega^{\ell\nu}\overline{\boldsymbol{\alpha}[\ell]}\boldsymbol{\alpha}[\ell-\tau]\right| \\ &= \frac{1}{L}\left|\sum_{\ell=0}^{L-1} \omega^{\ell\nu}\omega^{-\ell^3}\omega^{(\ell-\tau)^3}\right| \\ &= \frac{1}{L}\left|\sum_{\ell=0}^{L-1} \omega^{-3\tau\ell^2}\omega^{\left(3\tau^2+\nu\right)\ell}\omega^{-\tau^3}\right| \\ &= \frac{1}{L}\left|\sum_{\ell=0}^{L-1} \omega^{-3\tau\ell^2}\omega^{\left(3\tau^2+\nu\right)\ell}\right| \ .\end{aligned}$$

Now we set $a = -3\tau$ and $b = 3\tau^2 + \nu$ and consider the squared value of the absolute value of the inner product

$$\begin{aligned}|\langle \pi(\tau,\nu)\boldsymbol{\alpha}, \boldsymbol{\alpha}\rangle|^2 &= \frac{1}{L^2}\sum_{\ell,k=0}^{L-1} \omega^{a\ell^2}\omega^{b\ell}\omega^{-ak^2}\omega^{-bk} \\ &= \frac{1}{L^2}\sum_{\ell,k=0}^{L-1} \omega^{a(\ell^2-k^2)}\omega^{b(\ell-k)} \\ &= \frac{1}{L^2}\sum_{\ell,k=0}^{L-1} \omega^{(\ell-k)(a(l+k)+b)} \ .\end{aligned}$$

For the next step we set $\gamma = \ell - k$ and obtain

$$|\langle \pi(\tau,\nu)\boldsymbol{\alpha}, \boldsymbol{\alpha}\rangle|^2 = \frac{1}{L^2}\sum_{\gamma,k=0}^{L-1} \omega^{\gamma(a(\gamma+2k)+b)}$$

$$= \frac{1}{L^2} \sum_{\gamma=0}^{L-1} \omega^{\gamma(a\gamma+b)} \sum_{k=0}^{L-1} \omega^{2a\gamma k} .$$

From Lemma 4.2.2 we have for the final sum in the above equation,

$$\sum_{k=0}^{L-1} \omega^{2a\gamma k} = \begin{cases} L, & \text{if } 2a\gamma = 0 \mod L \\ 0, & \text{if } 2a\gamma \neq 0 \mod L \end{cases} .$$

Since $a = -3\tau$ and because $6 \mod L \neq 0$ the only case where the sum does not equal to zero is when $\tau = 0$ or when $\gamma = 0$.
We consider now these two cases separately.
Case 1 $\boldsymbol{\tau = 0}$: In this case we have $a = -3\tau = 0$ and $b = 3\tau + \nu = \nu$. Then we obtain

$$\begin{aligned} \left|\langle \pi(\tau,\nu)\boldsymbol{\alpha}, \boldsymbol{\alpha}\rangle\right|^2 &= \frac{1}{L^2} \sum_{\gamma=0}^{L-1} \omega^{\gamma\nu} \sum_{k=0}^{L-1} 1 \\ &= \frac{1}{L} \sum_{\gamma=0}^{L-1} \omega^{\gamma\nu} \\ &= \begin{cases} 1, & \text{if } \nu = 0 \mod L \\ 0, & \text{if } \nu \neq 0 \mod L \end{cases} . \end{aligned}$$

This proves the first two lines of the statement, and it remains to consider the case $\tau \neq 0$.
Case 2 $\boldsymbol{\tau \neq 0}$: Finally, we turn our attention to the case where $\gamma = 0$ and $\tau \neq 0$. Note that the sum $\sum_{k=0}^{L-1} \omega^{2a\gamma k}$ vanishes for all other cases, i.e. whenever $\gamma \neq 0$. Thus, in this final case we obtain

$$\begin{aligned} \left|\langle \pi(\tau,\nu)\boldsymbol{\alpha}, \boldsymbol{\alpha}\rangle\right|^2 &= \frac{1}{L^2} \sum_{\gamma,k=0}^{L-1} \omega^{\gamma(a(\gamma+2k)+b)} \\ &= \frac{1}{L} + \frac{1}{L^2} \sum_{\gamma=1}^{L-1} \omega^{\gamma(a\gamma+b)} \underbrace{\sum_{k=0}^{L-1} \omega^{2a\gamma k}}_{=0} \\ &= \frac{1}{L} . \end{aligned}$$

Hence, we obtain for $\tau \neq 0$,

$$\left|\langle \pi(\tau,\nu)\boldsymbol{\alpha}, \boldsymbol{\alpha}\rangle\right| = \frac{1}{\sqrt{L}} ,$$

which proves the claimed assertion that $\mu\left(\mathbf{G}_{\boldsymbol{\alpha}}\right) = \frac{1}{\sqrt{L}}$. □

It is interesting to see that the finite Gabor frames constructed with the Alltop window vector $\boldsymbol{\alpha}$ can be splitted into a disjoint collection of L orthogonal bases for $\mathbb{C}^L$. The L disjoint sets are given by $\{\pi\left(\tau,\nu\right)\boldsymbol{\alpha}\}_{\nu\in\mathbb{Z}_L}$ for $\tau = 0, 1, \ldots, L-1$. Adding the canonical basis, $\{\mathbf{e}_k\}_{k=0}^{L-1}$ we get in total $L+1$ disjoint orthogonal bases for $\mathbb{C}^L$. Furthermore, it is easy to see that we have

$$\left|\left\langle \pi\left(\tau,\nu\right),\mathbf{e}_k\right\rangle\right| = \frac{1}{\sqrt{L}},$$

for all $(\tau,\nu)\in\mathbb{Z}_L\times\mathbb{Z}_L$ and for all $k = 0, 1, \ldots, L-1$. Hence, consider the matrix $\boldsymbol{\Psi} = \left[\mathbf{G}_{\alpha}\middle|\mathrm{I}_L\right]$, then the coherence of $\boldsymbol{\Psi}$ is $\mu\left(\boldsymbol{\Psi}\right) = \frac{1}{\sqrt{L}}$. Additionally, the columns of $\boldsymbol{\Psi}$ form a unit norm tight frame for $\mathbb{C}^L$.
In fact, these type of collections of vectors, $\{\boldsymbol{\psi}_k\}_{k=1}^{L^2+L}$ which is just a collection of the sets $\{\pi\left(\tau,\nu\right)\boldsymbol{\alpha}\}_{\nu\in\mathbb{Z}_L}$ for $\tau = 0, 1, \ldots, L-1$ together with $\{\mathbf{e}_k\}_{k=0}^{L-1}$, and where $\boldsymbol{\psi}_k$ is the k-th column of $\boldsymbol{\Psi}$, are known as mutually unbiased bases in the literature [109, 110]. Two orthogonal bases are called mutually unbiased if the inner product of any two vectors, each from a distinct orthogonal basis, has the absolute value equal to $\frac{1}{\sqrt{L}}$, we assume that the vectors are normed with respect to the ℓ_2-norm. In fact, in $\mathbb{C}^L$ there are at most $L+1$ orthogonal bases that are mutually unbiased to each other [110, 111].

5.2. Statistical restricted isometry property of finite Gabor systems with Alltop window

In this subsection we are going to state and discuss the main result of this section. That is, the Gabor matrix generated with the Alltop window vector $\boldsymbol{\alpha}$ (5.1) has the statistical restricted isometry property with respect to Theorem 3.1.1. The main result will be proven later in Section 5.4.

Unfortunately, it is not possible to directly apply the framework of the statistical restricted isometry property as discussed in [70] or in Section 3.1 to Gabor matrices, due to the nature of the η-StRIP-ability as given in Definition 3.1.2. In fact, it is not difficult to see that a Gabor matrix $\mathbf{G}_{\mathbf{g}}$ cannot be η-StRIP-able (in the sense of the Definition 3.1.2) for any window vector $\mathbf{g}\in\mathbb{C}^L$. The easiest way to see that a Gabor matrix cannot be η-StRIP-able is to observe that Conditions S2 and S3 from Definition 3.1.2 cannot be satisfied simultaneously for a finite Gabor frame. The Condition S2 requires the columns of the matrix

$\mathbf{\Phi}$ to have a group structure under pointwise multiplication and that an identity vector with all ones exists. However, an all ones vector - which is also the first column of the discrete Fourier matrix, $\mathbf{f}_0$ - is an eigenvector of the translation matrix $\mathbf{T}$. Thus, the sets of the form $\{\pi(\tau,\nu)\,\mathbf{f}_0\}_{\tau\in\mathbb{Z}_L}$ for a fixed $\nu\in\mathbb{Z}_L$ all will have a one dimensional span. For instance the set $\{\pi(\tau,0)\,\mathbf{f}_0\}_{\tau\in\mathbb{Z}_L}$ will consist only of one vector which is the all ones vector, hence it is not possible to fulfill Condition S3 for the Gabor matrix $\mathbf{G}_{\mathbf{f}_0}$.
Nevertheless, it is reported that the Gabor matrix generated by the Alltop window vector has good sparse recovery performance, or equivalently, it is a suitable sensing matrix for compressed sensing. The Gabor matrix generated by the Alltop window has low coherence, as discussed in Section 5.1, however the performance evaluated in [43] outperforms the recovery guarantees that coherence based proofs can provide. Therefore, we will use the framework of statistical restricted isometry property to give a theoretical justification on the performance of the Gabor matrix generated with the Alltop window vector.
Before we begin with the discussion of the main result of this section we first need to introduce the following notation

$$\Sigma_k^N = \left\{\mathbf{x}\in\mathbb{C}^N \;:\; |\mathrm{supp}\,(\mathbf{x})|\le k\right\}.$$

Essentially, by Σ_k^N we denote the set of all N dimensional vectors with sparsity k.
Now we are going to discuss the main result of this section. The proofs of Theorem 5.2.1 and Theorem 5.2.2 will be discussed later in Section 5.4. For now we will state the main result of this section and discuss its implications. The proof is long and requires multiple auxiliary lemmas, therefore both theorems are going to be proven in a separate subsection.

Theorem 5.2.1 ([27]). *Let $L\ge 5$ be a prime number of the form $L=3n+2$ with $n\in\mathbb{N}$, and let $\mathbf{G}_{\boldsymbol{\alpha}}$ be the Gabor matrix generated by the Alltop window vector $\boldsymbol{\alpha}\in\mathbb{C}^L$.*
If $k<1+(L^2-1)\delta$ then $\mathbf{G}_{\boldsymbol{\alpha}}$ has (k,δ,ϵ)-StRIP, i.e.

$$(1-\delta)\,\|\mathbf{f}\|_2^2\le\|\mathbf{G}_{\boldsymbol{\alpha}}\|_2^2\le(1+\delta)\,\|\mathbf{f}\|_2^2$$

holds with probability exceeding $1-\epsilon$ for all random $\mathbf{f}\in\Sigma_k^{L^2}$ uniform distributed over the set $\mathcal{S}^{L^2-1}=\left\{\mathbf{x}\in\Sigma_k^{L^2}\;:\;\|\mathbf{x}\|_2^2=1\right\}$, where

$$\epsilon=2\exp\left(-\left(\delta-\frac{k-1}{L^2-1}\right)^2\frac{L}{8k}\right).$$

Theorem 5.2.1 shows that a Gabor matrix $\mathbf{G}_{\boldsymbol{\alpha}}$ generated by the Alltop window vector $\boldsymbol{\alpha}$ has statistical restricted isometry property. However, the statistical restricted isometry property does not guarantee unique recovery of the signals, even not with high probability. In fact, we know from Theorem 4.2.8 that $\mathbf{G}_{\mathbf{g}}$ has full spark for almost all $\mathbf{g} \in \mathbb{C}^L$. But the matrix $\mathbf{G}_{\boldsymbol{\alpha}}$ does not have this property. Consider the submatrix of $\mathbf{G}_{\boldsymbol{\alpha}}$ given by

$$\mathbf{M} = \left[\mathbf{T}^0\boldsymbol{\alpha}\middle|\ \ \mathbf{T}^1\boldsymbol{\alpha}\middle|\dots\middle|\mathbf{T}^{L-1}\boldsymbol{\alpha}\right].$$

The easiest way to see that $\mathbf{M}$ is singular is to consider $\mathbf{M}^*\mathbf{f}_0$ where $\mathbf{f}_0$ is the first column of the discrete Fourier matrix or the all ones vector. Note that if $\mathbf{M}^*$ is singular then $\mathbf{M}$ is also singular. Hence, we have

$$\mathbf{M}^*\mathbf{f}_0 = \begin{bmatrix} \boldsymbol{\alpha}^* \left(\mathbf{T}^0\right)^* \mathbf{f}_0 \\ \boldsymbol{\alpha}^* \left(\mathbf{T}^1\right)^* \mathbf{f}_0 \\ \vdots \\ \boldsymbol{\alpha}^* \left(\mathbf{T}^{L-1}\right)^* \mathbf{f}_0 \end{bmatrix},$$

since $\mathbf{f}_0$ is an eigenvector of $\mathbf{T}^k$ with eigenvalue 1 for $k = 0, 1, \dots L-1$, we obtain for each entry in $\mathbf{M}^*\mathbf{f}_0$ the same value which is $\langle \mathbf{f}_0, \boldsymbol{\alpha} \rangle$. Now, observe

$$\langle \mathbf{f}_0, \boldsymbol{\alpha} \rangle = \sum_{x=0}^{L-1} \omega^{-x^3} = 0$$

holds, where we used Lemma 5.4.2 and Lemma 4.2.2 in the last step. Thus, $\mathbf{f}_0$ is in the kernel of $\mathbf{M}$.
Nevertheless, unique statistical restricted isometry property (Definition 3.1.1) requires the uniqueness only up to high probability with respect to the uniform distribution of $\mathbf{f}$ over all $\mathcal{S}^{L^2-1}$. Therefore, even though $\mathbf{G}_{\boldsymbol{\alpha}}$ does not has full spark, it may still have unique statistical restricted isometry property. Indeed, this is the case as stated in the following Theorem.

Theorem 5.2.2 ([27]). *Let $L \geq 5$ be a prime number of the form $L = 3n + 2$ with $n \in \mathbb{N}$, and let $\mathbf{G}_{\boldsymbol{\alpha}}$ be the Gabor matrix generated with the Alltop window vector $\boldsymbol{\alpha} \in \mathbb{C}^L$.*
If $k < 1 + (L^2 - 1)\delta$ and $L > c\frac{k \log L}{\delta^2}$ for some constant $c > 0$, then $\mathbf{G}_{\boldsymbol{\alpha}}$ has $(k, \delta, 2\epsilon)$-UStRIP where ϵ is given by

$$\epsilon = 2 \exp\left(-\left(\delta - \frac{k-1}{L^2-1}\right)^2 \frac{L}{8k}\right).$$

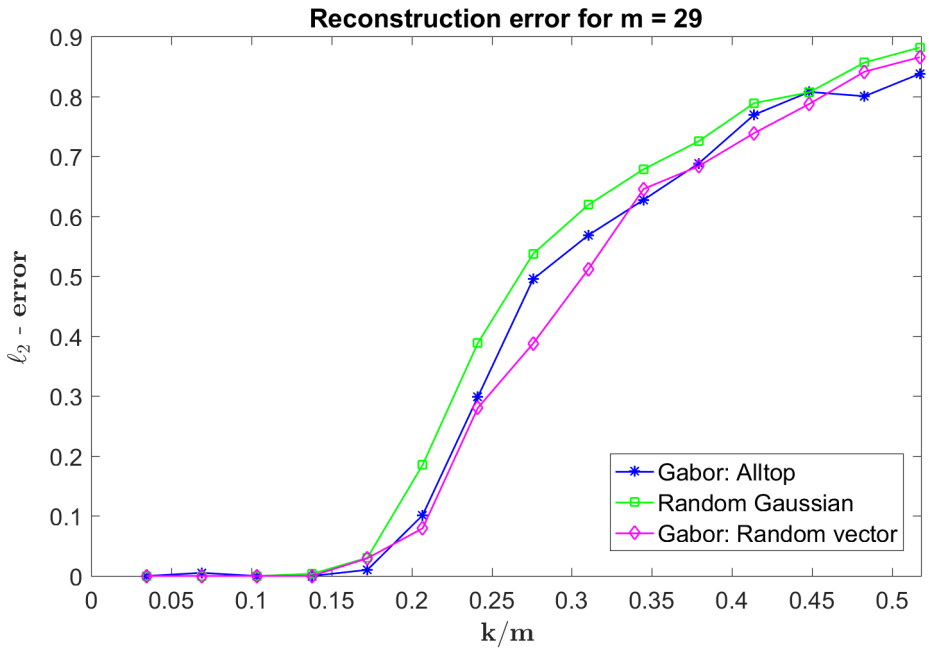

Figure 5.1.: Simulation results for $L = 29$ measurements. Depicted is the reconstruction error in the ℓ_2-norm. Each data point is averaged over 100 experiments.

As mentioned before, Theorem 5.2.2 will be proved in Section 5.4. For now we discuss some consequences of Theorem 5.2.2.
The difference between the result of Theorem 5.2.2 and known results on RIP of Gabor matrices generated by a random window vector [86, 88, 89] is that those results consider random matrices and arbitrary signals. We on the other hand consider deterministic matrices and random signals.
We want to emphasise that Theorem 5.2.2 shows that for the number of measurements $L > ck \log L\delta^{-2}$ the matrix $\mathbf{G}_{\boldsymbol{\alpha}}$ has uniqueness guaranteed statistical restricted isometry property (UStRIP). The necessary number of measurements L grows linearly (up to a logarithmic factor) with the sparsity k. This is a similar behaviour as for random matrices [37, 50].

5.3. Numerical experiments

This section presents some numerical experiments which should illustrate the performance of a Gabor matrix generated by the Alltop window. The numerical experiments support the theoretical observations made in Theorem 5.2.1 and

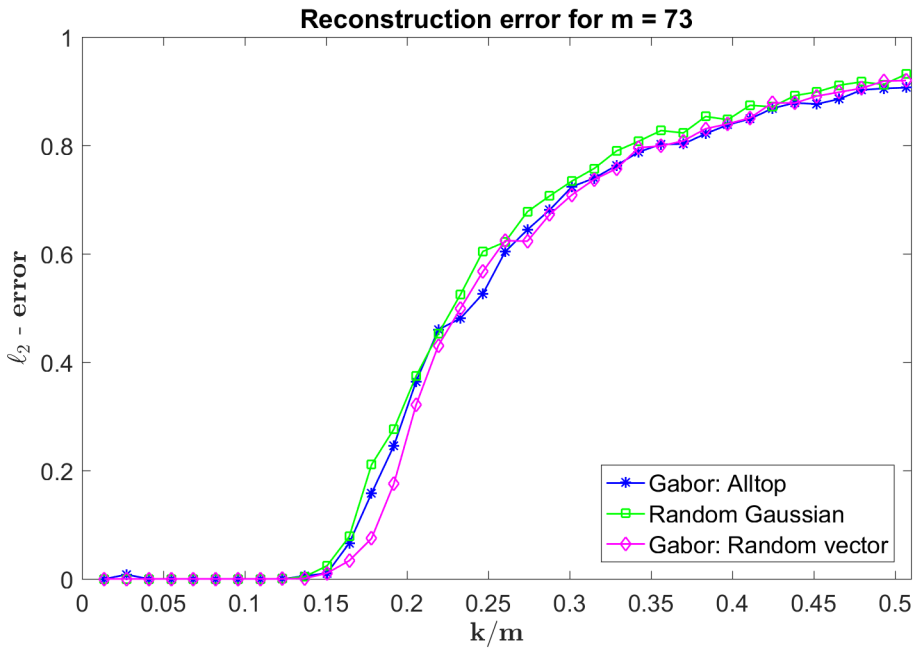

Figure 5.2.: Simulation results for $L = 73$ measurements. Depicted is the reconstruction error in the ℓ_2-norm. Each data point is averaged over 100 experiments.

Theorem 5.2.2. Figures 5.1 and 5.2 show simulation results where k-sparse vectors $\mathbf{f}$ were recovered from measurements of the form $\mathbf{g} = \mathbf{A}\mathbf{f}$ using basis pursuit, i.e.

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}\in\mathbb{C}^{L^2}} \|\mathbf{f}\|_1 \quad \text{subject to} \quad \mathbf{A}\mathbf{f} = \mathbf{g}\,.$$

The support of the vectors $\mathbf{f}$ were chosen uniformly random and the entries where chosen as i.i.d. normal random variables.
As measurement matrix $\mathbf{A}$, we compared the matrices:

1) The Gabor matrix $\mathbf{G}_{\boldsymbol{\alpha}}$ generated by the Alltop window vector.

2) A random Gaussian matrix whose entries are independent, identically distributed normal random variables.

3) A Gabor matrix with a random Gaussian window vector $\boldsymbol{\Phi}$. All entries of $\boldsymbol{\Phi}$ are independent, identically distributed normal random variables.

All matrices are of size $L \times L^2$. In the numerical simulations the sparsity k of the data vector $\mathbf{f}$ has been varied and for each value of k the normalized

quadratic reconstruction error $\frac{\left\|\mathbf{f}-\hat{\mathbf{f}}\right\|_2^2}{\|\mathbf{f}\|_2^2}$ has been evaluated. For each value of k, 100 experiments were conducted and the average value was computed and is illustrated as the data point in the Figures 5.1 and 5.2. Hence, each data point in the figures is the average reconstruction error of 100 experiments of randomly generated data vectors. The horizontal axis in the Figures 5.1 and 5.2 show the ratio of sparsity in relation to the measurement dimension - in our discussion the dimension of the measured signal $\mathbf{g}$, which is L. On the other hand, the vertical axis depicts the normalized error quadratic reconstruction error.

We observe in our simulations that the deterministic Gabor matrix generated with the Alltop window vector (dimensions $L = 29, 73$ were chosen) performs similarly good as random Gaussian matrices and Gabor matrices with random window vector. This observation aligns with the theoretical result given in Theorem 5.2.2, since the class of random matrices - in this case random Gaussian matrix and the random Gabor matrix - are known to have restricted isometry property with high probability for $L \geq ck\log(N/k)$, where we have $N = L^2$ and c is a constant. Clearly, our simulations show that the deterministic Gabor matrices generated with the Alltop window have similar behaviour as those random matrices.

5.4. Proof of Theorem 5.2.1 and Theorem 5.2.2

In order to prove Theorem 5.2.1 and Theorem 5.2.2, which is the main result of this section, we need the following lemma.

Lemma 5.4.1. *Let $L \geq 5$ be a prime number of the form $L = 3n + 2$, $n \in \mathbb{N}$, let $\boldsymbol{\alpha} \in \mathbb{C}^L$ be the Alltop window (5.1) and let $\mathbf{G}_{\boldsymbol{\alpha}}$ be the associated Gabor matrix. With $\boldsymbol{\alpha}$ we associate the $L \times L$ diagonal matrix*

$$\mathbf{S}_{\boldsymbol{\alpha}} = L\,\mathrm{diag}\Big(\overline{\boldsymbol{\alpha}[0]}, \overline{\boldsymbol{\alpha}[1]}, \ldots, \overline{\boldsymbol{\alpha}[L-1]}\Big)\,.$$

Then $\Phi := \mathbf{S}_{\boldsymbol{\alpha}}\mathbf{G}_{\boldsymbol{\alpha}} \in \mathbb{C}^{L\times L^2}$ has the following properties.

(P1) *For any $x, y \in \mathbb{Z}_L$,*

$$\sum_{\lambda\in\mathbb{Z}_L\times\mathbb{Z}_L} \boldsymbol{\phi}_\lambda[x]\,\overline{\boldsymbol{\phi}_\lambda[y]} = \begin{cases} L^3, & \text{if } x = y \\ 0, & \text{else} \end{cases}\,.$$

(P2) *For any λ_1, $\lambda_2 \in \mathbb{Z}_L \times \mathbb{Z}_L$, there exist unique $\gamma \in \mathbb{Z}_L$ and $\lambda_3 \in \mathbb{Z}_L \times \mathbb{Z}_L$ such that*

$$\boldsymbol{\phi}_{\lambda_1}[x]\,\overline{\boldsymbol{\phi}_{\lambda_2}[x]} = \omega^{\gamma}\boldsymbol{\phi}_{\lambda_3}[x] \quad \textit{for all } x \in \mathbb{Z}_L,$$

where $\omega = \mathrm{e}^{j\frac{2\pi}{L}}$. For later use, we introduce the notation $\gamma = \gamma[\lambda_1, \lambda_2]$ and $\lambda_3 = \sigma[\lambda_1, \lambda_2]$.

(P3) *For any $\lambda \in (\mathbb{Z}_L \times \mathbb{Z}_L)\backslash(0,0)$, we have*

$$\left|\sum_{x=0}^{L-1} \boldsymbol{\phi}_{\lambda}[x]\right| \leq L\,\mu\Big(\frac{1}{\sqrt{L}}\boldsymbol{\Phi}\Big) = \sqrt{L}.$$

(P4) *For all $x \in \mathbb{Z}_L$, we have*

$$\sum_{\lambda_1,\lambda_2 \in \mathbb{Z}_L \times \mathbb{Z}_L} \boldsymbol{\phi}_{\lambda_1}[x]\,\overline{\boldsymbol{\phi}_{\lambda_2}[x]} = 0. \tag{5.2}$$

Where we denote by $\boldsymbol{\phi}_{\lambda}$ the column of $\boldsymbol{\Phi}$ indexed by λ.

Note that P3 shows that $\boldsymbol{\Phi}$ satisfies S3 from Definition 3.1.1 with $\eta = 1$.
Before continuing we are going to need the following definition.

Definition 5.4.1. *Let p be a prime number and let $2 \leq g \leq p-1$ be an integer. Then g is called a primitive root if we have*

$$\left\{g^k\right\}_{k=0}^{p-2} = \{1, 2, \ldots, p-1\},$$

where all arithmetic operations are preformed mod p.

In other words a primitive root of a congruence class ($\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$) generates the congruence class apart from the zero element, i.e. $\mathbb{Z}_L \setminus \{0\} = \left\{g^k\right\}_{k=0}^{p-2}$.
Furthermore, if p is a prime number then there always exists a primitive root in $\mathbb{Z}_p$ [112].
In order to prove (P4) in Lemma 5.4.1 we are going to need the following lemma.

Lemma 5.4.2. *Let p be an odd prime number and let $r > 0$ be a positive integer that does not divide $p-1$ ($r \not| p-1$). Then we have*

$$\{1, 2, \ldots, p-1\} = \{1^r, 2^r, \ldots (p-2)^r\},$$

where all arithmetic operations are performed mod p.

For the proof of Lemma 5.4.2 we are going to need Fermat's Little Theorem [112, 113].

Lemma 5.4.3 (Fermat's Little Theorem)**.** *Let $p \in \mathbb{N}$ be a prime number then for any integer a we have,*

$$a = a^p \mod p\,.$$

Note that by dividing both sides by a Lemma 5.4.3 can equivalently be formulated as

$$1 = a^{p-1} \mod p\,,$$

which leads to a common formulation, that for a prime number p and an integer a not divisible by p, p divides $a^{p-1} - 1$ [94, 112].
The following proof of Lemma 5.4.3 can be found in [113].

Proof of Lemma 5.4.3. First, we are going to show for two integers r, k that

$$r^p + k^p = (r+k)^p \mod p\,, \tag{5.3}$$

holds. Then applying (5.3) we will use induction to prove Fermat's Little Theorem.
Using the binomial coefficient we get ,

$$(r+k)^p = \sum_{i=0}^{p} \binom{p}{i} r^{p-i}k^i = \sum_{i=0}^{p} \frac{p!}{i!\,(p-i)!} r^{p-i}k^i\,.$$

Note that for $i \neq p$ and $i \neq 0$ we have $\frac{p!}{i!(p-i)!} = 0 \mod p$, since the numerator is always a multiple of p and the denominator is a multiplication with two integer factors each smaller than p and therefore not divisible by p. Thus, we are only left with the cases where $i = 0$ and $i = p$, hence we end up with $r^p + k^p$.
Now we use induction to prove Fermat's Little Theorem. For the base step of the induction we obviously have $1^p = 1 \mod p$. For the induction step assume the claim is true for $a = r \in \mathbb{N}$. We have to show that the claim follows for $a = r + 1$. To this end, assume $r = r^p \mod p$ and observe

$$\begin{aligned} r+1 &= r^p + 1 \mod p \\ &= r^p + 1^p \mod p \\ &= (r+1)^p \mod p\,, \end{aligned}$$

where we used (5.3). □

Next we prove Lemma 5.4.2

Proof of Lemma 5.4.2. Let g be a primitive root of $\mathbb{Z}_p$, such a primitive root always exists because p is prime. Assume $1 \leq \ell, k \leq p-2$ and $\ell \neq k$ and without loss of generality $l < k$. Then assume for a contradiction that $g^{(r\cdot\ell)} = g^{(r\cdot k)}$. Using Lemma 5.4.3 we get

$$g^{r(k-\ell)} = g^{p-1} = 1\,,$$

hence we must have $r(k-\ell) = p-1$ which implies $k - \ell = \frac{p-1}{r}$. This is a contradiction since $k - \ell > 0$ is an integer and r does not divide $p-1$. □

Now we are ready to prove Lemma 5.4.1.

Proof of Lemma 5.4.1. The property (P1) follows directly from Proposition 4.2.4, observe

$$\mathbf{\Phi}\mathbf{\Phi}^* = \mathbf{S}_{\boldsymbol{\alpha}} \underbrace{\mathbf{G}_{\boldsymbol{\alpha}}\mathbf{G}_{\boldsymbol{\alpha}}^*}_{=L^2\|\boldsymbol{\alpha}\|_2^2\mathrm{I}_L} \mathbf{S}_{\boldsymbol{\alpha}}^* = L^2\mathbf{S}_{\boldsymbol{\alpha}}\mathbf{S}_{\boldsymbol{\alpha}}^* = L^3\mathrm{I}_L\,.$$

Thus, we see that the rows of $\mathbf{\Phi}$ are orthogonal to each other, which is the exact statement of (P1).
Next, for (P2) with $\lambda_1 = (\tau_1, \nu_1)$ and $\lambda_2 = (\tau_2, \nu_2)$ observe that the statement is equivalent to

$$\boldsymbol{\phi}_{\lambda_1}[x]\overline{\boldsymbol{\phi}_{\lambda_2}[x]} = \omega^{-x^3+(x-\tau_1)^3+\nu_1 x} \cdot \overline{\omega^{-x^3+(x-\tau_2)^3+\nu_2 x}}\omega^{\gamma-x^3+(x-\tau_3)^3+\nu_3 x}$$

for all $x \in \mathbb{Z}_L$. The argument in the power of ω is simply a polynomial (in $\mathbb{Z}_L[x]$), thus we can just check the coefficients to obtain the values for τ_3, ν_3 and γ, these are then given by

$$\begin{aligned}
\gamma &= -3\tau_1^2\tau_2 + 3\tau_1\tau_2^2\,,\\
\tau_3 &= \tau_1 - \tau_2\,,\\
\nu_3 &= \nu_1 - \nu_2 + 6\tau_1\tau_2 - 6\tau_2^2\,.
\end{aligned}$$

For (P3), first observe that $\boldsymbol{\phi}_{(0,0)} \in \mathbb{C}^L$ is the all ones vector, i.e. we have $\boldsymbol{\phi}_{(0,0)} = [1, 1, \ldots, 1]$. Hence, the statement of (P3) can be written in terms of

an inner product with $\boldsymbol{\phi}_{(0,0)}$,

$$\left|\left\langle \boldsymbol{\phi}_\lambda, \boldsymbol{\phi}_{(0,0)} \right\rangle\right| = \left|\boldsymbol{\phi}_{(0,0)}^* \boldsymbol{\phi}_\lambda\right| = \left|\sum_{x=0}^{L-1} \boldsymbol{\phi}_\lambda[x]\right| .$$

Now to obtain the claimed estimate observe,

$$\begin{aligned}\left|\left\langle \boldsymbol{\phi}_\lambda, \boldsymbol{\phi}_{(0,0)} \right\rangle\right| &= |\langle \mathbf{S}_{\boldsymbol{\alpha}} \pi(\lambda)\boldsymbol{\alpha}, \mathbf{S}_{\boldsymbol{\alpha}} \phi(0,0)\boldsymbol{\alpha}\rangle| = \left|\boldsymbol{\alpha}^* \underbrace{\pi(0,0)}_{=\mathrm{I}_L} \underbrace{\mathbf{S}_{\boldsymbol{\alpha}}^* \mathbf{S}_{\boldsymbol{\alpha}}}_{=L\cdot \mathrm{I}_L} \pi(\lambda)\boldsymbol{\alpha}\right| \\ &= L \cdot |\boldsymbol{\alpha}^* \pi(\lambda)\boldsymbol{\alpha}| \leq L\frac{1}{\sqrt{L}} = \sqrt{L}\,,\end{aligned}$$

where we used Proposition 5.1.1 to upper bound $|\boldsymbol{\alpha}^*\pi(\lambda)\boldsymbol{\alpha}|$.
To prove (P4) note that we have with $\lambda_1 = (\tau_1, \nu_1)$ and $\lambda_2 = (\tau_2, \nu_2)$,

$$\begin{aligned}&\sum_{\lambda_1,\lambda_2 \in \mathbb{Z}_L \times \mathbb{Z}_L} \boldsymbol{\phi}_{\lambda_1}[x]\overline{\boldsymbol{\phi}_{\lambda_2}[x]} \\ &\qquad = \sum_{(\tau_1,\nu_1),(\tau_2,\nu_2)\in \mathbb{Z}_L\times\mathbb{Z}_L} \omega^{-x^3+(x-\tau_1)^3+\nu_1 x} \cdot \omega^{x^3-(x-\tau_2)^3-\nu_2 x} \\ &\qquad = \sum_{(\tau_1,\nu_1),(\tau_2,\nu_2)\in \mathbb{Z}_L\times\mathbb{Z}_L} \omega^{3(-\tau_1+\tau_2)x^2+(3\tau_1^2-3\tau_2^2+\nu_1-\nu_2)x-\tau_1^3+\tau_2^3} \qquad (5.4) \\ &\qquad = \sum_{\tau_1\in\mathbb{Z}_L} \omega^{-3\tau_1 x^2+3\tau_1^2 x-\tau_1^3} \cdot \sum_{\nu_1\in\mathbb{Z}_L} \omega^{\nu_1 x} \\ &\qquad\quad \times \sum_{\tau_2\in\mathbb{Z}_L} \omega^{3\tau_2 x^2-3\tau_2^2 x+\tau_2^3} \cdot \sum_{\nu_2\in\mathbb{Z}_L} \omega^{\nu_2 x} . \qquad (5.5)\end{aligned}$$

Now we separate two cases.
First, if $x \neq 0$ then in (5.5) we have from Lemma 4.2.2

$$\sum_{\nu_1\in\mathbb{Z}_L} \omega^{\nu_1 x} = \sum_{\nu_2\in\mathbb{Z}_L} \omega^{\nu_2 x} = 0$$

hence (5.5) equals to zero.
For the second case, if we have $x = 0$ then (5.4) becomes

$$L^2 \cdot \sum_{\tau_1\in\mathbb{Z}_L} \omega^{-\tau_1^3} \sum_{\tau_2\in\mathbb{Z}_L} \omega^{\tau_2^3} . \qquad (5.6)$$

Since we have that L is a prime number of the form $L = 3n+2$ with $n \in \mathbb{N}$, we immediately see that $3 \not| L$ (3 does not divide L). Hence, we can apply Lemma

5.4.2 and obtain $\{0^3, 1^3, \dots (L-1)^3\} = \{0, 1, \dots L-1\}$. Therefore, the sums in (5.6) both are simply a reordering of the sum $\sum_{\tau \in \mathbb{Z}_L} \omega^\tau$. Finally, by applying Lemma 4.2.2 we see that both sums in (5.6) are equal to zero and consequently (5.4) is equal to zero. Which is the statement of P4. □

The subsequent proofs of Theorem 5.2.1 and Theorem 5.2.2 mainly follow the steps from [70]. Hence, we are going to need a version of the McDiarmid inequality [114] that is used in [70]. We will state the self avoiding McDiarmid inequality as in [70] and its proof can be found in the appendix of [70].

Lemma 5.4.4 (Theorem 11 in [70])**.** *Let $X_1, ..., X_K$ be probability spaces and define $\mathcal{X}$ as the probability space of all distinct K-tuples. In other words, the set $\mathcal{X}$ is the subset of the product set $\mathbb{X} = X_1 \times X_2 \times ... \times X_K$ given by*

$$\mathcal{X} = \left\{ (t_1, ..., t_K) \in \prod_{i=1}^{K} X_i \ : \ t_i \neq t_j \text{ whenever } i \neq j \right\},$$

where the probability measure on $\mathcal{X}$ is just the renormalization (so as to be a probability measure) of the restriction to $\mathcal{X}$ of the standard product measure on $\mathbb{X}$.
Let $h(t_1, ..., t_K)$ be a function from the set $\mathcal{X}$ to $\mathbb{C}$, such that for any coordinate i, given $t_1, ..., t_{\ell-1}$:

$$\begin{aligned} \Bigg(& \sup_{\substack{u \in X_\ell \text{ with} \\ u \neq t_i, \ i=1,...,\ell-1}} \mathbb{E}\left[h\left(t_1, ..., t_{\ell-1}, u, T_{\ell+1}, ..., T_K\right)\right] \\ - & \inf_{\substack{v \in X_\ell \text{ with} \\ v \neq t_i, \ i=1,...,\ell-1}} \mathbb{E}\left[h\left(t_1, ..., t_{\ell-1}, v, T_{\ell+1}, ..., T_K\right)\right] \Bigg) \leq c_\ell, \end{aligned} \tag{5.7}$$

where the expectations are taken over the random variables $\mathrm{T}_{\ell+1}, ..., T_K$ (conditioned on taking values that are all different from each other and from $t_1, ...t_{\ell-1}$ as well as u (first expectation) or v (second expectation). Then for any $\gamma > 0$.

$$\Pr\left[\left|h\left(T_1, ..., T_K\right) - \mathbb{E}\left[h\left(T_1, ..., T_K\right)\right]\right| \geq \gamma\right] \leq 2 \exp\left(\frac{-2\gamma^2}{\sum_{\ell=1}^{K} c_\ell^2}\right).$$

Before starting with the proofs let us describe the framework first. We assume that random vectors $\mathbf{f} \in \Sigma_k^{L^2}$ with $\|\mathbf{f}\|_2^2 = 1$ are generated as follows. First, the support set of $\mathbf{f}$ is chosen as $\mathcal{S} = \{\pi_0, \pi_2, \dots, \pi_{k-1}\}$ where $\{\pi_j\}_{j=0}^{L^2-1}$ is a random permutation of $\{0, 1, \dots, L^2\}$. Then one chooses the k entry values of $\mathbf{f}$ randomly so that $\|\mathbf{f}\|_2^2 = 1$ with no specific distribution.

Proof of Theorem 5.2.1. First, we establish the inequality from the theorem for the matrix $\mathbf{\Phi} = \mathbf{S}_{\boldsymbol{\alpha}}\mathbf{G}_{\boldsymbol{\alpha}}$ and then we use that $\|\mathbf{\Phi}\mathbf{f}\|_2^2 = L\,\|\mathbf{G}_{\boldsymbol{\alpha}}\mathbf{f}\|_2^2$ for all $\mathbf{f} \in \mathbb{C}^L$ to obtain the bounds for $\mathbf{G}_{\boldsymbol{\alpha}}$.
First, let $\mathbf{f} \in \Sigma_k^{L^2}$ be a random vector supported on $\mathcal{S} = \{\pi_0, \pi_1, \ldots, \pi_{k-1}\}$ with k nonzero entries denoted as $f_0, f_1, \ldots, f_{k-1}$, which is equivalent to say $\mathbf{f}[\pi_0] = f_0, \mathbf{f}[\pi_1] = f_1, \ldots, \mathbf{f}[\pi_{k-1}] = f_{k-1}$. Set $\mathbf{g} = \frac{1}{\sqrt{L}}\mathbf{\Phi}\mathbf{f}$. Then

$$\left(1 - \frac{k-1}{L^2-1}\right)\|\mathbf{f}\|_2^2 \leq \mathbb{E}_\pi\left[\|\mathbf{g}\|_2^2\right] \leq \left(1 + \frac{1}{L^2-1}\right)\|\mathbf{f}\|_2^2\,, \tag{5.8}$$

wherein $\mathbb{E}_\pi$ stands for the expectation over all possible choices π.
To see this, one writes the k-sparse vector $\mathbf{f}$ as $\mathbf{f} = \sum_{j=0}^{k-1} f_j \mathbf{e}_{\pi_j}$. This yields $\mathbf{g} = \frac{1}{\sqrt{L}}\sum_{j=0}^{k-1} f_j \boldsymbol{\phi}_{\pi_j}$ and

$$\|\mathbf{g}\|_2^2 = \sum_{x=0}^{L-1}|\mathbf{g}[x]|^2 = \frac{1}{L}\sum_{x=0}^{L-1}\left(\sum_{j=0}^{k-1}|f_j|^2 + \sum_{\substack{i,j\\ i\neq j}} f_i\overline{f_j}\boldsymbol{\phi}_{\pi_i}[x]\,\overline{\boldsymbol{\phi}_{\pi_j}[x]}\right) \tag{5.9}$$

$$= \|\mathbf{f}\|_2^2 + \frac{1}{L}\sum_{\substack{i,j\\ i\neq j}} f_i\overline{f_j}\left\langle\boldsymbol{\phi}_{\pi_i}, \boldsymbol{\phi}_{\pi_j}\right\rangle\,, \tag{5.10}$$

where we used that $|\boldsymbol{\phi}_i[x]| = 1$ for all $x \in \mathbb{Z}_L$ and i. By Property (P3) of Lemma 5.4.1, we have for $i \neq j$

$$\begin{aligned}\mathbb{E}_\pi\left[\left\langle\boldsymbol{\phi}_{\pi_i}, \boldsymbol{\phi}_{\pi_j}\right\rangle\right] &= \frac{1}{L^2(L^2-1)}\sum_{\substack{\lambda_1,\lambda_2\in\mathbb{Z}_\times\mathbb{Z}_L\\ \lambda_1\neq\lambda 2}}\left\langle\boldsymbol{\phi}_{\lambda_1}, \boldsymbol{\phi}_{\lambda_2}\right\rangle\\ &= \frac{1}{L^2(L^2-1)}\sum_{x=0}^{L-1}\left(-L^2\right) = -\frac{L}{L^2-1}\,,\end{aligned}$$

and therefore

$$\mathbb{E}_\pi\left[\|\mathbf{g}\|_2^2\right] = \|\mathbf{f}\|_2^2 - \frac{1}{L^2-1}\sum_{\substack{0\leq i,j\leq k-1\\ i\neq j}} f_i\overline{f_j}\,. \tag{5.11}$$

Then using the Cauchy-Schwarz inequality, we obtain

$$0 \leq \left|\sum_{j=0}^{k-1} f_j\right|^2 = \sum_{j=0}^{k-1}|f_j|^2 + \sum_{\substack{0\leq i,j\leq k-1\\ i\neq j}} f_i\overline{f_j} \leq k\sum_{j=0}^{k-1}|f_j|^2\,,$$

which is equivalent to

$$-(k-1)\left\|\mathbf{f}\right\|_2^2 \leq -\sum_{\substack{0\leq i,j\leq k-1\\ i\neq j}} f_i\overline{f_j} \leq \left\|\mathbf{f}\right\|_2^2 . \tag{5.12}$$

Finally, combining (5.11) and (5.12) we obtain (5.8).
Next, we are going to show that the random variable $\left\|\mathbf{g}\right\|_2^2$ is concentrated around its mean. The main idea is to upper bound $\Pr_\pi\left(\left|\left\|\mathbf{g}\right\|_2^2 - \left\|\mathbf{f}\right\|_2^2\right| \geq \delta\left\|\mathbf{f}\right\|_2^2\right)$, and we already know from (5.8) that $\mathbb{E}_\pi\left[\left\|\mathbf{g}\right\|_2^2\right]$ is close to $\left\|\mathbf{f}\right\|_2^2$. Recalling the dependence of $\mathbf{g}$ on $\pi_0,\ldots,\pi_{k-1}$ as shown in (5.9), we write $h(\pi_0,\ldots,\pi_{k-1}) = \left\|\mathbf{g}\right\|_2^2$. For $\beta > 0$ we consider

$$\Pr_\pi\left(\left|\left\|\mathbf{g}\right\|_2^2 - \mathbb{E}_\pi\left[\left\|\mathbf{g}\right\|_2^2\right]\right| \geq \beta\left\|\mathbf{f}\right\|_2^2\right) = \Pr_\pi\left(\left|h - \mathbb{E}\left[h\right]\right| \geq \beta\left\|\mathbf{f}\right\|_2^2\right) .$$

Our next goal is to establish the condition given in (5.7) for h. To this end we have

$$\begin{aligned}
&h(\pi_0,...,\pi_\ell,...,\pi_{k-1}) - h(\pi_0,...,\pi'_\ell,...,\pi_{k-1})\\
&= \frac{1}{L}\sum_{\substack{0\leq j\leq k-1\\ j\neq\ell}}\left(f_\ell\,\overline{f_j}\,\left[\boldsymbol{\phi}_{\pi_\ell} - \boldsymbol{\phi}_{\pi_{\ell'}}\right]^T\overline{\boldsymbol{\phi}_{\pi_j}}\right)\\
&\quad + \frac{1}{L}\sum_{\substack{0\leq j\leq k-1\\ j\neq\ell}}\left(f_j\,\overline{f_\ell}\,\boldsymbol{\phi}_{\pi_j}^T\overline{\left[\boldsymbol{\phi}_{\pi_\ell} - \boldsymbol{\phi}_{\pi_{\ell'}}\right]}\right)\\
&= \frac{1}{L}\sum_{\substack{0\leq j\leq k-1\\ j\neq\ell}}\left(f_\ell\overline{f_j}\sum_{x=0}^{L-1}\omega^{\gamma[\pi_\ell,\pi_j]}\boldsymbol{\phi}_{\sigma[\pi_\ell,\pi_j]}[x]\right)\\
&\quad - \frac{1}{L}\sum_{\substack{0\leq j\leq k-1\\ j\neq\ell}}\left(f_\ell\overline{f_j}\sum_{x=0}^{L-1}\omega^{\gamma[\pi_{\ell'},\pi_j]}\boldsymbol{\phi}_{\sigma[\pi_{\ell'},\pi_j]}[x]\right)\\
&\quad + \frac{1}{L}\sum_{\substack{0\leq j\leq k-1\\ j\neq\ell}}\left(f_j\overline{f_\ell}\sum_{x=0}^{L-1}\omega^{\gamma[\pi_\ell,\pi_j]}\boldsymbol{\phi}_{\sigma[\pi_\ell,\pi_j]}[x]\right)\\
&\quad - \frac{1}{L}\sum_{\substack{0\leq j\leq k-1\\ j\neq\ell}}\left(f_j\overline{f_\ell}\sum_{x=0}^{L-1}\omega^{\gamma[\pi'_\ell,\pi_j]}\boldsymbol{\phi}_{\sigma[\pi'_\ell,\pi_j]}[x]\right),
\end{aligned}$$

where we have used (P2) from Lemma 5.4.1 to write $\boldsymbol{\phi}_{\pi_j}\overline{\boldsymbol{\phi}_{\pi_j}} = \omega^{\gamma[\pi_\ell,\pi_j]}\boldsymbol{\phi}_{\sigma[\pi_\ell,\pi_j]}$.

Using Lemma 5.4.1, we now estimate

$$\begin{aligned}&|h(\pi_0,\dots,\pi_\ell,\dots,\pi_{k-1}) - h(\pi_0,\dots,\pi'_\ell,\dots,\pi_{k-1})| \\ &\quad\leq \frac{4}{L}\,|f_\ell| \sum_{\substack{0\leq j\leq k-1\\ j\neq\ell}} |f_j|\,\sqrt{L} = \frac{4}{\sqrt{L}}\,|f_\ell| \sum_{\substack{0\leq j\leq k-1\\ j\neq\ell}} |f_j| = c_\ell\,.\end{aligned}$$

Applying Lemma 5.4.4 [70], one obtains

$$\begin{aligned}&\Pr_\pi\left(|h-\mathbb{E}\,[h]| \geq \beta\,\|\mathbf{f}\|_2^2\right) \\ &\qquad\leq 2\exp\left(-\frac{\beta^2 L\,\|\mathbf{f}\|_2^4}{8\sum_{\ell=0}^{k-1}|f_\ell|\left(\sum_{\substack{0\leq j\leq k-1\\ j\neq\ell}}|f_j|\right)^2}\right).\end{aligned}$$

Together with the estimate

$$\sum_{\ell=0}^{k-1}|f_\ell|^2\left(\sum_{\substack{0\leq j\leq k-1\\ j\neq\ell}}|f_\ell|\right)^2 \leq \sum_{\ell=0}^{k-1}|f_\ell|^2\cdot\left(\sum_{\ell=0}^{k-1}|f_\ell|\right)^2 \leq k\,\|\mathbf{f}\|_2^4\,,$$

where we used the Cauchy-Schwarz inequality, we obtain

$$\Pr_\pi\left(|h-\mathbb{E}\,[h]| \geq \beta\,\|\mathbf{f}\|_2^2\right) \leq 2\exp\left(-\frac{\beta^2 L}{8k}\right).$$

Finally, substituting $\|\mathbf{g}\|_2^2$ for h and using (5.8) we obtain

$$\Pr_\pi\left(\left|\|\mathbf{g}\|_2^2 - \|\mathbf{f}\|_2^2\right| \geq \left(\beta+\frac{k-1}{L^2-1}\right)\|\mathbf{f}\|_2^2\right) \leq 2\exp\left(-\frac{\beta^2 L}{8k}\right).$$

For $\delta > \frac{k-1}{L^2-1}$, we set $\beta = \delta - \frac{k-1}{L^2-1}$ and $\epsilon = 2\exp\left(-\left(\delta-\frac{k-1}{L^2-1}\right)^2\frac{L}{8k}\right)$. Then, using that

$$\|\mathbf{g}\|_2^2 = \left\|\frac{1}{L}\mathbf{\Phi}\mathbf{f}\right\|_2^2 = \|\mathbf{G}_{\boldsymbol{\alpha}}\mathbf{f}\|_2^2\,,$$

one obtains

$$\Pr_\pi\left(\left|\|\mathbf{G}_{\boldsymbol{\alpha}}\mathbf{f}\|_2^2 - \|\mathbf{f}\|_2^2\right| \geq \delta\,\|\mathbf{f}\|_2^2\right) \leq \epsilon\,.$$

This show that $\mathbf{G}_{\boldsymbol{\alpha}}$ is a near-isometry for k-sparse vectors with probability at least $1-\epsilon$ and proves the statement of the Theorem 5.2.1. □

Now we continue with the proof of Theorem 5.2.2.

Proof of Theorem 5.2.2. From Theorem 5.2.1, $\mathbf{G}_{\boldsymbol{\alpha}}$ has (k,δ,ϵ)-StRIP and therefore it has also $(k,\delta,2\epsilon)$-StRIP. To prove that $\mathbf{G}_{\boldsymbol{\alpha}}$ has $(k,\delta,2\epsilon)$-UStRIP, it suffices to show that the probability of having $\mathbf{G}_{\boldsymbol{\alpha}}\mathbf{f} = \mathbf{G}_{\boldsymbol{\alpha}}\mathbf{f}'$ with distinct k-sparse vectors $\mathbf{f} \neq \mathbf{f}'$ is less than ϵ.
In the following the proof is separated into eight smaller steps in order to make it more accessible.

1) Fix any $r \in \{0,\dots,L^2-1\}$ and let $\mathcal{S} = \{\pi_0,\dots,\pi_{k-1}\}$ be the set of the first k elements in a random permutation of the L^2-1 elements $\{0,\dots,L^2-1\}\backslash\{r\}$. First we show that

$$\mathbb{E}_{\mathcal{S}}\left[\|\Phi_{\mathcal{S}}^*\boldsymbol{\phi}_r\|_2^2\right] = \frac{kL^2}{L+1}, \tag{5.13}$$

where the expectation is taken with respect to the choice of $\mathcal{S}$, $\boldsymbol{\Phi}_{\mathcal{S}}$ denotes the $L \times k$ submatrix of $\boldsymbol{\Phi}$ where the columns are indexed by the elements in $\mathcal{S}$. This equations follows from

$$\mathbb{E}_{\mathcal{S}}\left[\|\Phi_{\mathcal{S}}^*\,\boldsymbol{\phi}_r\|_2^2\right] = \sum_{i=0}^{k-1}\mathbb{E}_{\mathcal{S}}\left[|\langle\boldsymbol{\phi}_r,\boldsymbol{\phi}_{\pi_i}\rangle|^2\right] = k\,\mathbb{E}_{\mathcal{S}}\left[|\langle\boldsymbol{\phi}_r,\boldsymbol{\phi}_{\pi_1}\rangle|^2\right]$$

and by using Properties (P2) and (P3) of Φ.

2) Next, we show that with probability exceeding $1-\epsilon$, any random subset $\mathcal{S} \subset \{0,\dots,L^2-1\}$ of size k and any $r \in \mathcal{S}^c$ satisfy

$$\|\boldsymbol{\Phi}_{\mathcal{S}}^*\boldsymbol{\phi}_r\|^2 \leq k\,L + L\sqrt{2k\log\left(L^2/\epsilon\right)}. \tag{5.14}$$

To see this, let $y(t_0,\dots,t_{k-1}) = \sum_{i=0}^{k-1}\left|\boldsymbol{\phi}_{t_i}^*\boldsymbol{\phi}_r\right|^2$ where $t_0,\dots,t_{k-1}$ are k distinct elements chosen randomly from $\{0,\dots,L^2-1\}\setminus\{r\}$, with r fixed. For $t_i \neq t_i'$, we have

$$\begin{aligned}&\mathbb{E}\left[y(t_1,\dots,t_i,\dots,t_k)\right] - \mathbb{E}\left[y(t_1,\dots,t_i',\dots,t_k)\right]\\ &= \left||\boldsymbol{\phi}_i^*\boldsymbol{\phi}_r|^2 - |\boldsymbol{\phi}_{t_i'}^*\boldsymbol{\phi}_r|^2\right|\end{aligned}$$

$$= \left| \left| \sum_{x=0}^{L-1} \boldsymbol{\phi}_{\gamma(r,t_i)}(x) \right|^2 - \left| \sum_{x=0}^{L-1} \boldsymbol{\phi}_{\gamma(r,t_i')}(x) \right|^2 \right| \leq 2L,$$

using (P1) and (P2) together with the fact that $\gamma(t_i, r) \neq 0$ and $\gamma(t_i', r) \neq 0$, since t_i, t_i' are distinct from r. Similarly as in the proof of Theorem 5.2.1, we apply the McDiarmid inequality which yields

$$\begin{aligned} \Pr\left(\|\boldsymbol{\Phi}_{\mathcal{S}}^* \boldsymbol{\phi}_r\|^2 \geq kL + \xi \right) &\leq \Pr\left(\|\boldsymbol{\Phi}_{\mathcal{S}}^* \boldsymbol{\phi}_r\|^2 \geq \frac{kL^2}{L+1} + \xi \right) \\ &\leq \exp\left(-\frac{\xi^2}{2kL^2} \right). \end{aligned}$$

Taking the union bound over all L^2 choices of r yields (5.14).

3) When $L = \mathrm{O}(\frac{k \log L}{\delta^2})$, (5.14) becomes

$$\begin{aligned} &\left\| \frac{1}{\sqrt{L}} \boldsymbol{\Phi}_{\mathcal{S}}^* \frac{1}{\sqrt{L}} \boldsymbol{\phi}_w \right\|_2^2 \leq \frac{k}{L} + \frac{1}{L} \sqrt{\frac{2k \log L^2}{\epsilon}} \\ &= \mathrm{O}\left(\frac{\delta^2}{\log L} \right) + \mathrm{O}\left(\delta^2 \left(1 + \frac{|\log \delta|}{\log L} \right)^{1/2} \cdot \frac{1}{\sqrt{k \log L}} \right), \end{aligned}$$

which indicates a small coherence between the columns of $\boldsymbol{\Phi}_{\mathcal{S}}$ and the remaining columns.

4) Since $\mathbf{G}_{\boldsymbol{\alpha}}$ has (k, δ, ϵ)-*StRIP* (see Theorem 5.2.1), one can deduce the following statement, as in [70, Lemma 19]:

Let $\mathcal{S} = \{\pi_0, \ldots, \pi_{k-1}\}$ be the set of the first k elements in a random permutation of $\{0, \ldots, L^2\}$. With probability exceeding $1 - \epsilon$, any subset $\Lambda \subset \{0, \ldots, L^2 - 1\}$ of size k with $\Lambda \neq \mathcal{S}$ satisfies

$$\dim\left(\mathrm{range}\,(\mathbf{G}_{\boldsymbol{\alpha}})_{\Lambda} \cap \mathrm{range}\,(\mathbf{G}_{\boldsymbol{\alpha}})_{\mathcal{S}} \right) < k. \tag{5.15}$$

5) We consider $f \in \mathbb{C}^{L^2}$ as a random vector supported on $\mathcal{S} = \{\pi_0, \ldots, \pi_{k-1}\}$. The k nonzero entries are chosen randomly from a k-dimensional vectors space equipped with a measure which is absolutely continuous with respect to the Lebesgue measure.

6) Using Theorems B and D in [115] and the fact that $\mathbf{G}_{\boldsymbol{\alpha}}$ has (k, δ, ϵ)-

StRIP, it follows that with probability exceeding $1-\epsilon$, $(\mathbf{G}_{\boldsymbol{\alpha}})_{\mathcal{S}}$ is injective, i.e, $\dim(\text{range}(\mathbf{\Phi}_{\mathcal{S}})) = k$. This means that with probability exceeding $1-\epsilon$, no two signals supported on $\mathcal{S}$ can have the same measurement.

7) If there exists $\mathbf{f}'$ supported on $\Lambda \neq \mathcal{S}$ with $\mathbf{G}_{\boldsymbol{\alpha}}\mathbf{f}' = \mathbf{G}_{\boldsymbol{\alpha}}\mathbf{f}$, then (5.15) implies that with probability exceeding $1-\epsilon$, the vector $\mathbf{f}$ restricted to $\mathcal{S}$ lies in at most $(k-1)$-dimensional subspace of $\mathbb{C}^k$, which is of zero measure with respect to any absolutely continuous measure on $\mathbb{C}^k$. That is, with probability larger $1-\epsilon$, the set of vectors $\mathbf{f}$ satisfying $\mathbf{G}_{\boldsymbol{\alpha}}\mathbf{f}' = \mathbf{G}_{\boldsymbol{\alpha}}\mathbf{f}$ for some $\mathbf{f}'$ supported on $\Lambda \neq \mathcal{S}$ is a measure zero set.

8) Combining the above two statements, we conclude that with probability exceeding $1-2\epsilon$ (with respect to random choice of $\mathbf{f}$), $\mathbf{f}$ is the only k-sparse vector satisfying the equation $\mathbf{y} = \mathbf{G}_{\boldsymbol{\alpha}}\mathbf{f}$. □

6. Low Coherence Finite Gabor Frames

The theory of Gabor frames is important in communications since one is able to model the channel effects on the transmitted signal in terms of time-frequency shifts which directly leads to Gabor frames [8, 90, 108]. In wireless communications an often analysed case is the multipath propagation in which a wireless signal arrives at the receiver via multiple paths [9]. Hence, the receiver receives a superposition of time-frequency shifted copies of the transmit signal, i.e. the received signal is distorted. In order to reduce the distortion, one seeks to find signal sequences that have low correlation with time-frequency shifted copies of their own. In fact, we already have discussed such a sequence in Section 5.1, there the finite Gabor frame generated with the Alltop window vector had low coherence, which is equivalent to saying that the discrete periodic sequence - generated by repeating the entries in the Alltop vector - has low correlation. Indeed, since this problem is an everyday occurrence in wireless communications there is a fair amount of literature devoted on finding low correlation sequences, or equivalently, low coherence finite Gabor frames [90, 91, 116]. Additionally, finite Gabor frames have been considered as sensing matrices in Compressed Sensing [86, 88, 105] due to their occurrence in communications.

In this section we will discuss a new family of finite Gabor frames that has low coherence. We will analyse the constructed finite Gabor frames with regard to their suitability as sensing matrices in compressed sensing. We will prove that these Gabor matrices exhibit the statistical restricted isometry property and the strong coherence property.
We will first introduce some preliminary topics that are necessary to introduce the results of this section.
The contents of this section has already been partially published in the paper [29].

6.1. Introduction to finite fields

We start by giving a short introduction to finite fields. Let p be a prime number and $n \in \mathbb{N}$ then we denote the finite field of order p^n by $\mathbb{F}_{p^n}$. Formally, finite fields are defined as follows.

Definition 6.1.1. *A field $(F, +, \cdot)$ is a set F equipped with two operations usually called addition "$+$" and multiplication "$\cdot$", satisfying the following conditions*

- $a \cdot b \in F$ and $a + b \in F$, for all $a, b \in F$
- $a \cdot b = b \cdot a$ and $a + b = b + a$, for all $a, b \in F$
- $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ and $a + (b + c) = (a + b) + c$, for all $a, b, c \in F$
- $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(a + b) \cdot c = a \cdot c + b \cdot c$ for all $a, b, c \in F$
- $\exists 0 \in F$ s.t. $0 + a = a$ and $\exists 1 \in F$ s.t. $1 \cdot a = a$, for all $a \in F$
- $\forall a \in F \ \exists - a$ s.t. $a + (-a) = 0$ and $\forall a \in F \ \exists a^{-1}$ s.t. $a \cdot a^{-1} = 1$

In general if $|F| = q$ then this field is denoted by $\mathbb{F}_q$.

For example the finite field $\mathbb{F}_7$ is given by the integers modulo 7, $\mathbb{Z}_7 := \{0, 1, ..., 6\}$ with the usual operations addition and multiplication on natural numbers $\mod 7$. In engineering it is also common to denote a finite field of order q, $\mathbb{F}_q$ by $\mathrm{GF}(q)$ and call it a Galois field [117]. Let p be a prime number and $n \in \mathbb{N}$ then the n-dimensional vector space over the finite field $\mathbb{F}_p$, denoted by $\mathbb{F}_p^n$, forms also a finite field. Furthermore, the finite field $\mathbb{F}_p$ is referred to as a prime field [94]. We will denote the finite field $\mathbb{F}_p^n$ by $\mathbb{F}_{p^n}$. We have the following result characterizing the order of finite fields.

Proposition 6.1.1. *All finite fields have order p^n with p a prime number and n a natural number, i.e. $n \in \mathbb{N}$.*

Proof. The proposition is an immediate consequence of Theorem 1.78 and Theorem 1.84 in [94]. □

Furthermore, we call the finite field $\mathbb{F}_{p^n}$ a finite extension of the primitive field $\mathbb{F}_p$. The characteristic of a finite field characterizes how often an element of the field can be added to itself until one obtains zero. Let p be a prime number for a primitive field $\mathbb{F}_p$ the characteristic is p and for a field which is an extension of a primitive field $\mathbb{F}_{p^n}$ the characteristic is the same as the characteristic of the

vector	polynomial	powers of x
00	0	$x^{-\infty}$
10	1	x^0
01	x	x^1
12	$1+2x$	x^2
22	$2+2x$	x^3
20	2	x^4
02	$2x$	x^5
21	$2+x$	x^6
11	$1+x$	x^7

Table 6.1.: Elements of the finite field $\mathbb{F}_{3^2}$ generated by the irreducible polynomial $r(x) = x^2 + x + 2$.

base primitive field, in this case p. Note that Proposition 6.1.1 together with Theorem 1.78 in [94] implies that a finite field is either a primitive field or a field extension of a primitive field. Furthermore, we have that the characteristic of a finite field is always a prime number p [94].

For the construction of finite fields the so called irreducible polynomials are needed. These are polynomials with coefficients form $\mathbb{F}_p$ and with degree n. In short one can state that a polynomial over a field is irreducible if it is not the product of two polynomials of lower degree in this field. Assume an irreducible polynomial $r(x)$ is given over $\mathbb{F}_{p^n}$ then every element in $\mathbb{F}_{p^n}$ can be represented with calculations modulo $r(x)$.

In Table 6.1 an example of a finite field $\mathbb{F}_{3^2}$ is given. The finite field is constructed with the irreducible polynomial $r(x) = x^2 + x + 2$. Note that $r(x)$ has no root in $\mathbb{F}_3$, i.e. it is irreducible over $\mathbb{F}_3$. However, $r(x)$ has two roots in $\mathbb{F}_{3^2}$. Furthermore, as indicated in the Table 6.1 the finite field $\mathbb{F}_{3^2}$ consist of two groups. One group is the additive group and the other one is the multiplicative group.

Indeed, this is true for all finite fields $\mathbb{F}_{p^n}$. Additionally, the multiplicative group of a finite field is cyclic [94]. Hence, there always exists a generator element of the multiplicative group of the finite field. In the example given in Table 6.1 this generator element is denoted by x and is generally referred as the primitive element of $\mathbb{F}_{p^n}$. In fact, we already introduced the notion of primitive elements for prime fields already in Definition 5.4.1, since the congruence class $\mathbb{Z}_p$ is equivalent to the prime field $\mathbb{F}_p$, $\mathbb{Z}_p \cong \mathbb{F}_p$. We denote the multiplicative group of the finite field of order p^n by $\mathbb{F}^*_{p^n}$, i.e. $\mathbb{F}_{p^n} = \mathbb{F}_{p^n} \setminus \{0\}$ where 0 is the additive neutral element.

We now define the trace map in finite fields which is a linear mapping from $\mathbb{F}_{p^n}$ to $\mathbb{F}_p$.

Definition 6.1.2. *The trace map* $\mathrm{tr} : \mathbb{F}_{p^n} \to \mathbb{F}_p$ *is given by*

$$\mathrm{tr}(\alpha) = \alpha + \alpha^p + \cdots + \alpha^{p^{n-1}} = \sum_{i=0}^{n-1} \alpha^{p^i} ,$$

with $\alpha \in \mathbb{F}_{p^n}$.

Further note that the trace $\mathrm{tr}(\alpha)$ with $\alpha \in \mathbb{F}_{p^n}$ will always be an element of $\mathbb{F}_p$, $\mathrm{tr}(\alpha) \in \mathbb{F}_p$. In general $\mathbb{F}_p$ in Definition 6.1.2 does not necessarily need to be a prime field, but for our purposes we can always assume that $\mathbb{F}_p$ to be prime field.
Next we will discuss some properties of the trace map.

Theorem 6.1.2 ([94])**.** *The trace function defined in Definition 6.1.2 satisfies the following properties:*

(i) $\mathrm{tr}(\alpha + \beta) = \mathrm{tr}(\alpha) + \mathrm{tr}(\beta)$ *for all* $\alpha, \beta \in \mathbb{F}_{p^n}$*;*

(ii) $\mathrm{tr}(c\alpha) = c \cdot \mathrm{tr}(\alpha)$ *for all* $c \in \mathbb{F}_p$ *and* $\alpha \in \mathbb{F}_{p^n}$*;*

(iii) tr *is a linear transformation from* $\mathbb{F}_{p^n}$ *onto* $\mathbb{F}_p$*, where both* $\mathbb{F}_{p^n}$ *and* $\mathbb{F}_p$ *are viewed as vector spaces over* $\mathbb{F}_p$*;*

(iv) $\mathrm{tr}(\alpha) = n \cdot \alpha$ *for all* $\alpha \in \mathbb{F}_p$*;*

(v) $\mathrm{tr}(\alpha^p) = \mathrm{tr}(\alpha)$ *for all* $\alpha \in \mathbb{F}_{p^n}$.

Proof. See proof of Theorem 2.23 in [94]. □

The following theorem from [94] characterizes the kernel of the trace map.

Theorem 6.1.3 ([94])**.** *Let the trace map* $\mathrm{tr} : \mathbb{F}_{p^n} \to \mathbb{F}_p$ *be defined as in Definition 6.1.2. Then for* $\alpha \in \mathbb{F}_{p^n}$ *we have* $\mathrm{tr}(\alpha) = 0$ *if and only if* $\alpha = \beta^p - \beta$ *for some* $\beta \in \mathbb{F}_{p^n}$.

Proof. See proof of Theorem 2.25 in [94]. □

6.2. Characters and Gaussian sums

We consider a character as a mapping from an Abelian group to the unit circle, where the unit circle T is defined by $T = \{x \in \mathbb{C} : |x| = 1\}$.

Definition 6.2.1. *Let $(G, \bullet)$ be an Abelian group, then a character is a mapping $\chi : G \to T$ that preserves the group structure under multiplication. Specifically, we have*

$$\chi(a \bullet b) = \chi(a)\chi(b)\,, \quad \textit{for all } a, b \in G.$$

Further, the trivial character χ_t is defined as the mapping that maps all elements of G to 1, i.e. $\chi_t(a) = 1$ for all $a \in G$.

As previously mentioned, finite fields consist of two different groups an additive group and a multiplicative group. Therefore, we differentiate between two different characters on finite fields, the additive and the multiplicative characters.
The additive characters $\{\chi_a : a \in \mathbb{F}_{p^n}\}$ in $\mathbb{F}_{p^n}$ are given by

$$\chi_a(b) = \exp\left\{j\tfrac{2\pi}{p}\operatorname{tr}(ab)\right\}\,, \quad b \in \mathbb{F}_{p^n}. \tag{6.1}$$

As the name suggest the multiplication of an additive character with two different arguments is equivalent to the addition of the arguments. Let $a, b, c \in \mathbb{F}_{p^n}$ then we have

$$\begin{aligned}\chi_a(b)\,\chi_a(c) &= \exp\left\{j\tfrac{2\pi}{p}\operatorname{tr}(ab)\right\}\cdot\exp\left\{j\tfrac{2\pi}{p}\operatorname{tr}(ac)\right\}\\ &= \exp\left\{j\tfrac{2\pi}{p}\operatorname{tr}(ab)+\operatorname{tr}(ac)\right\}\\ &= \exp\left\{j\tfrac{2\pi}{p}\operatorname{tr}\left(a\left(b+c\right)\right)\right\}\\ &= \chi_a(b+c)\,,\end{aligned}$$

where we used the linearity of the trace map from Theorem 6.1.2.
Let α be a primitive element of $\mathbb{F}_{p^n}$. Then $\mathbb{F}_{p^n}^* = \{\alpha^0, \alpha^1, ..., \alpha^{p^n-2}\}$, and the multiplicative characters $\{\psi_{\alpha^l} : l = 0, 1, \ldots, p^n - 2\}$ in $\mathbb{F}_{p^n}$ can be written as

$$\psi_{\alpha^l}(\alpha^k) = \exp\left\{j\frac{2\pi}{p^n-1}kl\right\}\,, \quad k \in \{0, 1, ..., p^n - 2\}. \tag{6.2}$$

Analogously to the additive character the multiplication of an multiplicative character with two different arguments is equivalent to the multiplication of

the arguments. Let $l, k, r \in \{0, 1, \ldots, p^n - 2\}$ then we have

$$\begin{aligned}\psi_{\alpha^l}\left(\alpha^k\right)\psi_{\alpha^l}\left(\alpha^r\right) &= \exp\left\{j\frac{2\pi}{p^n-1}kl\right\}\cdot\exp\left\{j\frac{2\pi}{p^n-1}kr\right\}\\ &= \exp\left\{j\frac{2\pi}{p^n-1}k(l+r)\right\}\\ &= \psi_{\alpha^l}\left(\alpha^{(k+r)}\right)\\ &= \psi_{\alpha^l}\left(\alpha^k\cdot\alpha^r\right).\end{aligned}$$

We denote the trivial multiplicative character by ψ_1, i.e. $\psi_1(a) = 1$ for all $a \in \mathbb{F}_{p^n}^*$ and the trivial additive character by χ_0, i.e. $\chi_0(a) = 1$ for all $a \in \mathbb{F}_{p^n}$.

Next we continue with Gaussian sums which will be essential in our analysis later on.
Let ψ be a multiplicative character and χ be an additive character of $\mathbb{F}_{p^n}$. Then

$$G(\psi, \chi) = \sum_{a \in \mathbb{F}_{p^n}^*} \psi(a)\,\chi(a) \tag{6.3}$$

is said to be a Gaussian sum. Note that the sum in (6.3) runs over the multiplicative group of the finite field, i.e. $\mathbb{F}_{p^n}^*$. The following theorem will be of importance later and can be found with its proof in [94, Chapter 5].

Theorem 6.2.1. *Let ψ be a multiplicative and χ an additive character of $\mathbb{F}_{p^n}$. Then the Gaussian sum $G(\psi, \chi)$ satisfies*

$$G(\psi, \chi) = \begin{cases} p^n - 1 & \text{if } \psi = \psi_1,\ \chi = \chi_0, \\ -1 & \text{if } \psi = \psi_1,\ \chi \neq \chi_0, \\ 0 & \text{if } \psi \neq \psi_1,\ \chi = \chi_0. \end{cases}$$

If $\psi \neq \psi_1$ and $\chi \neq \chi_0$, then

$$|G(\psi, \chi)| = \sqrt{p^n}.$$

It is possible to derive multiple useful identities of Gaussian sums under various transformations of the additive or multiplicative characters. This is discussed in the next theorem.

Theorem 6.2.2. *Gaussian sums for the finite field $\mathbb{F}_{q^n}$ satisfy the following properties:*

(i) $G(\psi, \chi_{ab}) = \overline{\psi(a)} G(\psi, \chi_b)$ *for* $a \in \mathbb{F}_{p^n}^*$ *and* $b \in \mathbb{F}_{p^n}$*;*

(ii) $G(\psi, \overline{\chi}) = \psi(-1) G(\psi, \chi)$*;*

(iii) $G\left(\overline{\psi}, \chi\right) = \psi(-1) \overline{G(\psi, \chi)}$*;*

(iv) $G(\psi, \chi) G\left(\overline{\psi}, \chi\right) = \psi(-1) q$ *for* $\psi \neq \psi_1$, $\chi \neq \chi_0$*;*

(v) $G(\psi^p, \chi_b) = G\left(\psi, \chi_{\sigma(b)}\right)$ *for* $b \in \mathbb{F}_{p^n}$*, where* p *is the characteristic of* $\mathbb{F}_{p^n}$ *and* $\sigma(b) = b^p$*.*

Proof. See proof of Theorem 5.12 in [94]. □

6.3. Cyclic difference sets

Gaussian sums will play a key role in constructing and analysing low coherence finite frames in Section 6.5. A subclass of this family of low coherence Gabor frames can be derived using cyclic difference sets [118, 119]. Thus, in this section we introduce the notion of cyclic difference sets and their properties. We begin by defining the cyclic difference sets.

Definition 6.3.1. *Let* $\mathbb{Z}_L$ *be the integers* $\bmod L$*. Let* k *and* l *be positive integers such that* $2 \leq k < L$*. Then an* (L, k, l)*-cyclic difference set in* $\mathbb{Z}_L$ *is a subset* $\mathcal{D} \subset \mathbb{Z}_L$ *that satisfies the following properties:*

(i) $|\mathcal{D}| = k$*,*

(ii) *the multiset* $\{x - y : x, y \in \mathcal{D}, x \neq y\}$ *contains every element in* $\mathbb{Z}_L \setminus \{0\}$ *exactly* l *times.*

In Definition 6.3.1 the condition that we consider $\mathbb{Z}_L$ is not necessary. In fact there exist difference sets for groups that are not even Abelian [118]. However in the cases we are interested in the groups considered will be cyclic. In fact, the groups will be equivalent to an n-dimensional vector space over $\mathbb{Z}_p$ with p is a prime number and $n \in \mathbb{N}$. Out of convenience we will also refer to cyclic difference sets as simply difference sets.
Further, note that $l(L-1) = k(k-1)$ holds if a (L, k, l)-difference set exists. Before continuing let us first give an example, a $(7, 3, 1)$-difference set in $\mathbb{Z}_7$ (i.e. integers modulo 7) is given by the set

$$\mathcal{D} = \{0, 1, 3\} .$$

To see that $\mathcal{D}$ is indeed a difference set we compute the differences

$$\begin{aligned} 0-1=6, \quad 1-0=1, \quad 3-0=3, \\ 0-3=4, \quad 1-3=5, \quad 3-1=2, \end{aligned}$$

note that all computations are done $\mod 7$ since we work with integers modulo 7. Indeed, we see that all elements of $\mathbb{Z}_7$ apart from 0 are obtained exactly once from mutual differences of the elements in $\mathcal{D}$.
We can clearly see that $a+\mathcal{D}=\{a+x : x \in \mathcal{D}\}$ remains a difference set for all $a \in \mathbb{Z}_7$.
Let $\mathcal{D}$ be a (L,k,l)-difference set then its complement, i.e. the set $\mathcal{D}^c = \mathbb{Z}_L \setminus \mathcal{D}$, also forms a difference set with parameters $(L, L-k, L-2k+l)$ [119]. Now continuing with our previous example of $\mathcal{D}=\{0,1,3\}$, we have for $\mathcal{D}^c=\{2,4,5,6\}$ and since $\mathcal{D}$ is a $(7,3,1)$-difference set we have that $\mathcal{D}^c$ is a $(7,4,2)$-difference set. Further, by computing the mutual differences in $\mathcal{D}^c$ we get

$$\begin{aligned} 2-4=5, \quad 2-5=4, \quad 2-6=3, \\ 4-2=2, \quad 4-5=6, \quad 4-6=5, \\ 5-2=3, \quad 5-4=1, \quad 5-6=6, \\ 6-2=4, \quad 6-4=2, \quad 6-5=1, \end{aligned}$$

which verifies that $\mathcal{D}^c$ is indeed a $(7,4,2)$-difference set.
Next we look at combinatorial aspects of the sets $a+\mathcal{D}$ with $a \in \mathbb{Z}_L$. More precisely we want to study the cardinality of the intersection of the sets $\mathcal{D}$ with $a+\mathcal{D}$, i.e. the number $|\mathcal{D} \cap a+\mathcal{D}|$. Although, this property is not difficult to study it will play a crucial role in the upcoming discussion of low coherence finite Gabor frames.

Proposition 6.3.1. *Let $\mathcal{D}$ be a (v,k,l)-difference set in $\mathbb{Z}_L$, then*

$$|\mathcal{D} \cap a+\mathcal{D}| = l\,,$$

for all $a \in \mathbb{Z}_L \setminus \{0\}$.

Proof. Our aim is to determine how many pairs x and y exists with $x \in \mathcal{D}$ and $y \in a+\mathcal{D}$ such that $x=y$, this equality can equivalently be reformulated as $x-y=0$. Now set $y'=y-a$ thus we get $x-y'+a=0$. Notice that $y' \in \mathcal{D}$, hence observing that the multiset

$$\{x-y'+a : x \in \mathcal{D} \text{ and } y' \in \mathcal{D}\}$$

contains the 0 element exactly l times implies the claim. □

In fact, a difference set $\mathcal{D}$ of $\mathbb{Z}_L$ gives rise to a symmetric (L, k, l)-balanced incomplete block design (BIBD) [118, 119]. However, we don't need the notion of BIBDs here and Proposition 6.3.1 will be enough for us.
A direct consequence of Proposition 6.3.1 is that we have

$$|(b + \mathcal{D}) \cap (a + \mathcal{D})| = l\,,$$

for $a, b \in \mathbb{Z}_L$ and $a \neq b$.
For a better understanding let us visualize the statement of Proposition 6.3.1 using the $(7, 3, 1)$-difference set $\mathcal{D} = \{0, 1, 3\}$ in $\mathbb{Z}_7$ from the previous example:

$$\begin{aligned}
0 + \mathcal{D} &= \{0, 1, 3\},\\
1 + \mathcal{D} &= \{1, 2, 4\},\\
2 + \mathcal{D} &= \{2, 3, 5\},\\
3 + \mathcal{D} &= \{3, 4, 6\},\\
4 + \mathcal{D} &= \{0, 4, 5\},\\
5 + \mathcal{D} &= \{1, 5, 6\},\\
6 + \mathcal{D} &= \{0, 6, 2\}.
\end{aligned}$$

As we see any two sets intersect in exactly 1 element.

Let $\mathcal{D}$ be a (L, k, l)-difference set for $\mathbb{Z}_L$, we define the vector $\mathbf{v}_\mathcal{D} \in \mathbb{C}^L$ as follows

$$\mathbf{v}_\mathcal{D}[i] = \begin{cases} 1, & \text{if } i \in \mathcal{D},\\ 0, & \text{else .} \end{cases} \tag{6.4}$$

Another important aspect of cyclic difference sets is given in connection with the columns of the discrete Fourier transformation matrix.

Proposition 6.3.2. *Let $\mathcal{D}$ be a (L, k, l)-difference set for $\mathbb{Z}_L$ and the vector $\mathbf{v}_\mathcal{D} \in \mathbb{C}^L$ defined as in (6.4), and denote by $\mathbf{f}_r$ the r-th column of the discrete Fourier transformation matrix in $\mathbb{C}^L$. Then*

$$|\langle \mathbf{v}_\mathcal{D}, \mathbf{f}_r \rangle| = \sqrt{k - l}\,,$$

for $r = 1, \ldots L - 1$.

Proof. Set $\omega = \mathrm{e}^{j\frac{2\pi}{L}}$ then by straight forward computation we obtain

$$|\langle \mathbf{v}_\mathcal{D}, \mathbf{f}_r \rangle|^2 = |\mathbf{f}_r^* \mathbf{v}_\mathcal{D}|^2 = \left| \sum_{a \in \mathcal{D}} \omega^{r \cdot a} \right|^2 = \sum_{a,b \in \mathcal{D}} \omega^{ra} \omega^{-rb}$$

$$= \sum_{a,b\in\mathcal{D}} \omega^{r(a-b)} = k + \sum_{\substack{a,b\in\mathcal{D}\\ a\neq b}} \omega^{r(a-b)} = k + l \cdot \sum_{a=1}^{L-1} w^{ra}$$
$$= k - l\,,$$

where we applied Lemma 4.2.2 in the last step. □

A well known class of difference sets are the Singer difference sets [120] whose parameters are given in the following theorem.

Theorem 6.3.3 ([118, 119])**.** *Let $q = p^m$ be a prime power (i.e. p is a prime number and $m \in \mathbb{N}$) and $n \geq 2$ an integer. Then there exists a $\left(\frac{q^{n+1}-1}{q-1}, \frac{q^n-1}{q-1}, \frac{q^{n-1}-1}{q-1}\right)$ -difference set in $\mathbb{Z}_{(q^{n+1}-1)/(q-1)}$.*

In particular we are interested in the Singer difference sets with parameter $q = 2$, i.e. $(2^{n+1}-1, 2^n-1, 2^{n-1}-1)$. For $n = 2$ we have that $\mathcal{D} = \{1, 2, 4\}$ in $\mathbb{Z}_L$ is a Singer difference set. These Singer difference sets can be constructed via the trace map. Now we are going to shortly discuss the construction of these difference sets as given in [119].
Let α be a primitive element of $\mathbb{F}_{2^{n+1}}$ with $n \geq 2$, then using the trace map we obtain the Singer difference set $\mathcal{D} \subset \mathbb{Z}_{2^{n+1}-1}$ as follows:

$$\mathcal{D} = \left\{b \,:\, \mathrm{tr}\left(\alpha^b\right) = 0\right\}. \tag{6.5}$$

If a difference set $\mathcal{D}$ is obtained as described in (6.5) then a shift of the difference set can be obtained using the trace map. For instance let $\alpha^a \in \mathbb{F}^*_{2^{n+1}}$ then the difference set $\mathcal{D} + a$ is given by

$$\mathcal{D} + a = \left\{b \,:\, \mathrm{tr}\left(\alpha^{b-a}\right) = 0\right\}.$$

6.4. Construction of low coherence finite Gabor frames based on difference sets

Let us fix a Singer difference set $\mathcal{D}$ with parameters $(2^{n+1}-1, 2^n-1, 2^{n-1}-1)$ and let $\mathbf{v}_\mathcal{D}$ be defined as in (6.4). Additionally, let $\mathcal{D}^c$ denote the complementary set of $\mathcal{D}$, i.e. $\mathcal{D}^c = \mathbb{Z}_{2^{n+1}-1} \setminus \mathcal{D}$, and with $\mathbf{v}_{\mathcal{D}^c}$ defined analogous to $\mathbf{v}_\mathcal{D}$. Next we define the window vector $\mathbf{v}$ as the difference between $\mathbf{v}_\mathcal{D}$ and $\mathbf{v}_{\mathcal{D}^c}$, $\mathbf{v} = \mathbf{v}_\mathcal{D} - \mathbf{v}_{\mathcal{D}^c}$,

more precisely

$$\mathbf{v}[k] = \begin{cases} 1, & \text{if } k \in \mathcal{D}, \\ -1, & \text{if } k \in \mathcal{D}^c. \end{cases} \tag{6.6}$$

Now we are ready to state the main result of this subsection.

Theorem 6.4.1. *Let $n \geq 2$ and $\mathcal{D}$ be a Singer difference set with parameters $(2^{n+1}-1, 2^n-1, 2^{n-1}-1)$. Further, set $\phi = \frac{1}{\sqrt{2^{n+1}-1}}\mathbf{v} \in \mathbb{C}^{2^{n+1}-1}$ with $\mathbf{v}$ as defined in* (6.6) *and denote by $\mathbf{G}_\phi$ the Gabor matrix generated by ϕ. Then the for the coherence of the Gabor matrix we have $\mu(\mathbf{G}_\phi) = \frac{\sqrt{2^{n+1}}}{2^{n+1}-1}$.*
Explicitly, for the absolute values of the inner products of the finite Gabor frame generated by the window vector ϕ we have

$$|\langle \phi, \mathbf{M}^\nu \mathbf{T}^\tau \phi \rangle| = \begin{cases} 1, & \text{if } \tau = 0,\ \nu = 0, \\ 0, & \text{if } \tau = 0,\ \nu \neq 0, \\ \frac{1}{2^{n+1}-1}, & \text{if } \tau \neq 0,\ \nu = 0, \\ \frac{\sqrt{2^{n+1}}}{2^{n+1}-1}, & \text{if } \tau \neq 0,\ \nu \neq 0. \end{cases}$$

Moreover, let $\mathbf{F}$ be the discrete Fourier transform matrix in $\mathbb{C}^{2^{n+1}-1}$ then we have

$$\frac{1}{\sqrt{2^{n+1}-1}} |\langle \mathbf{f}_k, \mathbf{M}^\nu \mathbf{T}^\tau \phi \rangle| = \begin{cases} \frac{1}{2^{n+1}-1}, & \text{if } k = \nu, \\ \frac{\sqrt{2^{n+1}}}{2^{n+1}-1}, & \text{if } k \neq \nu, \end{cases} \tag{6.7}$$

where $\mathbf{f}_k$ denotes the k-th column of $\mathbf{F}$. Note that we use the factor $\frac{1}{\sqrt{2^{n+1}-1}}$ to ℓ_2-normalize the columns of the discrete Fourier transform matrix.

Before we prove the theorem, let us remark that the coherence of the Gabor matrix in Theorem 6.4.1 is very close to the best possible coherence given by the Welch bound in Theorem 2.4.3. Since the finite Gabor frame $\{\pi(\lambda)\phi\}_{\lambda \in \mathbb{Z}_{2^{n+1}-1} \times \mathbb{Z}_{2^{n+1}-1}}$ has $(2^{n+1}-1)^2$ elements the lower bound for the coherence derived from the Welch bound in (2.1) for the parameters of the Gabor frame is $\frac{1}{\sqrt{2^{n+1}}}$. The coherence of the Gabor matrix constructed in Theorem 6.4.1 is $\frac{\sqrt{2^{n+1}}}{2^{n+1}-1}$, hence the construction given in Theorem 6.4.1 is very close to the best possible coherence achievable.
Furthermore, the sensing matrix obtained from concatenating the Gabor matrix of Theorem 6.4.1 with the discrete Fourier transform matrix, $[\mathbf{G}_\mathbf{v}|\mathbf{F}]$, fulfils the statistical restricted isometry property and the strong coherence property.

This will be proven in a later subsection.
We now continue with the proof of Theorem 6.4.1.

Proof of Theorem 6.4.1. We start by verifying (6.7). The case $\nu = r$ is trivial, we continue with the case $r \neq \nu$.
Let $\omega = \mathrm{e}^{j\frac{2\pi}{2^{n+1}-1}}$, since $\mathcal{D}$ is a $(2^{n+1}-1, 2^n-1, 2^{n-1}-1)$ difference set we have from Proposition 6.3.2

$$|\langle \mathbf{v}_{\mathcal{D}}, \mathbf{f}_r \rangle| = \left| \sum_{a \in \mathcal{D}} \omega^{-r \cdot a} \right| = \sqrt{2^{n-1}}\,.$$

Further, we have that $\mathcal{D}^c$ is a $(2^{n+1}-1, 2^n, 2^{n-1})$ difference set. Thus we have

$$|\langle \mathbf{v}_{\mathcal{D}^c}, \mathbf{f}_r \rangle| = \left| \sum_{a \in \mathcal{D}^c} \omega^{-r \cdot a} \right| = \sqrt{2^{n-1}}.$$

In fact, simply by using the definition of $\mathcal{D}^c$ and Lemma 4.2.2 we conclude

$$\sum_{a \in \mathcal{D}} \omega^{-r \cdot a} = -\sum_{a \in \mathcal{D}^c} \omega^{-r \cdot a}\,.$$

Now note that $\mathbf{v} = \mathbf{v}_{\mathcal{D}} - \mathbf{v}_{\mathcal{D}^c}$, hence we have

$$\begin{aligned} |\langle \mathbf{v}, \mathbf{f}_r \rangle| &= |\langle \mathbf{v}_{\mathcal{D}}, \mathbf{f}_r \rangle - \langle \mathbf{v}_{\mathcal{D}^c}, \mathbf{f}_r \rangle| \\ &= \left| \sum_{a \in \mathcal{D}} \omega^{-r \cdot a} - \sum_{a \in \mathcal{D}^c} \omega^{-r \cdot a} \right| = \left| 2 \cdot \sum_{a \in \mathcal{D}} \omega^{-r \cdot a} \right| = 2\sqrt{2^n}\,. \end{aligned}$$

Finally, observing

$$|\langle \mathbf{M}^{\nu} \mathbf{T}^{\tau} \mathbf{v}, \mathbf{f}_r \rangle| = |\langle \mathbf{v}, \mathbf{f}_{r+\nu} \rangle|\,,$$

where we used that the columns of the discrete Fourier transform matrix form eigenvectors of the translation operators. Now together with $\phi = \frac{1}{\sqrt{2^{n+1}-1}}\mathbf{v}$ delivers the claimed assertion.

In the next step we want to analyse $\langle \mathbf{v}, \mathbf{T}^{\tau} \mathbf{v} \rangle$ for $\tau \neq 0$.
Since the support of $\mathbf{v}$ can be divided into two disjoint sets as $\mathcal{D}$ and $\mathcal{D}^c$, the multiplication with the translation operator $\mathbf{T}^{\tau}$ has the consequence that the vector $\mathbf{T}^{\tau}\mathbf{v}$ can now be splitted into the disjoint sets $\mathcal{D} + \tau$ and $\mathcal{D}^c + \tau$. To determine the value of $\langle \mathbf{v}, \mathbf{T}^{\tau} \mathbf{v} \rangle$ we just need to look at the sizes of four different intersections of sets, these are:

- $|\mathcal{D} \cap \mathcal{D} + \tau|$,
- $|\mathcal{D}^c \cap \mathcal{D}^c + \tau|$,
- $|\mathcal{D}^c \cap \mathcal{D} + \tau|$,
- $|\mathcal{D} \cap \mathcal{D}^c + \tau|$.

Since $\mathcal{D}$ and $\mathcal{D}^c$ both are difference sets we have from Proposition 6.3.1 the values

$$\begin{aligned} |\mathcal{D} \cap \mathcal{D} + \tau| &= 2^{n-1} - 1\,, \\ |\mathcal{D}^c \cap \mathcal{D}^c + \tau| &= 2^{n-1}\,, \end{aligned}$$

and due to symmetry we get for the two remaining cases

$$|\mathcal{D} \cap \mathcal{D}^c + \tau| = |\mathcal{D}^c \cap \mathcal{D} + \tau| = 2^{n-1}\,.$$

Next looking at the multiplications

$$\mathbf{v}\,[k] \cdot (\mathbf{T}^\tau \mathbf{v})\,[k] = \begin{cases} 1, & \text{if } \; k \in \{\mathcal{D} \cap \mathcal{D} + \tau\} \cup \{\mathcal{D}^c \cap \mathcal{D}^c + \tau\}\;, \\ -1, & \text{if } \; k \in \{\mathcal{D}^c \cap \mathcal{D} + \tau\} \cup \{\mathcal{D} \cap \mathcal{D}^c + \tau\}\;, \end{cases}$$

we see that the value of $\langle \mathbf{v}, \mathbf{T}^\tau \mathbf{v} \rangle$ is simply determined by the sizes of these four sets. Hence, we have

$$\langle \mathbf{v}, \mathbf{T}^\tau \mathbf{v} \rangle = -1 \quad \text{for } \; \tau \neq 0\,.$$

Finally, ℓ_2-normalization of $\mathbf{v}$ delivers the claim of the theorem.

The only relation left to show now is

$$|\langle \mathbf{v}, \mathbf{M}^\nu \mathbf{T}^\tau \mathbf{v} \rangle| = \sqrt{2^{n+1}} \quad \text{for } \; \tau \neq 0\,, \nu \neq 0\,.$$

Following the construction of difference sets given in (6.5) we can write the entries of $\mathbf{v}$ as $\mathbf{v}\,[k] = (-1)^{\mathrm{tr}\left(\alpha^k\right)}$ with α a primitive element of $\mathbb{F}_{2^{n+1}}$. For a $\tau \neq 0$ set $\alpha^{-\gamma} = 1 + \alpha^{-\tau}$. Note that $1 + \alpha^{-\tau} \neq 0$ since the additive inverse of 1 in $\mathbb{F}_{2^{n+1}}$ is 1 and $\alpha^{-\tau}$ cannot be equal to 1 since $\tau \neq 0$. Now by straight

forward computation we get

$$\begin{aligned}
|\langle \mathbf{v}, \mathbf{M}^{\nu}\mathbf{T}^{\tau}\mathbf{v}\rangle| &= \left| \sum_{k=0}^{2^{n+1}-2} \omega^{-k\nu} (-1)^{\mathrm{tr}\left(\alpha^{k}\right)} (-1)^{\mathrm{tr}\left(\alpha^{k-\tau}\right)} \right| \\
&= \left| \sum_{k=0}^{2^{n+1}-2} \omega^{-k\nu} (-1)^{\mathrm{tr}\left(\left(1+\alpha^{-\tau}\right)\alpha^{k}\right)} \right| \\
&= \left| \sum_{k=0}^{2^{n+1}-2} \omega^{-k\nu} (-1)^{\mathrm{tr}\left(\alpha^{-\gamma}\alpha^{k}\right)} \right| \\
&= \left| \sum_{k\in\mathcal{D}+\gamma} \omega^{-k\nu} - \sum_{k\in(\mathcal{D}+\gamma)^{c}} \omega^{-k\nu} \right| \\
&= 2 \cdot \left| \sum_{k\in\mathcal{D}+\gamma} \omega^{-k\nu} \right| \\
&= \sqrt{2^{n+1}}\,,
\end{aligned}$$

where we used Proposition 6.3.2 and that $\mathcal{D}+\gamma$ forms a difference set. Normalizing the vector $\mathbf{v}$ we obtain the statement of the theorem which finishes our proof. □

Our proof of Theorem 6.4.1 intimately relies on the fact that we are able to write the entries of $\mathbf{v}_{\mathcal{D}}$ and $\mathbf{v}_{\mathcal{D}^c}$ as powers of -1 which depended on the trace map in finite fields. This type of a connection between the entries of $\mathbf{v}_{\mathcal{D}}$ and $\mathbf{v}_{\mathcal{D}^c}$ is not possible if $\mathcal{D}$ is not a Singer difference set with parameters $(2^{n+1}-1, 2^n-1, 2^{n-1}-1)$. Hence, the suggested construction fails for general difference sets $\mathcal{D}$. However, in the next subsection we will see that the construction discussed here is indeed a special case of a larger family of low coherence finite Gabor frames.

6.5. Low coherence finite Gabor frames based on Gaussian sums

In this section our aim is to generalize the construction given in the previous section. In the last section we had the observation in Proposition 6.3.2 and made a good guess to construct the window vector $\mathbf{v}$. We constructed the window vector by disjointly indexing the entries either by 1 or by -1 depending on whether it was indexed by a certain cyclic difference set $\mathcal{D}$ or by its

complementary set $\mathcal{D}^c$ which is also a difference set. This construction however does not work for other cyclic difference sets. In fact, the proof of Theorem 6.4.1 relied on the fact that the elements of $\mathbf{v}$ could be written by means of the trace map in the finite fields.
These observations raises the question whether the frame work of cyclic difference sets is the correct one to analyse the low coherence Gabor frames constructed in the previous section. Indeed, as we will see Gaussian sums over finite fields are more suited for the analysis, and the construction in the previous section turns out to be a special case of a more general construction.

Let us begin by first inspecting the result from the previous section. There the vector $\mathbf{v} \in \mathbb{C}^L$ (with $L = 2^{n+1} - 1$) was disjointly indexed by $\mathcal{D}$ and $\mathcal{D}^c$ with entries 1 or -1. In the proof of Theorem 6.4.1 we made the observation that each entry of $\mathbf{v}$ can be written as

$$\mathbf{v}\,[k] = (-1)^{\mathrm{tr}\left(\alpha^k\right)}\ ,$$

for $k = 0, \ldots, L-1$ and α is a primitive element of the finite field $\mathbb{F}_{2^{n+1}}$. Now comparing it with the additive character of $\mathbb{F}_{2^{n+1}}$ given in (6.1) we see that

$$(-1)^{\mathrm{tr}\left(\alpha^k\right)} = \chi_1\left(\alpha^k\right) = \exp\left\{j\frac{2\pi}{2}\mathrm{tr}\left(\alpha^k\right)\right\}\ .$$

Let $\omega = \mathrm{e}^{j\frac{2\pi}{L}}$ and observe

$$\begin{aligned}|\langle \mathbf{v}, \mathbf{f}_r\rangle| &= \left|\sum_{k=0}^{L-1} \chi_1\left(\alpha^k\right)\omega^{rk}\right| = \left|\sum_{k=0}^{L-1} \chi_1\left(\alpha^k\right)\psi_{\alpha^r}\left(\alpha^k\right)\right| \\ &= \left|\sum_{a\in\mathbb{F}^*_{2^{n+1}}} \chi_1(a)\,\psi_{\alpha^r}(a)\right| = |G\left(\psi_{\alpha^r}, \chi_1\right)| = \sqrt{2^{n+1}}\,,\end{aligned}$$

where we used the notion of multiplicative characters in (6.2) and Gaussian sums in (6.3). A natural question to ask now is whether this construction generalizes.
Indeed, a generalization is possible: In the following, we provide a construction for Gabor frames in $\mathbb{C}^L$ with $L = p^n - 1$ where p is a prime number and $n \in \mathbb{N}$. Let α be a primitive element of $\mathbb{F}_{p^n}$ and define the vector $\phi \in \mathbb{C}^L$ by

$$\boldsymbol{\phi}\,[k] = \frac{1}{\sqrt{L}}\chi_1\left(\alpha^k\right)\ ,\ \text{with } k \in \{0, \ldots, p^n - 2\}\,. \tag{6.8}$$

For convenience we introduce the column-wise ℓ_2-normalized discrete Fourier transform matrix $\mathbf{W} = \frac{1}{\sqrt{L}}\mathbf{F}$ with columns $\mathbf{w}_k = \frac{1}{\sqrt{L}}\mathbf{f}_k$.

Theorem 6.5.1. *Let $\{\pi(\lambda)\phi\}_{\lambda\in\mathbb{Z}_L\times\mathbb{Z}_L}$ be the full Gabor system generated by the window vector* (6.8) *and let $\mathbf{W} \in \mathbb{C}^{L\times L}$ be the ℓ_2-normalized discrete Fourier matrix. Then*

$$|\langle \boldsymbol{\phi}, \mathbf{M}^\nu \mathbf{T}^\tau \boldsymbol{\phi}\rangle| = \begin{cases} 1 & \text{if } \tau = 0\ ,\ \nu = 0 \\ 0 & \text{if } \tau = 0\ ,\ \nu \neq 0 \\ \frac{1}{L} & \text{if } \tau \neq 0\ ,\ \nu = 0 \\ \frac{\sqrt{L+1}}{L} & \text{if } \tau \neq 0\ ,\ \nu \neq 0 \end{cases} \tag{6.9}$$

and

$$|\langle \mathbf{w}_k, \mathbf{M}^\nu \mathbf{T}^\tau \boldsymbol{\phi}\rangle| = \begin{cases} \frac{1}{L} & \text{if } k = -\nu \\ \frac{\sqrt{L+1}}{L} & \text{if } k \neq \nu \end{cases}\ .$$

So in particular $\mu(\mathbf{G}_{\mathbf{\Phi}}) = \frac{\sqrt{L+1}}{L}$. Moreover, also for the concatenation $\mathbf{\Phi} := [\mathbf{G}_{\boldsymbol{\phi}} \,|\, \mathbf{W}]$ holds $\mu(\mathbf{\Phi}) = \frac{\sqrt{L+1}}{L}$.

Proof. Let α be a primitive element of $\mathbb{F}_{p^n}$, let $\boldsymbol{\phi} \in \mathbb{C}^L$ be constructed as in (6.8) with the chosen α, and recall that $L = p^n - 1$. Then for arbitrary $\nu, \tau \in \mathbb{Z}_L$, we get

$$\begin{aligned} \langle \boldsymbol{\phi}, \mathbf{M}^\nu \mathbf{T}^\tau \boldsymbol{\phi}\rangle &= \sum_{x=0}^{L-1} \overline{\boldsymbol{\phi}[x]} \boldsymbol{\phi}[x-\tau]\ \mathrm{e}^{j\frac{2\pi}{p^n-1}\nu x} \\ &= \frac{1}{L} \sum_{x=0}^{L-1} \overline{\chi_1(\alpha^x)}\, \chi_1(\alpha^{x-\tau})\, \psi_{\alpha^\nu}(\alpha^x)\ . \end{aligned} \tag{6.10}$$

Setting for the moment $\omega_0 = \exp\left\{j\frac{2\pi}{p}\right\}$, and using the linearity of the trace map, we get from the definition of the additive characters

$$\begin{aligned} \overline{\chi_1(\alpha^x)}\, \chi_1(\alpha^{x-\tau}) &= \omega_0^{-\mathrm{tr}(\alpha^x)} \omega_0^{\mathrm{tr}(\alpha^{x-\tau})} = \omega_0^{\mathrm{tr}(\alpha^{x-\tau}-\alpha^x)} \\ &= \omega_0^{\mathrm{tr}([\alpha^{-\tau}-1]\alpha^x)} = \chi_{\alpha^{-\tau}-1}(\alpha^x) \end{aligned}$$

Inserting this relation into (6.10) and remembering the definition of the Gauss sums in (6.3) yields

$$\langle \boldsymbol{\phi}, \mathbf{M}^\nu \mathbf{T}^\tau \boldsymbol{\phi}\rangle = \frac{1}{L} \sum_{x=0}^{m-1} \chi_{\alpha^{-\tau}-1}(\alpha^x)\, \psi_{\alpha^\nu}(\alpha^x)$$

$$= \frac{1}{L} G\left(\psi_{\alpha^\nu}, \chi_{\alpha^{-\tau}-1}\right) \ .$$

Then (6.9) follows by applying Theorem 6.2.1.
To prove the second part, notice that $\mathbf{w}_k$ can be written as $\mathbf{w}_k = \mathbf{M}^{-k}\mathbf{w}_0$. Then for arbitrary $\nu, \tau \in \mathbb{Z}_L$ follows

$$\left|\left\langle \mathbf{w}_k, \mathbf{M}^{-\nu}\mathbf{T}^\tau \boldsymbol{\phi}\right\rangle\right| = \left|\tfrac{1}{L}\mathbf{G}\left(\psi_1, \chi_1\right)\right| = \frac{1}{L} \ ,$$

and for $k \neq -\nu$ and $\tau \in \mathbb{Z}_L$ we get

$$|\langle \mathbf{f}_k, \mathbf{M}^\nu \mathbf{T}^\tau \boldsymbol{\phi}\rangle| = \left|\frac{1}{L}\mathbf{G}\left(\psi_{\alpha^{\nu+k}}, \chi_1\right)\right| = \frac{\sqrt{L+1}}{L},$$

which proves the second statement and finishes the proof. □

A full Gabor system in $\mathbb{C}^L$ has L^2 vectors. Thus the Welch bound for such a system is equal to $1/\sqrt{L+1}$. Theorem 6.5.1 shows now that the coherence $\mu(\mathbf{G}_\phi)$ of the Gabor system generated by vector (6.8) is very close to this lower bound, especially for large dimensions L. Moreover, joining the L columns of the Fourier matrix to this system does not increase the coherence. So the coherence of this system of $L^2 + L$ vectors is even closer to the corresponding Welch bound. Moreover, we note that this system of $L^2 + L$ vectors has a very remarkable structure. Indeed from the above proof we can deduce two additional symmetries of the constructed frames described in the following corollary.

Corollary 6.5.2. *The frame* $[\mathbf{G}_\phi|\mathbf{W}]$ *can be written as*

1. *a union of* $L+1$ *orthogonal bases, namely*

$$\mathbf{W} \quad \textit{and} \quad \{\mathbf{M}^\nu \mathbf{T}^\tau \boldsymbol{\phi}\}_{\nu=0}^{L-1} \ , \quad \tau = 0, 1, \ldots, L-1 \ .$$

2. *the union of* L *sets of equiangular lines in* $\mathbb{C}^L$

$$\{\mathbf{M}^\nu \mathbf{T}^\tau \boldsymbol{\phi}\}_{\tau=0}^{L-1} \cup \mathbf{w}_{-\nu} \ , \qquad \nu = 0, 1, \ldots, L-1 \ ,$$

each of which is of size $L+1$ *and achieves the Welch bound with equality.*

Remark 6.5.1. *Note also that the canonical basis* $\{\mathbf{e}_i\}_{i=0}^{L-1}$ *of* $\mathbb{C}^L$ *is unbiased to all the orthogonal bases described in Corollary 6.5.2, i.e.* $|\langle \mathbf{r}, \mathbf{e}_i\rangle| = \frac{1}{\sqrt{L}}$ *for all* $\mathbf{r} \in \left\{\{\pi_\lambda \boldsymbol{\phi}\}_{\lambda \in \mathbb{Z}_L \times \mathbb{Z}_L} \cup \{\mathbf{w}_i\}_{i \in \mathbb{Z}_L}\right\}$ *and for every* $i = 0, \ldots, L-1$.

Joining the Fourier basis $\mathbf{W}$ to the Gabor frame $\mathbf{G}_{\boldsymbol{\phi}}$ seems to be very natural. Because this way, we are able to partition the whole frame $[\mathbf{G}_{\boldsymbol{\phi}}|\mathbf{W}]$ into equiangular sets as described by Part 2) of Corollary 6.5.2. Moreover the Fourier basis completes naturally the group structure of the vectors in the frame under pointwise multiplication. This will become of importance in the next section. If, on the other hand, the canonical basis would be joined to $\mathbf{G}_{\boldsymbol{\phi}}$ then the nice property of partitioning the frame into equiangular line sets would not be possible. Moreover, the canonical basis would also change the number of absolute values of the inner products within the frame.

6.6. Statistical sparse signal recovery guarantees

In this section we will verify that the Gabor systems constructed in Section 6.5 are good measurement matrices for compressed sensing, in the sense that these matrices satisfy the statistical restricted isometry property (Section 3.1) and the strong coherence property (Section 3.2.

Statistical restricted isometry property

The following proposition shows that the Gabor frames considered in Theorem 6.5.1 fulfill Conditions (S1)–(S3) of Definition 3.1.2, i.e. that they are η-StRIP-able.

Proposition 6.6.1. *Let $\mathbf{G}_{\boldsymbol{\phi}} := \{\pi(\lambda)\boldsymbol{\phi}\}_{\lambda\in\mathbb{Z}_L\times\mathbb{Z}_L}$ be the full Gabor system generated by the window vector $\boldsymbol{\phi}$ given in* (6.8)*, and consider the concatenation $\mathbf{\Phi} = \sqrt{L}\,[\mathbf{\Pi}_{\boldsymbol{\phi}}\,|\,\mathbf{W}]$. Then $\mathbf{\Phi}$ fulfills (S1)–(S3) of Definition 3.1.2.*

Proof. (S1) Since finite Gabor systems are tight frames (Proposition 4.2.4) and because the discrete Fourier transform matrix is orthogonal, the rows of their concatenation $\mathbf{\Phi}$ are orthogonal.
For the second condition in (S1), set $\omega = \exp\{j\frac{2\pi}{L}\}$, let χ_1 be the additive character of $\mathbb{F}_{p^n}$ as defined in (6.1), and let α be a primitive element of $\mathbb{F}_{p^n}$. Observe for the x-th row

$$\sum_{\nu=0}^{L-1}\sum_{\tau=0}^{L-1}\omega^{x\nu}\chi_1(\alpha^{x-\tau}) + \sum_{j=0}^{L-1}\omega^{xj} = \sum_{\tau=0}^{L-1}\chi_1(\alpha^{x-\tau})\sum_{\nu=0}^{L-1}\omega^{x\nu} + \sum_{j=0}^{L-1}\omega^{xj} = 0$$

where we used $G(\psi_1,\chi) = -1$ for $\chi \neq \chi_0$ from Theorem 6.2.1.

(S2): First observe that pointwise multiplication of columns of the discrete Fourier transformation matrix is equivalent to $L \cdot \mathrm{diag}(\mathbf{w}_i)\mathbf{w}_j = \sqrt{L} \cdot \mathbf{w}_k$ with $k = i + j$.
For the Gabor part, we need to prove that for any choice of (ν_1, τ_1) and (ν_2, τ_2) there exists a pair (ν_3, τ_3) such that

$$\omega^{x\nu_1}\chi_1(\alpha^{x-\tau_1}) \cdot \omega^{x\nu_2}\chi_1(\alpha^{x-\tau_2}) = \omega^{x\nu_3}\chi_1(\alpha^{x-\tau_3}). \tag{6.11}$$

First consider $\alpha^{-\tau_1} + \alpha^{-\tau_2} \neq 0$ then (6.11) can be written as

$$\begin{aligned}&\omega^{x\nu_1}\chi_1(\alpha^{x-\tau_1}) \cdot \omega^{x\nu_2}\chi_1(\alpha^{x-\tau_2}) \\ &\quad = \omega^{x(\nu_1+\nu_2)}\chi_1(\alpha^x(\alpha^{-\tau_1} + \alpha^{-\tau_2})) = \omega^{x\nu_3}\chi_1(\alpha^{x-\tau_3})\end{aligned}$$

and a τ_3 exists based on our assumption. Now consider $\alpha^{-\tau_1} + \alpha^{-\tau_2} = 0$, then (6.11) becomes

$$\omega^{x\nu_1}\chi_1(\alpha^{x-\tau_1})\omega^{x\nu_2}\chi_1(\alpha^{x-\tau_2}) = \omega^{x(\nu_1+\nu_2)}\chi_1(0) = \omega^{x(\nu_1+\nu_2)}$$

which is just the description of the $\nu_1 + \nu_2 = \nu_3$-th column of the discrete Fourier transformation matrix.

(S3): The third condition can be stated by means of an inner product in our case i.e.

$$L^2 \left|\langle \mathbf{w}_0, \mathbf{M}^\nu \mathbf{T}^\tau \boldsymbol{\phi}\rangle\right|^2 = \mu(\boldsymbol{\Phi})^2 = L + 1 \leq L^{2-\eta}.$$

By assuming equality, we deduce $\eta = 2 - \frac{\log(L+1)}{\log(L)}$. □

Strong coherence property

The following proposition verifies that the Gabor system, constructed in Section 6.5, has the strong coherence property as described in Definition 3.2.1.

Proposition 6.6.2. *Let* $\mathbf{G}_\phi := \{\pi(\lambda)\boldsymbol{\phi}\}_{\lambda \in \mathbb{Z}_L \times \mathbb{Z}_L}$ *be the full Gabor system generated by the window vector* $\boldsymbol{\phi}$ *given in* (6.8). *Consider the concatenation* $\boldsymbol{\Phi} = [\mathbf{G}_\phi \,|\, \mathbf{W}]$ *then* $\boldsymbol{\Phi}$ *fulfills the strong coherence property for sufficiently large* L.

Proof. Plugging in the known values for μ and N, one can rewrite Condition 1) from (3.2) as $164\mu\log(N) \sim \frac{\log(L)}{\sqrt{L}} \leq 1$. This shows that there exists an m_0 such that the first condition in (3.2) is satisfied for all $L \geq m_0$.

To prove 2) in (3.2), we are going to show that for any pair (k, l) always

$$\sum_{\nu,\tau} \left\langle \mathbf{M}^k \mathbf{T}^l \boldsymbol{\phi}, \mathbf{M}^\nu \mathbf{T}^\tau \boldsymbol{\phi} \right\rangle + \sum_i \left\langle \mathbf{M}^k \mathbf{T}^l \boldsymbol{\phi}, \mathbf{w}_i \right\rangle = 0$$

holds, and that for any i one has $1 + \sum_{\nu,\tau} \langle \mathbf{w}_i, \mathbf{M}^\nu \mathbf{T}^\tau \boldsymbol{\phi} \rangle = 0$. This would imply that the sum in (3.1) is equal to 1. To this end, we recall first that

$$\begin{aligned} \left\langle \mathbf{M}^k \mathbf{T}^l \boldsymbol{\phi}, \mathbf{M}^\nu \mathbf{T}^\tau \boldsymbol{\phi} \right\rangle &= \omega^{l(\nu-k)} \left\langle \boldsymbol{\phi}, \mathbf{M}^{\nu-k} \mathbf{T}^{\tau-l} \boldsymbol{\phi} \right\rangle \\ &= \frac{1}{L} \psi_{\alpha^{\nu-k}} \left(\alpha^l\right) G\left(\psi_{\alpha^{\nu-k}}, \chi_{\alpha^{l-\tau}-1}\right) \end{aligned}$$

where the commutation relation of the modulation and translation operator was used. Now consider

$$\begin{aligned} &\sum_{\nu,\tau} \left\langle \mathbf{M}^k \mathbf{T}^l \boldsymbol{\phi}, \mathbf{M}^\nu \mathbf{T}^\tau \boldsymbol{\phi} \right\rangle = \sum_{\nu,\tau} \omega^{l(\nu-k)} \left\langle \boldsymbol{\phi}, \mathbf{M}^{\nu-k} \mathbf{T}^{\tau-l} \boldsymbol{\phi} \right\rangle \\ &= \frac{1}{L} \sum_{\gamma,\eta} \psi_{\alpha^\gamma} \left(\alpha^l\right) G\left(\psi_{\alpha^\gamma}, \chi_{\alpha^\eta-1}\right) \\ &= \frac{1}{L} + \frac{1}{L} \sum_{\gamma\neq 0} \psi_{\alpha^\gamma} \left(\alpha^l\right) G\left(\psi_{\alpha^\gamma}, \chi_1\right) \sum_{\eta\neq 0} \overline{\psi_{\alpha^\gamma}\left(1-\alpha^\eta\right)} \\ &= \frac{1}{L} - \frac{1}{L} \sum_{\gamma\neq 0} \psi_{\alpha^\gamma} \left(\alpha^l\right) G\left(\psi_{\alpha^\gamma}, \chi_1\right) \end{aligned} \tag{6.12}$$

where we used $\sum_{\eta\neq 0} \overline{\psi_{\alpha^\gamma}\left(1-\alpha^\eta\right)} = -1$. For the inner products with the Fourier basis one obtains

$$\sum_i \left\langle \mathbf{M}^k \mathbf{T}^l \boldsymbol{\phi}, \mathbf{w}_i \right\rangle = \sum_i \omega^{l(i-k)} \langle \boldsymbol{\phi}, \mathbf{w}_{i-k} \rangle = \sum_h \omega^{lh} \langle \boldsymbol{\phi}, \mathbf{w}_h \rangle \tag{6.13}$$

$$= -\frac{1}{L} + \frac{1}{L} \sum_h \psi_{\alpha^h} \left(\alpha^l\right) G\left(\psi_{\alpha^h}, \chi_1\right). \tag{6.14}$$

Combining (6.12) and (6.13) we get

$$\sum_{\nu,\tau} \left\langle \mathbf{M}^k \mathbf{T}^l \boldsymbol{\phi}, \mathbf{M}^\nu \mathbf{T}^\tau \boldsymbol{\phi} \right\rangle + \sum_i \left\langle \mathbf{M}^k \mathbf{T}^l \boldsymbol{\phi}, \mathbf{w}_i \right\rangle = 0$$

which is what we wanted to show. If the initial choice of the column is an

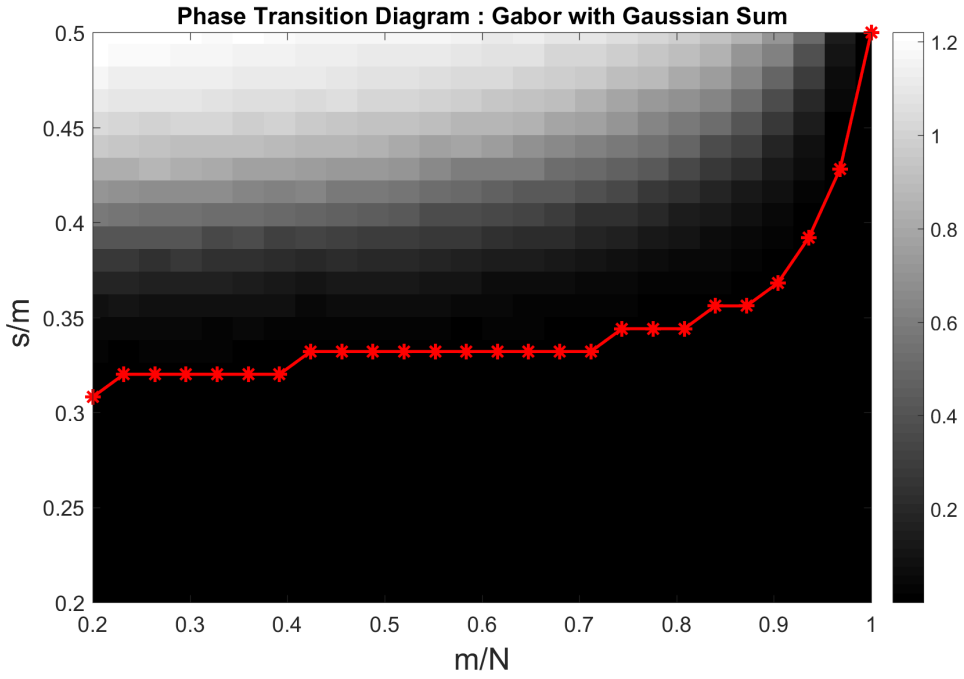

Figure 6.1.: ℓ_2-recovery error as phase transition diagram and transition curve (in red) of a Gabor frame generated by the window in (6.8).

element of the Fourier basis, one gets

$$\begin{aligned} 1+\sum_{\tau,\nu} \langle \mathbf{f}_i, \mathbf{M}^\nu \mathbf{T}^\tau \boldsymbol{\phi} \rangle &= 1+\frac{1}{L}\sum_{\tau,\nu} \psi_{\alpha^{\nu-i}}(\alpha^\tau)\, G(\psi_{\alpha^{\nu-i}}, \chi_1) \\ &= \frac{1}{L}\sum_{\gamma\neq 0} G(\psi_{\alpha^\gamma}, \chi_1) \sum_{\eta} \psi_{\alpha^\gamma}(\alpha^\eta) = 0 \end{aligned}$$

finally, using that the columns of $\mathbf{\Phi}$ are normed, we arrive at $\Delta = \frac{1}{L^2+L-1}$. □

6.7. Numerical experiments

Some numerical experiments were performed to illustrate that the family of Gabor matrices constructed in Section 6.5 provides good measurement matrices for CS.

Fig. 6.1 shows a *phase transition diagram* for sparse signal recovery using orthogonal matching pursuit (OMP) [37, 121, 122] as recovery algorithm and where we used the Gabor frame $\mathbf{G}_{\boldsymbol{\phi}}$, generated by (6.8), as measurement matrix

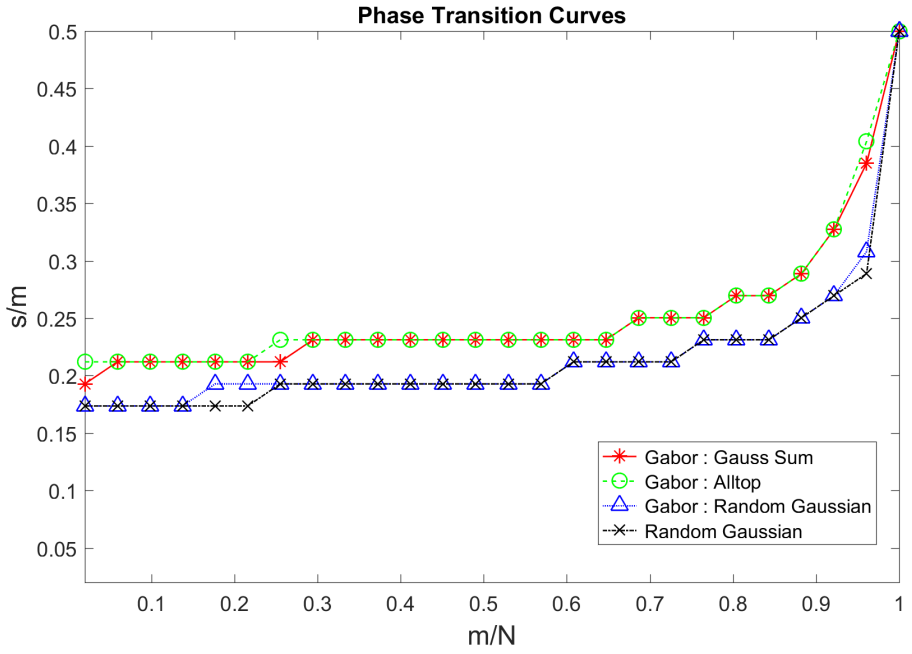

Figure 6.2.: Phase transition curves of Gabor frames generated by the windows red: (6.8), green: Alltop, blue: normed Random Gaussian and the phase transition curve of a random Gaussian sensing matrix.

$\mathbf{\Phi}$.
Black indicates a small and white a high average recovery error, respectively. The *phase transition curve* (in red) corresponds to an average error of 1%.
The graphs in Fig. 6.2 are obtained as the red curve in Fig. 6.1 but for different measurement matrices $\mathbf{\Phi}$. For the Gabor measurement matrix with Alltop window and with window (6.8), each data point value is averaged over 1000 experiments. For the Gabor matrix with random Gaussian window and for the random Gaussian measurement matrix, 100 different measurement matrices were generated and for each matrix 100 different experiments were done for each data point. Then the final value was averaged over all experiments. The matrix size for the Alltop measurements was $L = 53$ (denoted by m) and $N = 2809$ ($N = L^2$) for all the other matrices the size was $L = 52$ (denoted by L) and $N = 2704$ ($N = 52^2$).
Figure 6.2 indicates that the Gabor system $\mathbf{G}_\phi$ constructed in Section 6.5 performs as well as the construction with the Alltop window which is known to be a good sensing matrix [43]. Surprisingly, the deterministic matrix constructions seem to outperform the random constructions, although best results concerning

sparse recovery are proven for random constructions. A simple explanation is, that the results for random constructions are usually accompanied by some universal constants [37, 53, 89, 105]. It is likely that the dimension $L \sim 50$ is just not sufficiently large enough.

6.8. Number of inner products in Mersenne prime dimensions

Gaussian sums behave unpredictable when it comes to their phases. Nevertheless, an interesting aspect of the construction in Section 6.5 is that in dimensions $L = 2^p - 1$ where L is a prime number, also called Mersenne prime, we are able characterize how often a phase appears. We state and prove this observation in the following proposition.

Proposition 6.8.1. *Let $L = 2^p - 1$ be a prime number and let $\boldsymbol{\phi} \in \mathbb{C}^L$ be given as in* (6.8) *and consider the inner products*

$$\langle \boldsymbol{\phi}, \mathbf{M}^\nu \mathbf{T}^\tau \boldsymbol{\phi} \rangle \quad \textit{and} \quad \langle \mathbf{w}_k, \boldsymbol{\phi} \rangle \ .$$

Then varying over τ, ν, k there appears exactly $\frac{L^2-L}{p}$ different phases with modulus $\frac{\sqrt{L+1}}{L}$ and each of this phases appear exactly p times.

Proof. Consider the integers modulo $2^p - 1$, i.e. $\mathbb{Z}_{2^p-1} \simeq \mathbb{F}_{2^p-1} \simeq \mathbb{F}_{2^p}^*$ which is a field since $2^p - 1$ is a prime number and at the same time isomorphic to the multiplicative character group of $\mathbb{F}_{2^p}$. Take an element $c_1 \in \mathbb{F}_{2^p-1} \setminus \{0\}$ and consider the set $\mathcal{C}_1 = \{c_1, 2c_1, 2^2c_1, ..., 2^{p-1}c_1\}$ note that $c_1 2^p = c_1$ since $1 \equiv 2^p \mod 2^p - 1$ and since we are working over a field there does not exist $a \neq b$ s.t. $ac = cb$ with $a, b, c \in \mathbb{F}_{2^p-1} \setminus \{0\}$. That implies the existence of a $c_2 \notin \mathcal{C}_1$ and therefore the existence of a second set $\mathcal{C}_2 = \{c_2, 2c_2, 2^2c_2, ..., 2^{q-1}c_2\}$ with $\mathcal{C}_1 \bigcap \mathcal{C}_2 = \emptyset$. Now by the same line of argumentation and Lemma 5.4.3 we see that there exist a disjoint tiling of $\mathbb{F}_{2^p-1} \setminus \{0\}$ in $\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_{\frac{2^p-2}{p}}$.

The dimension L being a prime number implies that in each row of the Gramian matrix will be the same values just in different positions. This fact can be deduced from the proof of Proposition 6.6.2 and that the columns form a group under pointwise multiplication. Next we consider the commutative group $\{\mathbf{M}^\nu\}_{\nu=0}^{2^p-2}$ and its cosets with the innerproducts $\langle \mathbf{w}_0, \mathbf{M}^\nu \mathbf{T}^\tau \boldsymbol{\phi} \rangle$, the case when $\nu = 0$ is clear therefore we will consider just the cases when $\nu \neq 0$. From previous discussion we can partition $\{\mathbf{M}^\nu\}_{\nu=0}^{2^p-2}$ into $\frac{2^p-2}{p}$ different sets. Note that p must be a prime number since $L = 2^p - 1$ is a prime number [113] and

by Lemma 5.4.3 we have that p divides $2^p - 2$.

Now by Theorem 6.2.2 for $\nu_1 \neq \nu_2$ we can write

$$\langle \mathbf{w}_0 \mathbf{M}^{\nu_1} \boldsymbol{\phi} \rangle = \langle \mathbf{w}_0 \mathbf{M}^{\nu_2} \boldsymbol{\phi} \rangle \quad \text{if } \nu_1, \nu_2 \in \mathcal{C}_i, i \in \{1, ..., \frac{2^p - 2}{p}\},$$

additionally, by Theorem 6.2.2 and by the commutation relation in Lemma 4.2.1 we observe

$$\langle \mathbf{w}_0, \mathbf{M}^{\nu_1} \mathbf{T}^{\tau_1} \boldsymbol{\phi} \rangle = \omega^{\nu_1 \tau_1} \langle \mathbf{w}_0, \mathbf{M}^{\nu_1} \boldsymbol{\phi} \rangle = \langle \mathbf{w}_0, \mathbf{M}^{\nu_2} \mathbf{T}^{\tau_2} \boldsymbol{\phi} \rangle$$
$$\text{if } \nu_1, \nu_2 \in \mathcal{C}_i \text{ and } \nu_1 \tau_1 = \nu_2 \tau_2.$$

Next fixing $\nu \neq 0$ with $\nu \in \mathcal{C}_i$, going over all $\tau \in \{1, ..., 2^p - 2\}$ and looking at $\omega^{\nu\tau}$ we will get every primitive root of unity, since $2^p - 1$ is prime and ω a $2^p - 1$-th root of unity. Now varying ν over all elements of $\mathcal{C}_i$ we will get each primitive root exactly p times since $|\mathcal{C}_i| = p$. Finally these considerations imply that the number of different values of innerproducts in the normed frame is exactly $(2^p - 1)\frac{2^p-2}{p} + 3$, where the latter 3 values are, 1 appearing exactly once, 0 appearing $L - 1$ times and $-\frac{1}{L}$ appearing L times. There are $(2^p - 1)\frac{2^p-2}{p}$ different phases, closed under complex conjugation, corresponding to the innerproducts with modulus $\frac{\sqrt{L+1}}{L}$ each appearing exactly p times. □

7. On Compressed Sensing of Sparse Covariance Matrices using Deterministic Sensing Matrices

In this section we consider the problem of determining the sparse covariance matrix $\mathbf{X}$ of an unknown data vector $\mathbf{x}$ by observing the covariance matrix $\mathbf{Y}$ of a compressive measurement vector $\mathbf{y} = \mathbf{A}\mathbf{x}$. Our aim is to construct deterministic sensing matrices $\mathbf{A}$ for which the recovery of a k-sparse covariance matrix $\mathbf{X}$ from L values of $\mathbf{Y}$ is guaranteed with high probability. This section has been partially published in [30].

7.1. Problem formulation

Let $\mathbf{x} = (x_1, x_2, \ldots, x_N)^{\mathrm{T}} \in \mathbb{C}^N$ be a vector of N independent, zero-mean random variables with covariance matrix $\mathbf{X} = \mathbb{E}[\mathbf{x}\mathbf{x}^*]$, and let $\mathbf{y} = \mathbf{A}\mathbf{x}$ be L linear measurements of $\mathbf{x}$ with the $L \times N$ measurement matrix $\mathbf{A} \in \mathbb{C}^{L\times N}$. We consider the problem of determining $\mathbf{X}$ from the known covariance matrix

$$\mathbf{Y} = \mathbb{E}[\mathbf{y}\mathbf{y}^*] = \mathbf{A}\,\mathbb{E}[\mathbf{x}\mathbf{x}^*]\,\mathbf{A}^* = \mathbf{A}\mathbf{X}\mathbf{A}^* \tag{7.1}$$

of the observed measurements $\mathbf{y}$. This problem, also known as matrix sketching, appears in several problems of signal processing, for instance in array signal processing or communications [123].
In many applications, the covariance matrix $\mathbf{X}$ can assumed to be sparse in some sense. For example, if two random variables x_i and x_j are known to be uncorrelated then the corresponding entries in the covariance matrix, namely $[\mathbf{X}]_{i,j} = \mathbb{E}[x_i\overline{x_j}]$ and $[\mathbf{X}]_{j,i} = \mathbb{E}[x_j\overline{x_i}]$, are zero. So in cases where only a few entries of $\mathbf{x}$ are correlated, the matrix $\mathbf{X}$ will be sparse. Therefore, ideas from compressive sensing (CS) may be applied to find efficient sampling schemes which only need a few measurements to determine $\mathbf{X}$ [124]. In particular, it is

natural to ask whether it is possible to find sensing matrices $\mathbf{A}$ with $L < N$ rows and such that $\mathbf{X}$ can uniquely recovered from $\mathbf{Y}$.
One common approach [125] is based on rewriting (7.1) as a standard linear compressed sensing problem by stacking the columns of $\mathbf{X}$ and $\mathbf{Y}$ into vectors $\widetilde{\mathbf{x}} = \mathrm{vec}(\mathbf{X}) \in \mathbb{C}^{N^2}$ and $\widetilde{\mathbf{y}} = \mathrm{vec}(\mathbf{Y}) \in \mathbb{C}^{L^2}$ respectively. This yields

$$\widetilde{\mathbf{y}} = \mathbf{C}\,\widetilde{\mathbf{x}} \qquad \text{with} \qquad \mathbf{C} = \overline{\mathbf{A}} \otimes \mathbf{A} \tag{7.2}$$

and wherein $\otimes$ stands for the usual Kronecker product of matrices. The problem is now to find a measurement matrix $\mathbf{A}$ such that the corresponding matrix $\mathbf{C} = \overline{\mathbf{A}} \otimes \mathbf{A}$ in (7.2) is a good measurement matrix for compressed sensing. Then the problem (7.2) can be solved uniquely by standard compressed sensing algorithms [37, 121].

7.2. Statistical restricted isometry property for Kronecker structured matrices

Since we are interested in deterministic sensing matrices in (7.2) we can employ the frame work of statistical restricted isometry property discussed in Section 3.1.
We now consider the following problem. Let $\mathbf{A}$ be a matrix which is known to be η-StRIP-able for some positive η. Is the Kronecker product $\overline{\mathbf{A}} \otimes \mathbf{A}$ again η'-StRIP-able for some $\eta' > 0$? If so, what would be the value of η'? We give a complete answer to these questions in the following theorem.

Theorem 7.2.1. *Assume $\boldsymbol{A} \in \mathbb{C}^{n\times N}$ is η_A-StRIP-able and $\boldsymbol{B} \in \mathbb{C}^{m\times M}$ is η_B-StRIP-able. Then the following holds.*

(a) $\overline{\boldsymbol{A}}$ is $\eta_{\overline{A}}$-StRIP-able with $\eta_{\overline{A}} = \eta_A$.

(b) The matrix $\boldsymbol{C} = \boldsymbol{A} \otimes \boldsymbol{B} \in \mathbb{C}^{nm\times NM}$ is η_C-StRIP-able with

$$\eta_C = \begin{cases} \eta_A \frac{\ln(n)}{\ln(nm)} & \text{if } n^{\eta_A} \le m^{\eta_B}\,, \\ \eta_B \frac{\ln(m)}{\ln(nm)} & \text{if } n^{\eta_A} > m^{\eta_B}. \end{cases} \tag{7.3}$$

Proof. Let us write $\mathbf{A} = \frac{1}{\sqrt{n}}\boldsymbol{\Phi}$, $\mathbf{B} = \frac{1}{\sqrt{m}}\boldsymbol{\Psi}$, and $\mathbf{C} = \frac{1}{\sqrt{nm}}\boldsymbol{\Gamma}$. Clearly, the matrices $\boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Gamma}$ are related by $\boldsymbol{\Gamma} = \boldsymbol{\Phi} \otimes \boldsymbol{\Psi}$ and the entry of $\boldsymbol{\Gamma}$ in the (k,ℓ)-th column and (x,y)-th row is given by $\boldsymbol{\gamma}_{(k,\ell)}[x,y] = \boldsymbol{\phi}_k[x]\boldsymbol{\psi}_\ell[y]$.

Part (a) is immediate by observing that each conditions (S1), (S2), (S3) for $\boldsymbol{\Phi}$ implies the respective condition for $\overline{\boldsymbol{\Phi}}$. Therefore $\overline{\mathbf{A}}$ is $\eta_{\overline{A}}$-StRIP-able with $\eta_{\overline{A}} = \eta_A$ (the same constant as $\mathbf{A}$).

To prove (b), we check conditions (S1), (S2) and (S3) for $\boldsymbol{\Gamma}$. First, observe that for $(x, y) \neq (x', y')$,

$$\begin{aligned}\sum_{k=1}^{N}\sum_{\ell=1}^{M} \boldsymbol{\gamma}_{(k,\ell)}[x,y]\,\overline{\boldsymbol{\gamma}_{(k,\ell)}[x',y']} &= \sum_{k=1}^{N}\sum_{\ell=1}^{M} \boldsymbol{\phi}_k[x]\,\boldsymbol{\psi}_\ell[y]\,\overline{\boldsymbol{\phi}_k[x']\,\boldsymbol{\psi}_\ell[y']} \\ &= \sum_{k=1}^{N} \boldsymbol{\phi}_k[x]\,\overline{\phi_k[x']} \cdot \sum_{\ell=1}^{M} \psi_\ell[y]\,\overline{\psi_\ell[y']} = 0,\end{aligned}$$

using the fact that $\sum_{k=1}^{N} \boldsymbol{\phi}_k[x]\overline{\boldsymbol{\phi}_k[x']} = 0$ if $x \neq x'$, and that $\sum_{\ell=1}^{M} \boldsymbol{\psi}_\ell[y]\overline{\boldsymbol{\psi}_\ell[y']} = 0$ if $y \neq y'$. Moreover, for any (x, y),

$$\begin{aligned}\sum_{k=1}^{N}\sum_{\ell=1}^{M} \boldsymbol{\gamma}_{(k,\ell)}[x,y] &= \sum_{k=1}^{N}\sum_{\ell=1}^{M} \boldsymbol{\phi}_k[x]\boldsymbol{\psi}_\ell[y] \\ &= \sum_{k=1}^{N} \boldsymbol{\phi}_k[x] \cdot \sum_{\ell=1}^{M} \boldsymbol{\psi}_\ell[y] = 0,\end{aligned}$$

where we have used that $\sum_{k=1}^{N} \boldsymbol{\phi}_k[x] = \sum_{\ell=1}^{M} \boldsymbol{\psi}_\ell[y] = 0$. Therefore, $\boldsymbol{\Gamma}$ satisfies the condition (S1).

To verify the (S2) for $\boldsymbol{\Gamma}$, fix any (k, ℓ) and (k', ℓ'), where $1 \leq k, k' \leq N$, $1 \leq \ell, \ell' \leq M$. Since $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ satisfy (S2) there exist $1 \leq k'' \leq N$ and $1 \leq \ell'' \leq M$ such that $\boldsymbol{\phi}_k[x]\boldsymbol{\phi}_{k'}[x] = \boldsymbol{\phi}_{k''}[x]$ for all x and $\boldsymbol{\psi}_\ell[y]\boldsymbol{\psi}_{\ell'}[y] = \boldsymbol{\psi}_{\ell''}[y]$ for all y. Then

$$\begin{aligned}\boldsymbol{\gamma}_{(k,\ell)}[x,y] \cdot \boldsymbol{\gamma}_{(k',\ell')}[x,y] &= \boldsymbol{\phi}_k[x]\boldsymbol{\psi}_\ell[y] \cdot \boldsymbol{\phi}_{k'}[x]\boldsymbol{\psi}_{\ell'}[y] \\ &= \boldsymbol{\phi}_k[x]\boldsymbol{\phi}_{k'}[x] \cdot \boldsymbol{\psi}_\ell[y]\boldsymbol{\psi}_{\ell'}[y] \\ &= \boldsymbol{\phi}_{k''}[x] \cdot \boldsymbol{\psi}_{\ell''}[y] = \boldsymbol{\gamma}_{(k'',\ell'')}[x,y]\end{aligned}$$

which proves (S2).

Finally, we verify (S3) for $\boldsymbol{\Gamma}$. For $(k, \ell) \neq (1, 1)$ and with $1 \leq k \leq N$, $1 \leq \ell \leq M$, we have

$$\left|\sum_{x=1}^{n}\sum_{y=1}^{m} \boldsymbol{\gamma}_{(k,\ell)}[x,y]\right|^2 = \left|\sum_{x=1}^{n}\sum_{y=1}^{m} \boldsymbol{\phi}_k[x]\,\boldsymbol{\psi}_\ell[y]\right|^2$$

$$= \left|\sum_{x=1}^{n} \boldsymbol{\phi}_k[x]\right|^2 \cdot \left|\sum_{y=1}^{m} \boldsymbol{\psi}_\ell[y]\right|^2$$
$$= \begin{cases} n^2 \cdot m^{2-\eta_B} & \text{if } k=1,\ \ell \neq 1, \\ n^{2-\eta_A} \cdot m^2 & \text{if } k=1,\ \ell \neq 1, \\ n^{2-\eta_A} \cdot m^{2-\eta_B} & \text{if } k \neq 1,\ \ell \neq 1, \end{cases}$$

and since $\eta_A, \eta_B > 0$, we have

$$\max_{(k,\ell)\neq(1,1)} \left|\sum_{x=1}^{n}\sum_{y=1}^{m} \boldsymbol{\gamma}_{(k,\ell)}[x,y]\right|^2 = \max\left\{n^2 m^{2-\eta_B},\ n^{2-\eta_A} m^2\right\}.$$

Setting $(nm)^{2-\eta_C} = \max\{n^2 m^{2-\eta_B}, n^{2-\eta_A} m^2\}$ gives

$$\eta_C = \begin{cases} \eta_A \frac{\ln(n)}{\ln(nm)} & \text{if } n^{\eta_A} \leq m^{\eta_B} \\ \eta_B \frac{\ln(m)}{\ln(nm)} & \text{if } n^{\eta_A} > m^{\eta_B}. \end{cases}$$

which finishes the proof. □

Theorem 7.2.1 shows that the Kronecker product $\mathbf{C} = \mathbf{A} \otimes \mathbf{B}$ of two StRIP-able matrices $\mathbf{A}$ and $\mathbf{B}$ is again StRIP-able. However, the constant η_C given in (7.3) is always strictly smaller than both η_A and η_B, i.e., $\eta_C < \min\{\eta_A, \eta_B\}$. This means that the coherence of a Kronecker product matrix is always worse (i.e., larger) than the coherence of the original matrices.
Motivated from the applications described in Introduction, we are interested in sensing matrices of the form $\overline{\mathbf{A}} \otimes \mathbf{A}$ (cf. (7.2)). For such matrices, Theorem 7.2.1 immediately yields the following.

Corollary 7.2.2. *If $\mathbf{A} \in \mathbb{C}^{m\times N}$ is an η-StRIP-able matrix, then the matrix $\overline{\mathbf{A}} \otimes \mathbf{A} \in \mathbb{C}^{m^2\times N^2}$ is $(\eta/2)$-StRIP-able.*

Combining Theorem 3.1.1 and Corollary 7.2.2, we obtain a sufficient condition under which a Kronecker product matrix $\overline{\mathbf{A}} \otimes \mathbf{A} \in \mathbb{C}^{m^2\times N^2}$ satisfies UStRIP.

Corollary 7.2.3. *Let $\mathbf{A} = \frac{1}{\sqrt{m}}\boldsymbol{\Phi} \in \mathbb{C}^{m\times N}$ be an η-StRIP-able matrix with $\eta > 1$. If $k < 1 + (N^2-1)\delta$ and $m^2 \geq c\,(2k \log N)/\delta^2$ for some constant $c > 0$, then $\overline{\mathbf{A}} \otimes \mathbf{A}$ has $(k, \delta, 2\epsilon)$-UStRIP with*

$$\epsilon = 2\exp\left(-\left(\delta - \frac{k-1}{N^2-1}\right)^2 \frac{m^{2\eta}}{8k}\right). \tag{7.4}$$

So if a matrix $\mathbf{A}$ satisfies the conditions of Corollary 7.2.3, then one needs in the order of $\widetilde{m} := m^2 \geq c\,k\,\log N$ measurements of the form $\mathbf{y} = \left(\overline{\mathbf{A}} \otimes \mathbf{A}\right)\mathbf{x}$ to recover k-sparse vectors $\mathbf{x} \in \mathbb{C}^{N^2}$ with high probability. Equivalently, every covariance matrix $\mathbf{X} \in \mathbb{C}^{N \times N}$ can be recovered from $\widetilde{m}$ values of the covariance matrix $\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{A}^*$.

7.3. Kronecker structured matrices with recovery guarantee

Corollary 7.2.3 requires the StRIP constant η of an StRIP-able matrix $\mathbf{A} \in \mathbb{C}^{L \times N}$ has to be larger than 1 for $\overline{\mathbf{A}} \otimes \mathbf{A}$ to have UStRIP. To get a first idea which matrices might satisfy this condition, we recall from Section 3.1 that the coherence of an η-StRIP-able matrix is upper bounded by $\mu(\mathbf{A}) \leq L^{-\eta/2}$. On the other side, from Theorem 2.4.3 $\mu(\mathbf{A})$ is known to be lower bounded by the Welch bound [64]. So $\mu(\mathbf{A})$ always satisfies

$$\sqrt{\frac{N-L}{L(N-1)}} \leq \mu(\mathbf{A}) \leq \frac{1}{\sqrt{L^{\eta}}}$$

From these inequalities, one easily derives an upper bound on η:

$$\eta \leq 1 + \ln\left(\frac{N-1}{N-L}\right)\frac{1}{\ln(L)}$$

For $L > 1$, this upper bound is strictly larger than 1 but it gets very close to 1 for $L << N$ (as usually desired in CS). Since we are looking for matrices with $\eta > 1$, this means that we are searching for matrices $\mathbf{A}$ whose coherence is very close to the Welch bound, which means that the columns of $\mathbf{A}$ have to be close to an *equiangular tight frame* (ETF). In particular, we observe that if the coherence of $\mathbf{A}$ would achieve the Welch bound with equality then the Kronecker product $\overline{\mathbf{A}} \otimes \mathbf{A}$ would have UStRIP. Additionally, such an equal norm ETF needs to fulfill (S1) – (S3). A class of ETFs fulfilling these conditions are *equiangular harmonic frames (EHF)* [126]. This frames are constructed by picking out certain rows from the DFT matrix. The rows are indexed by difference sets. Note that by picking some arbitrary rows (apart from the first all ones row) of the DFT matrix, the partial DFT matrix fulfills (S1) and (S2), but not necessarily (S3).

Proposition 7.3.1. *Let $\mathcal{K} \subset \mathbb{Z}_N$ be an (N, L, ρ)-difference set. Then the*

partial Fourier matrix $\mathbf{F}_{\mathcal{K}} \in \mathbb{C}^{L\times N}$ *is* η*-StRIP-able with*

$$\eta = 2 - \frac{\ln(L-\rho)}{\ln(L)} > 1,$$

where $\mathbf{F}_{\mathcal{K}} = [\mathrm{e}^{-j\frac{2\pi}{N}rk}]_{k\in\mathcal{K},r=0,\ldots,N-1}$ *is the partial Fourier matrix formed with the rows indexed by* $\mathcal{K}$*.*

Proof. Conditions (S1) and (S2) follow from the properties of the $N\times N$ discrete Fourier transform matrix. To verify condition (S3), let $\mathcal{K} = \{a_1, a_2, \ldots, a_L\} \subset \mathbb{Z}_N$ be given in the increasing order. The entry of $\mathbf{F}_{\mathcal{K}}$ at the r-th column and k-th row is given by $\mathbf{f}_r[k] = \omega^{ra_k}$, where $\omega := \mathrm{e}^{-j2\pi/N}$. Then for any $r \in \{2, 3, ..., N\}$,

$$\begin{aligned}\left|\sum_{k=1}^{L}\mathbf{f}_r[k]\right|^2 &= \left|\sum_{k=1}^{L}\omega^{ra_k}\right|^2 = \sum_{k,\ell=1}^{L}\omega^{r(a_k-a_\ell)}\\ &= L + \sum_{k\neq\ell}\omega^{r(a_k-a_\ell)} = L + \rho(\omega+\omega^2+\ldots+\omega^{L-1})\\ &= L-\rho,\end{aligned}$$

where we used the property of (N, L, ρ)-difference set $\mathcal{K}$ and Lemma 4.2.2. Setting this value equal to $L^{2-\eta}$ yields the desired expression for η. □

Since the constructed matrix $\mathbf{F}_{\mathcal{K}}$ is η-StRIP-able with $\eta > 1$, the corresponding Kronecker product $\overline{\mathbf{F}_{\mathcal{K}}} \otimes \mathbf{F}_{\mathcal{K}}$ is η-StRIP-able with $\eta > 1/2$ (cf. Corollary 7.2.2). Consequently $\overline{\mathbf{F}_{\mathcal{K}}} \otimes \mathbf{F}_{\mathcal{K}}$ has UStRIP according to Theorem 3.1.1, i.e. we have the following statement.

Corollary 7.3.2. *Let* $\mathbf{F}_{\mathcal{K}} \in \mathbb{C}^{L\times N}$ *be a matrix constructed as in Proposition 7.3.1 and let* $\mathbf{C}_{\mathcal{K}} = \overline{\mathbf{F}_{\mathcal{K}}} \otimes \mathbf{F}_{\mathcal{K}}$*. If* $k < 1 + (N^2-1)\,\delta$ *and if* $L^2 \geq c\,(2k\log N)/\delta^2$ *for some constant* $c > 0$ *then* $\mathbf{C}_{\mathcal{K}}$ *has* $(k, \delta, 2\epsilon)$*-UStRIP with* ϵ *given by* (7.4)*.*

We remark again, that Corollary 7.3.2 implies in particular a statistical recovery guarantee (in the sense of Def. 3.1.1) for a Kronecker structured measurement matrix and where the number of measurements $\widetilde{L} = L^2$ scales *linearly* with the sparsity k.

7.4. Numerical experiments

This section presents numerical experiments showing the effectiveness of the proposed measurement matrices. Before comparing the recovery performance

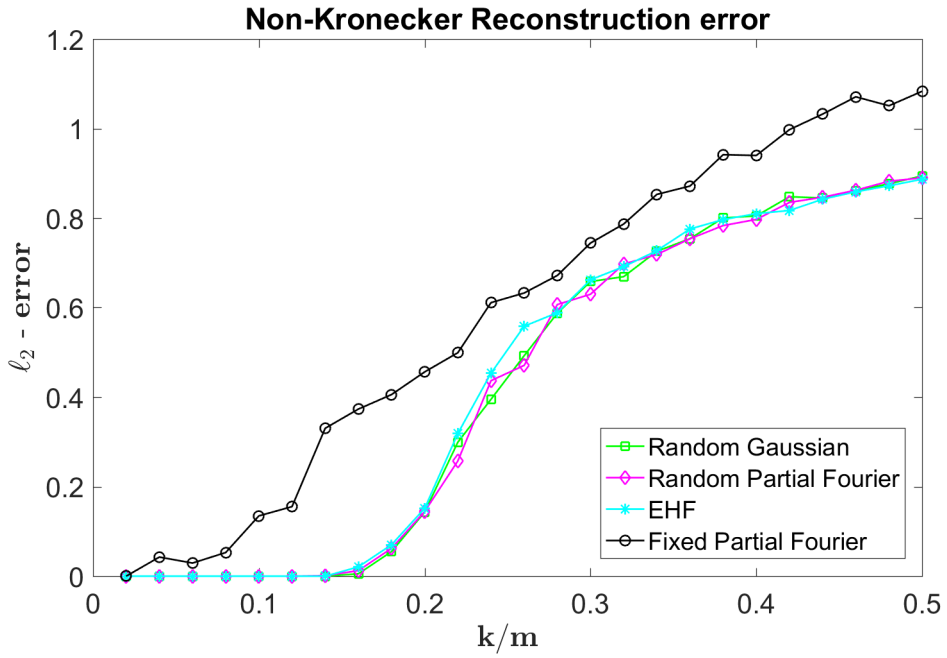

Figure 7.1.: Quadratic error of the optimal solutions of (7.5) for non-Kronecker structured matrices $\mathbf{C}$. Horizontal axis: k/L (sparsity over number of measurements). Vertical axis: ℓ_2-error ($\|\widehat{\mathbf{x}} - \mathbf{x}\|^2 / \|\mathbf{x}\|^2$).

of Kronecker product matrices $\mathbf{C} = (\overline{\mathbf{A}} \otimes \mathbf{A})$, we first check the performance of matrices $\mathbf{C}$ without Kronecker structure. To this end, we consider the following matrices $\mathbf{C}$ all of them having $L = 50$ rows and $N = 2451$ columns.

(i) *EHF*: $\mathbf{C} = \mathbf{F}_{\mathcal{K}}$ is the matrix constructed according to Prop. 7.3.1.

(ii) *deterministic partial Fourier*: $\mathbf{C}$ whose columns coincide with the first L columns of the $N \times N$ DFT matrix.

(iii) *random partial Fourier*: the rows of $\mathbf{C}$ are randomly chosen from the $N \times N$ DFT matrix.

(iv) *random Gaussian*: the entries of $\mathbf{C}$ are i.i.d normal distributed random variables.

Fig. 7.1 shows the corresponding simulation result for recovering a k-sparse vector $\mathbf{x} \in \mathbb{C}^N$ from linear measurements $\mathbf{y} = \mathbf{C}\,\mathbf{x}$, using basis pursuit (7.5), i.e. by solving

$$\widehat{\mathbf{x}} = \arg\min \|\mathbf{z}\|_1 \quad \text{subject to} \quad \mathbf{C}\mathbf{z} = \mathbf{y}\ , \mathbf{z} \in \mathbb{C}^N\ . \tag{7.5}$$

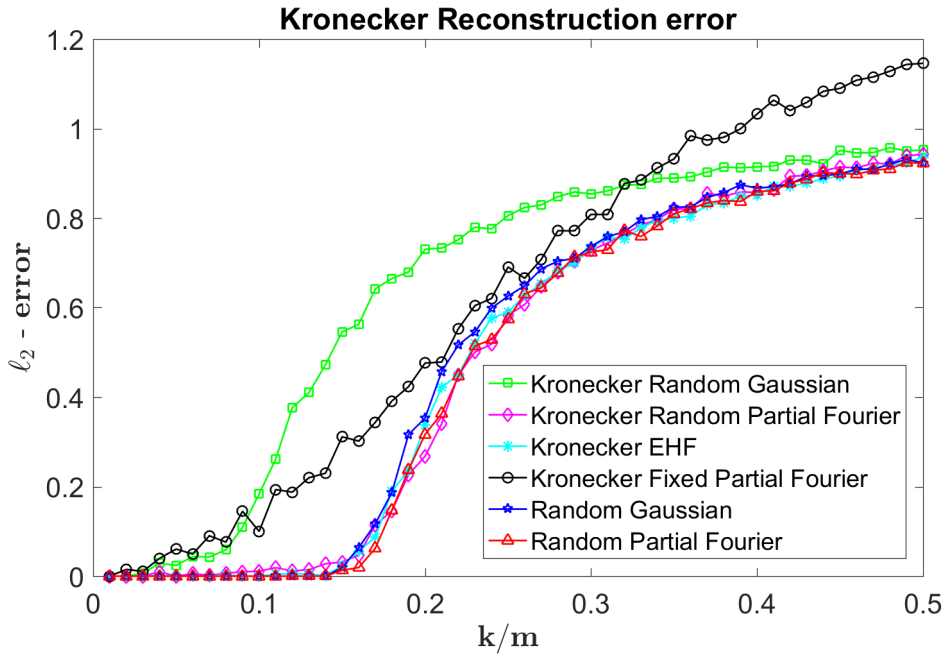

Figure 7.2.: Quadratic error of the optimal solutions of (7.5) for Kronecker structured matrices $\mathbf{C}$, and comparison with random Gaussian and random partial Fourier matrices. Axis as in Fig. 7.1.

For these simulation we varied the sparsity k of the data vector $\mathbf{x}$. On the horizontal axis we plot the normalized ℓ_2 reconstruction error $\|\widehat{\mathbf{x}} - \mathbf{x}\|^2 / \|\mathbf{x}\|^2$. For each k we generated 100 random k-sparse vectors $\mathbf{x}$ and averaged the reconstruction error over these 100 experiments. The simulation result for the deterministic partial Fourier matrix in Fig. 7.1 shows that not every choice of rows from the discrete Fourier transform matrix yields a good compressed sensing matrix. However, for a choice of rows that corresponds to an EHF the resulting measurement matrix is essentially as good as random Gaussian and random partial Fourier matrices, which are known to be good sensing matrices. In Fig 7.2, we compare the recovery performance of Kronecker product matrices $\mathbf{C} = (\overline{\mathbf{A}} \otimes \mathbf{A})$ for matrices $\mathbf{A} \in \mathbb{C}^{m \times N}$ as under (i)-(iv), denoted respectively by (i')-(iv'). Additionally, we consider random matrices $\mathbf{C}$ of size $L^2 \times N^2$ (without Kronecker product):

(v') *random partial Fourier*: the rows of $\mathbf{C}$ are randomly chosen from the $N^2 \times N^2$ DFT matrix.

(vi') *random Gaussian*: the entries of $\mathbf{C}$ are i.i.d normal distributed random

variables.

For these simulations, we fixed $L = 10$ and $N = 91$, and the results where averaged over 100 random vectors $\mathbf{x}$. We observe that the Kronecker structure destroys the good behaviour of the random Gaussian matrix which now performs worse. On the other side, we see that the Kronecker structured EHF matrix performs almost as good as the non-Kronecker-structured random partial Fourier and random Gaussian matrices. So for our *deterministic* EHF matrix the Kronecker structure does not harm its good compressed sensing properties. A similar behaviour is observed for the random partial Fourier matrix.

Part III.

Data Transmission over LTV Channels

8. Channel Model

In this section we introduce the communication channel model which will form the basis for discussions in the later sections. Parts of the Section in this form has been published in [31, 32, 34].
We consider a linear time-varying (LTV) multipath wireless communication channel. Commonly, LTV channels are modelled by linear time-varying systems [3, 10, 127]. The input-output relation of a LTV system is described by

$$(\mathrm{H}s)(t) = \iint_{\mathbb{R}^2} \eta(\tau,\nu) s(t-\tau)\, \mathrm{e}^{j2\pi\nu t}\, d\tau\, d\nu\,. \tag{8.1}$$

Wherein, $s(t)$ is the input signal and $\eta(\tau,\nu)$ describes the complex attenuation factor associated with the corresponding delay-Doppler pair (τ,ν), which is the result of multipath propagation and Doppler spread. The function $\eta(\tau,\nu)$ is called the spreading function of H since it describes how the signal is spread in the delay-Doppler domain. The representation of a LTV system as formulated in (8.1) emphasizes the physical effects of a communication channel, i.e. multipath propagation and Doppler effect. To describe a LTV system a time-varying impulse response representation of (8.1) is commonly used in the literature [3, 11, 128].

$$(\mathrm{H}s)(t) = \int_{\mathbb{R}} A(t,\tau) s(t-\tau) d\tau, \tag{8.2}$$

with

$$A(t,\tau) = \int_{\mathbb{R}} \eta(\tau,\nu) \mathrm{e}^{j2\pi\nu t} d\nu.$$

Another common representation of a LTV system is given by its time-varying transfer function [3, 8, 128]. Analogously to time-invariant systems the time-varying transfer function is obtained by taking the Fourier transform of the time-varying impulse response on the second argument,

$$\sigma(t,\omega) = \int_{\mathbb{R}} A(t,\tau) \mathrm{e}^{-j2\pi\omega\tau} d\tau,$$

and the LTV system representation is given by

$$(\mathrm{H}s)(t) = \int_{\mathbb{R}} \sigma(t,\omega)\hat{s}(\omega)\mathrm{e}^{j2\pi\omega t}d\omega, \tag{8.3}$$

where $\hat{s}(\omega)$ is the Fourier transform of $s(t)$. The time-varying impulse response and time-varying transfer function representations can be formulated from one form into the other by straight forward computation. For instance we obtain (8.2) from (8.3) simply by straight forward computation,

$$\begin{aligned}(\mathrm{H}s)(t) &= \int_{\mathbb{R}} \sigma(t,\omega)\hat{s}(\omega)\mathrm{e}^{j2\pi\omega t}d\omega \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} A(t,\tau)\mathrm{e}^{-j2\pi\omega\tau}\hat{s}(\omega)\mathrm{e}^{j2\pi\omega t}d\omega d\tau \\ &= \int_{\mathbb{R}} A(t,\tau)\int_{\mathbb{R}} \mathrm{e}^{-j2\pi\omega\tau}\hat{s}(\omega)\mathrm{e}^{j2\pi\omega t}d\omega d\tau \\ &= \int_{\mathbb{R}} A(t,\tau)s(t-\tau)d\tau.\end{aligned}$$

Furthermore, since $\sigma(t,\omega)$ is the Fourier transform of $A(t,\tau)$ we have

$$A(t,\tau) = \int_{\mathbb{R}} \sigma(t,\omega)\mathrm{e}^{j2\pi\omega\tau}d\omega.$$

Thus, we can also get (8.2) from (8.3) as follows

$$\begin{aligned}(\mathrm{H}s)(t) &= \int_{\mathbb{R}} A(t,\tau)s(t-\tau)d\tau \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \sigma(t,\omega)\mathrm{e}^{j2\pi\omega\tau}s(t-\tau)d\tau d\omega \\ &= \int_{\mathbb{R}} \sigma(t,\omega)\int_{-\infty}^{\infty} \mathrm{e}^{j2\pi\omega\tau}s(t-\tau)d\tau d\omega \\ &= \int_{\mathbb{R}} \sigma(t,\omega)\mathrm{e}^{j2\pi\omega t}(-1)\int_{\infty}^{-\infty} s(z)\mathrm{e}^{-j2\pi\omega z}dz d\omega \\ &= \int_{\mathbb{R}} \sigma(t,\omega)\hat{s}(\omega)\mathrm{e}^{j2\pi\omega t}d\omega,\end{aligned}$$

where we substituted $z = t - \tau$.
Although all of the representations for LTV systems are used in the literature here we will mainly use the spreading function representation as given in (8.1) since this representation emphasizes multipath propagation and Doppler effects which appear in wireless communication channels [3, 10].
The representation given in (8.1) can also be written using the translation and modulation operators [129]. Now we are going to define the continuous

analogues to the finite modulation and translation operators which we have discussed in Section 4.2.1. For $x, \nu \in \mathbb{R}$, the translation and modulation operators are defined as

$$\mathrm{T}_x s(t) = s(t-x), \quad t \in \mathbb{R} \tag{8.4}$$

and

$$\mathrm{M}_\nu s(t) = \mathrm{e}^{j2\pi\nu t} s(t), \quad t \in \mathbb{R} \tag{8.5}$$

respectively.
The translation operator T_x shifts the signal $s(t)$ by x and the modulation operator M_ν modulates the signal $s(t)$ by ν. The modulation operator corresponds to a shift of the function in the frequency domain, i.e.

$$\mathcal{F}\left(\mathrm{M}_\nu s(t)\right) = \hat{s}(\omega - \nu),$$

where $\mathcal{F}(\cdot)$ is the Fourier transform operator and $\hat{s}(\omega)$ denotes the Fourier transform of $s(t)$. Additionally, the following commutation relation holds [129]

$$\mathrm{T}_x \mathrm{M}_\nu = \mathrm{e}^{-j2\pi x\nu} \mathrm{M}_\nu \mathrm{T}_x. \tag{8.6}$$

The relation (8.6) can be verified by straight forward computation

$$\begin{aligned} \mathrm{T}_x \mathrm{M}_\nu s(t) &= (\mathrm{M}_\nu s)(t-x) \\ &= \mathrm{e}^{j2\pi\nu(t-x)} s(t-x) \\ &= \mathrm{e}^{-j2\pi x\nu} \mathrm{e}^{j2\pi\nu t} s(t-x) \\ &= \mathrm{e}^{-j2\pi x\nu} \mathrm{M}_\nu \mathrm{T}_x s(t). \end{aligned}$$

Finally, we can write (8.1) using the translation and modulation operators as defined in (8.4) and (8.5), respectively and we obtain

$$(\mathrm{H}s)(t) = \iint_{\mathbb{R}^2} \eta(\tau, \nu) \mathrm{T}_\tau \mathrm{M}_\nu s(t)\, d\tau\, d\nu\,. \tag{8.7}$$

8.1. Operator sampling

The identification of LTV systems, that is the problem of identifying the spreading function η in (8.1), has been the focus of research throughout history [8, 11, 13, 16, 107, 128, 130]. The main result of the identification problem sates that the channel H can be identified if the support area of the spreading

function is less then one. If, on the other hand, the support area is larger than one, channel identification is impossible. The channel can be identified by sending a periodic delta train

$$s_{\mathbf{c}}(t) = \sum_{r=0}^{L-1} \mathbf{c}[r] \sum_{k\in\mathbb{Z}} \delta(t - kLT + rT) \; . \tag{8.8}$$

with a specific vector $\mathbf{c} \in \mathbb{C}^L$ as a pilot over the channel. Then η can be determined form the channel output $\mathrm{H}s_{\mathbf{c}}$ at the receiver.

We now review some results on operator sampling developed in [13, 107, 128, 130]. We mainly follow the setup and notation in [130]. We consider the spreading function representation of a bounded linear operator $\mathrm{H} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$, which is given by (8.1). We denote the set of bounded linear operators by $\mathcal{L}(L^2(\mathbb{R}), L^2(\mathbb{R}))$.

Moreover, we note that any bounded linear operator on $L^2(\mathbb{R})$ has such a spreading function representation [130]. Let $M \subset \mathbb{R}^2$ be a subset in the delay-Doppler plain. Then $OPW(M)$ denotes the *operator Paley-Wiener space* of all operators of the form (8.1) with a spreading function η_{H} whose support set is contained in M, that is,

$$OPW(M) = \{\mathrm{H} \in \mathcal{L}(L^2(\mathbb{R}), L^2(\mathbb{R})) : \mathrm{supp}(\eta) \subset M\}.$$

In the cases we consider M is always compact. We say that $OPW(M)$ is *identifiable* using a single pilot signal s if there exist constants $A, B > 0$ such that

$$A \left\|H\right\|_{L^2(\mathbb{R})\to L^2(\mathbb{R})} \leq \left\|Hs\right\|_{L^2(\mathbb{R})} \leq B \left\|H\right\|_{L^2(\mathbb{R})\to L^2(\mathbb{R})} ,$$

for all $H \in OPW(M)$.

Operator identification is referred to *operator sampling* if a discretely supported distribution s is used, that is if pilot signal of the form (8.8) is used.

The *Zak transform* of $f \in L^2(\mathbb{R})$ with parameter T is defined to be

$$\mathcal{Z}^T f(t,\nu) = \sum_{r\in\mathbb{Z}} f(t - rT)\, \mathrm{e}^{j2\pi rT\nu} , \quad t,\nu \in \mathbb{R}. \tag{8.9}$$

This transformation plays a key role in the analysis of operator sampling. Based on (8.9), one defines the L-dimensional vector valued Zak transform

with parameter LT as

$$\mathcal{Z}_L^{LT} f(t,\nu) = \begin{pmatrix} \mathcal{Z}^{LT} f(t,\nu) \\ \mathcal{Z}^{LT} f(t+T,\nu) \mathrm{e}^{-j2\pi T\nu} \\ \vdots \\ \mathcal{Z}^{LT} f(t+(L-1)T,\nu) \mathrm{e}^{-j2\pi (L-1)T\nu} \end{pmatrix}. \tag{8.10}$$

It is easy to check that the vectorized Zak transform is $(LT, \frac{1}{T})$-quasi-periodic, i.e., $\mathcal{Z}_L^{LT} f(t+LT,\nu) = \mathrm{e}^{j2\pi LT\nu} \mathcal{Z}_L^{LT} f(t,\nu)$ and $\mathcal{Z}_L^{LT} f(t,\nu+\frac{1}{T}) = \mathcal{Z}_L^{LT} f(t,\nu)$. Additionally, one has the norm equality

$$\|f\|_{L^2(\mathbb{R}^2)} = \sqrt{T} \left\| \left\| \mathcal{Z}_L^{LT} f \right\|_2 \right\|_{L^2\left([0,T]\times[0,\frac{1}{LT}]\right)}.$$

Next, set $\Omega = \frac{1}{LT}$ and define the $(LT, L\Omega)$-*quasi-periodic spreading function* $\eta(\tau,\nu)$ by

$$\eta^{(LT,L\Omega)}(\tau,\nu) = \Omega \sum_{p\in\mathbb{Z}} \sum_{q\in\mathbb{Z}} \eta(\tau+pLT, \nu+pL\Omega) \mathrm{e}^{-j2\pi pLT\nu}. \tag{8.11}$$

The following two lemmas restate key results from operator sampling, which will frequently be needed in later sections. The first lemma shows that the identification problem can be reduced to a finite dimensional linear equation. The translation and modulation operators, $\mathbf{M}^k$, $\mathbf{T}^l$, are defined in Section 4.2.1.

Lemma 8.1.1 ([130])**.** *Let* $\mathrm{H} \in OPW(M)$ *be of the form of* (8.1) *with spreading function* η *and let* $s_{\mathbf{c}}(t)$ *be given as in* (8.8)*. Then for the vectorized Zak transform with parameter* LT *of the channel output* $y = (\mathrm{H}s_{\mathbf{c}})$ *holds*

$$\mathcal{Z}_L^{LT} y(\tau,\nu) = \sum_{k,l=0}^{L-1} \eta^{(LT,L\Omega)}(\tau+kT, \nu+l\Omega) \mathrm{e}^{-j2\pi kT\nu} \, \mathbf{M}^k \mathbf{T}^l \mathbf{c}\,, \tag{8.12}$$

where $\eta^{(LT,L\Omega)}$ *is the* $(LT, L\Omega)$*-quasi-periodic spreading function given in* (8.11)*. Furthermore, rearranging the right-hand side of* (8.12) *to a* vector valued bivariate spreading function $\boldsymbol{\eta}(\tau,\nu) \in \mathbb{C}^{L^2}$ *we get*

$$\mathcal{Z}_L^{LT} y(\tau,\nu) = \mathbf{G}_{\mathbf{c}}\, \boldsymbol{\eta}(\tau,\nu), \tag{8.13}$$

where $\mathbf{G}_{\mathbf{c}} \in \mathbb{C}^{L\times L^2}$ *denotes the Gabor matrix generated by the window vector* $\mathbf{c}$ *as defined in Section 4.2.2.*

To recover $\boldsymbol{\eta}(\tau,\nu)$ from $\mathcal{Z}_L^{LT} y(\tau,\nu)$ in (8.13), one needs that $\boldsymbol{\eta}(\tau,\nu)$ has at

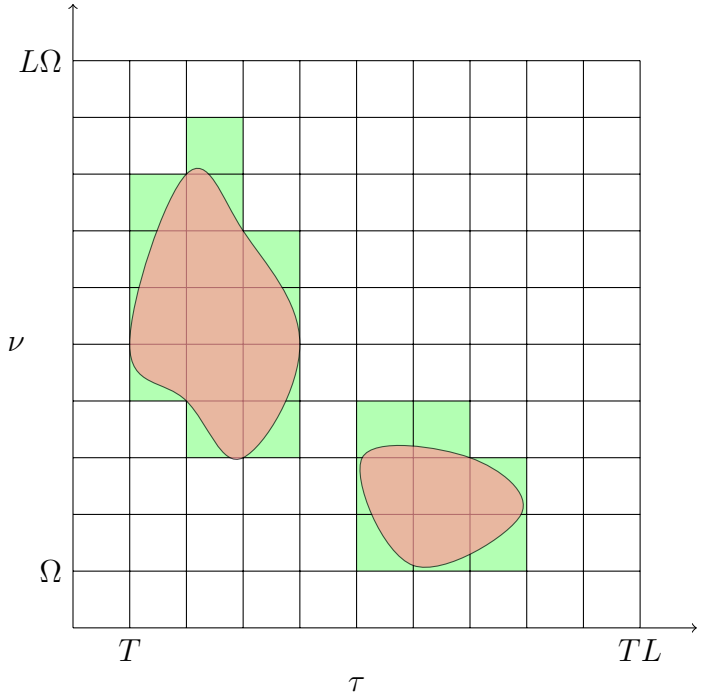

Figure 8.1.: Rectification of the channel support region M (red) in the delay-Doppler (time-frequency) domain. Here a rectification with $L = 10$ is chosen and the covering set U with respect to this rectification is shown in green.

most L nonzero entries and that the corresponding columns of $\mathbf{G_c}$ are linearly independent. As seen in Theorem 4.2.8 and Theorem 4.2.9, there always exists a vector $\mathbf{c} \in \mathbb{C}^L$ such that every L column vectors of $\mathbf{G_c}$ are linearly independent. In fact, the vectors $\mathbf{c} \in \mathbb{C}^L$ that do not satisfy this property form a set of measure zero in $\mathbb{C}^L$. Note that (8.13) has the form of the matrix identification problem already discussed in Section 4.2.3. This observation leads to the following fundamental result in operator sampling.

Definition 8.1.1 (see Definition 2.1 in [13] or [131])**.** *For $K, L \in \mathbb{N}$, let $R_{K,L} = [0, \frac{1}{K}) \times [0, \frac{K}{L})$ and*

$$\mathcal{U}_{K,L} = \left\{ \bigcup_{j=1}^{J} \left(R_{K,L} + \left(\tfrac{k_j}{K}, \tfrac{\ell_j K}{L} \right) \right) : k_j, \ell_j \in \mathbb{Z},\ J \in \mathbb{N} \right\}.$$

Let $M \subset \mathbb{R}^2$ be a bound set and let μ be the Lebesgue measure on $\mathbb{R}^2$. The

inner *and* outer content *of* M *are defined by*

$$\mathcal{A}^-(M) = \sup\{\mu(U) : U \subset M \text{ and } U \in \mathcal{U}_{K,L} \text{ for some } K, L \in \mathbb{N}\}$$

and

$$\mathcal{A}^+(M) = \inf\{\mu(U) : U \supset M \text{ and } U \in \mathcal{U}_{K,L} \text{ for some } K, L \in \mathbb{N}\},$$

respectively. Clearly, it holds $\mathcal{A}^-(M) \leq \mathcal{A}^+(M)$*. If* $\mathcal{A}^-(M) = \mathcal{A}^+(M)$*, then* M *is a Lebesgue measurable set with* $\mu(M) = \mathcal{A}(M) := \mathcal{A}^-(M) = \mathcal{A}^+(M)$ *and we call* M *a* Jordan domain *with* Jordan content $\mathcal{A}(M)$*. In general, a Jordan domain is a bounded Lebesgue measurable set whose boundary is a Lebesgue zero set (see e.g., [13, Proposition 2.2] or [131]).*

Lemma 8.1.2 ([13])**.** *Let* $M \subset [0,\infty)^2$ *be a Jordan domain. If* $\mathcal{A}(M) < 1$*, then* $OPW(M)$ *is identifiable, that is, any operator* H *with* $\operatorname{supp}(\eta_{\mathrm{H}}) \subset M$ *is identifiable, using an input signal* $s_{\mathbf{c}}(t)$ *of the form* (8.8)*. If* $\mathcal{A}(M) > 1$*, then* $OPW(M)$ *is not identifiable by means of any single input signal.*

In Lemma 8.1.2, the support set of the spreading function is assumed to be a subset of a known set M. A similar result was obtained in [107] for unknown support sets: An operator H with unknown $\operatorname{supp}(\eta)$ is identifiable if and only if $\mathcal{A}(\operatorname{supp}(\eta)) < 1/2$.

8.1.1. Rectification for known support sets M

If the support M of the channel is known, then the parameters $K, L \in \mathbb{N}$ and $\mathbf{c} \in \mathbb{C}^L$ for the input signal $s_{\mathbf{c}}(t)$ in (8.8) are chosen in the following way. First, based on the knowledge of M a rectification of the time-frequency plane is chosen. Since $\mathcal{A}(M) < 1$, there exists some $U \in \mathcal{U}_{K,L}$ with sufficiently large $K \in \mathbb{N}$ and $L \gg K$ such that $U \supset M$, $\mathcal{A}(M) < \mu(U) < 1$ and $M \subset [0, \frac{L}{K}) \times [0, K)$. Setting $T = \frac{1}{K}$ and $\Omega = \frac{K}{L}$, we have $M \subset [0, LT) \times [0, L\Omega)$ and since $\mu(U) < 1$, the set M intersects at most L of the rectangles $[kT, (k+1)T) \times [\ell\Omega, (\ell+1)\Omega)$, $k, \ell = 0, \ldots, L-1$. Note that each rectangle has area $T\Omega = \frac{1}{L}$ and the rectangles that cover M have total area of at most $L \cdot \frac{1}{L} = 1$ [107, Section 2]. For example, Figure 8.1 illustrates such a covering U of a support set M with area $\mathcal{A}(U) = 22/10^2 = 0.22$.
After a rectification is fixed (i.e., $K, L \in \mathbb{N}$ are fixed), we choose $\mathbf{c} \in \mathbb{C}^L$ so that every L column vectors of $\mathbf{G}_{\mathbf{c}}$ are linearly independent. The linear independence of every L columns of $\mathbf{G}_{\mathbf{c}}$ plays an important role in the proof of Lemma 8.1.2, as it allows one to solve $\boldsymbol{\eta}$ from (8.13) whenever the support set of $\boldsymbol{\eta}$ is known and has cardinality at most L.

8.1.2. Fixed Rectification for unknown support sets M

In a practical scenario, the support M of the spreading function will strongly depend on environmental factors, such as the relative movement of the transmitter and receiver. The set M may evolve in time and therefore change between consecutive communications. Nevertheless, one can make some reasonable assumptions on M. For instance, consider short distance communications between two cars on the road. Due to the speed limit of vehicles and the sparse scattering environment, it is reasonable to assume that $M \subset [0, \tau_{\max}) \times [0, \nu_{\max})$ for some $\tau_{\max} > 0$ and $\nu_{\max} > 0$. Then as in Section 8.1.1, we choose a sufficiently fine rectification of $[0, \tau_{\max}) \times [0, \nu_{\max})$, that is, we choose small $T = \frac{1}{K}$ and $\Omega = \frac{K}{L}$ (correspondingly, large $K \in \mathbb{N}$ and $L \gg K$), so that $[0, \tau_{\max}) \times [0, \nu_{\max}) \subset [0, LT) \times [0, L\Omega)$.
To simplify our analysis, we will restrict our attention to LTV channels each of which takes only a limited number of rectangles in the (a priori fixed) rectification to cover its spreading support M. This excludes some pathological sets M which, for a given rectification, intersect every rectangles but have area less than 1; such sets are certainly non-realistic in practice. Note that choosing a finer rectification (thus, a larger $L \in \mathbb{N}$) would increase the class of LTV channels under consideration, but the computational cost involved with (8.13) would also be increased.
With a fixed rectification, the support size of the LTV channels under consideration can be measured by the number of rectangles that are needed to cover M. We will later derive an upper bound on the mentioned number of rectangles for successful message transmission.

Remark 8.1.2. *We would like to emphasize that there is an important difference between the two rectification methods discussed in Section 8.1.1 and 8.1.2. Section 8.1.1 shows how to choose a rectification so that M is covered by at most L rectangles of area $\frac{1}{L}$. On the other hand, Section 8.1.2 assumes a* fixed rectification *of the estimated feasible region $[0, \tau_{\max}) \times [0, \nu_{\max})$, and restricts the class of LTV channels to those with spreading support covered by a limited number of rectangles in the rectification.*

8.2. Virtual channel representation

The previous Section 8.1 can be seen as more of a mathematical analysis and perspective on LTV systems. In engineering, rather than considering a continuous function, it is more common to interpret the time-varying channel as a multipath propagation and each path is considered to be a point scatterer

[3, 8, 9]. Hence, the channel simply becomes a superposition of multiple paths and each path shifts the signal in time and frequency. Thus, the time-varying transfer function for a channel H for a multi-antenna receiver can be written as

$$\tilde{\sigma}_{\mathrm{H}}(t,\omega) = \sum_{n=0}^{N_P-1} \mathbf{a}\left(\theta_n\right) \tilde{h}_n \, \mathrm{e}^{j2\pi\nu_n t} \mathrm{e}^{-j2\pi\tau_n\omega}, \tag{8.14}$$

wherein N_P is the number of different paths, $\mathbf{a}(\cdot)$ is an N_R dimensional function defined on $[-\pi,\pi)$ so that $\mathbf{a}\left(\theta_n\right) \in \mathbb{C}^{N_R}$ is the array steering vector in the direction θ_n which is associated with the n-th path, and $\tilde{h}_n$ and (τ_n,ν_n) are the complex gain and the delay-Doppler shift associated with the n-th path, respectively [8]. We assume that $(\tau_n,\nu_n) \in [0,\tau_{\max}]\times[0,\nu_{\max}]$ and $\theta_n \in \left[-\frac{\theta_{\max}}{2},\frac{\theta_{\max}}{2}\right)$, where $\theta_{\max} < 2\pi$ is the maximum angle range of the antenna array, and $\tau_{\max}$ and $\nu_{\max}$ are the maximum delay and maximum Doppler spread of the channel, respectively. Note that if the receiver is equipped with only one antenna then the array steering vector $\mathbf{a}$ is simply a scalar equal to 1, $\mathbf{a}\left(\theta_n\right) = 1$.

Although (8.14) is a physically precise description of the channel, it is quite impractical to work with. Instead we are going to consider an approximation of (8.14) called the *virtual channel model* [9] which is easier to work with. The idea is to round the points (τ_n,ν_n,θ_n) to the nearest point on a uniform grid, for instance, with uniform spacing $\Delta\tau$ in time, $\Delta\nu$ in frequency, and $\Delta\theta$ in angle. More precisely, we approximate $\tilde{\sigma}_{\mathrm{H}}\left(t,\omega\right)$ by

$$\sigma_{\mathrm{H}}\left(t,\omega\right) = \sum_{d=-D}^{D-1} \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} \mathbf{a}\left(d\Delta\theta\right) h_{k,m,d} \mathrm{e}^{-j2\pi k\Delta\tau\omega} \, \mathrm{e}^{j2\pi m\Delta\nu t} \tag{8.15}$$

where $D,K,M \in \mathbb{N}$ satisfy $\theta_{\max} \leq 2D\cdot\Delta\theta$, $\tau_{\max} \leq K\cdot\Delta\tau$, and $\nu_{\max} \leq M\cdot\Delta\nu$. Moreover, $h_{k,m,d}$ is the sum of all complex gains $\tilde{h}_n$ such that

$$(\tau_n,\nu_n,\theta_n) \in [k\Delta\tau,(k+1)\Delta\tau)\times[m\Delta\nu,(m{+}1)\Delta\nu)\,[d\Delta\theta,(d{+}1)\Delta\theta) \ .$$

Following the notation in [8], we write

$$h_{k,m,d} = \sum_{n\in\mathcal{S}_{\tau,k}\cap\mathcal{S}_{\nu,m}\cap\mathcal{S}_{\theta,d}} \tilde{h}_n,$$

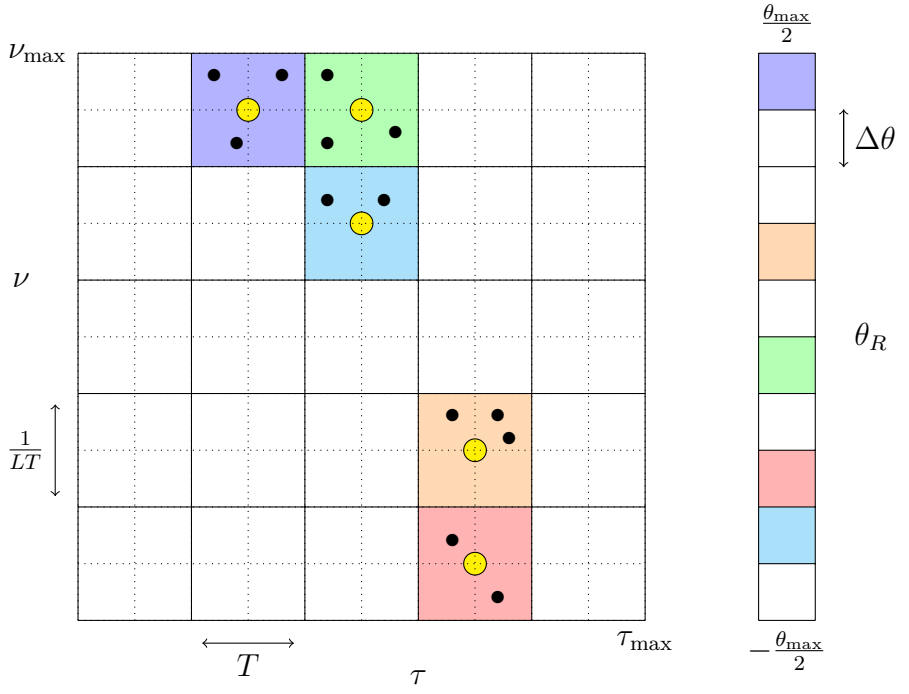

Figure 8.2.: Here $D = M = K = L = 5$ is chosen and it is assumed that $\tau_{\max} = LT$ and $\nu_{\max} = \frac{1}{T}$. On the right the direction of arrivals of each propagation path is shown with the corresponding colours.

where

$$\begin{aligned}\mathcal{S}_{\theta,d} &= \{n \ : \ \theta_n \in [d\Delta\theta, (d{+}1)\Delta\theta)\},\\ \mathcal{S}_{\tau,k} &= \{n \ : \ \tau_n \in [k\Delta\tau, (k{+}1)\Delta\tau)\},\\ \mathcal{S}_{\nu,m} &= \{n \ : \ \nu_n \in [m\Delta\nu, (m{+}1)\Delta\nu)\}.\end{aligned}$$

Replacing $\tilde{\sigma}_{\mathrm{H}}(t,\omega)$ in (8.14) with its approximation $\sigma_{\mathrm{H}}(t,\omega)$ in (8.15) gives

$$\left(\mathrm{H}f\right)(t) = \sum_{d=-D}^{D-1} \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} \mathbf{a}\left(d\Delta\theta\right) h_{k,m,d} f\left(t - k\Delta\tau\right) \mathrm{e}^{j2\pi m\Delta\nu t}. \tag{8.16}$$

In Figure 8.2 the virtual channel representation is illustrated. Each dot corresponds to a delay-Doppler shift of the signal caused by a propagation path. All paths which are in the same delay-Doppler region (i.e. in the same rectangle) are substituted by a virtual path which is assumed to be placed in the middle of the rectangle, illustrated by the yellow dots in Figure 8.2. On the right hand side of the figure each direction of arrival corresponding to each propagation

path is shown with the corresponding colours.

From virtual channel representation to matrix identification

Interestingly, the problem of channel identification that arises from (8.16) is the same as we get in Section 8.1 where the finite Gabor frames play a crucial role.
For the next part we assume a single antenna, i.e. $\mathbf{a}$ is a scalar and is equal to 1. Consider the channel model in (8.16) with $\Delta\nu = \frac{1}{TL}$ and $\Delta\tau = T$, where L is a sufficiently large and T is chosen sufficiently small so that $\tau_{\max} < TL$ and $\nu_{\max} < \frac{1}{T}$ (see Figure 8.2). With these parameters, (8.16) becomes

$$\tilde{y}(t) = (\mathrm{H}f)(t) = \sum_{k=0}^{L-1}\sum_{m=0}^{L-1} h_{k,m}\, f(t-kT)\, \mathrm{e}^{j2\pi m\frac{1}{TL}t}. \tag{8.17}$$

Now, we choose the input f as

$$f(t) = \sum_{\ell=0}^{L-1} z_\ell\, \psi_\epsilon(t-\ell T)\,, \quad t\in\mathbb{R}\,, \tag{8.18}$$

where ψ_ϵ is a C^∞ function which is supported on $[-\epsilon,\epsilon]$ and has value 1 on $[-\epsilon/2,\epsilon/2]$. Here, the coefficient vector $\mathbf{z} = \{z_\ell\}_{l=0}^{L-1} \in \mathbb{C}^L$ carries the pilot signal for the purpose of channel identification. Substituting (8.18) into (8.17) gives

$$\begin{aligned}
\tilde{y}(t) &= \sum_{k=0}^{L-1}\sum_{m=0}^{L-1}\sum_{\ell=0}^{L-1} h_{k,m}\, \mathbf{z}\,[\ell]\ \psi_\epsilon(t-\ell T-kT)\, \mathrm{e}^{j2\pi m\frac{1}{TL}t} \\
&= \sum_{k=0}^{L-1}\sum_{m=0}^{L-1}\sum_{n=k}^{k+L-1} h_{k,m}\, \mathbf{z}\,[n-k]\ \psi_\epsilon(t-nT)\, \mathrm{e}^{j2\pi m\frac{1}{TL}t} \\
&= \sum_{k=0}^{L-1}\sum_{n=k}^{k+L-1} \Big(\sum_{m=0}^{L-1} h_{k,m}\, \mathbf{z}\,[n-k]\ \mathrm{e}^{j2\pi m\frac{1}{TL}t}\Big)\, \psi_\epsilon(t-nT) \\
&= \sum_{n=0}^{L-1}\sum_{k=0}^{n} \Big(\sum_{m=0}^{L-1} h_{k,m}\, \mathbf{z}\,[n-k]\ \mathrm{e}^{j2\pi m\frac{1}{TL}t}\Big)\, \psi_\epsilon(t-nT) \\
&\quad + \sum_{n=L}^{2L-2}\sum_{k=n-L+1}^{L-1} \Big(\sum_{m=0}^{L-1} h_{k,m}\, \mathbf{z}\,[n-k]\ \mathrm{e}^{j2\pi m\frac{1}{TL}t}\Big)\, \psi_\epsilon(t-nT)
\end{aligned}$$

which is a linear combination of $\psi_\epsilon(t-nT)$, $n=0,\dots,2L-2$. Thus, the whole transmission cycle is $(2L-2)T$ long, where the interval $[TL,(2L-2)T]$ can be interpreted as a guard interval. By means of a convolution filter, one can easily read off the coefficient

$$\tilde{y}_n = \begin{cases} \sum_{k=0}^{n}\sum_{m=0}^{L-1} h_{k,m}\,\mathbf{z}\,[n-k]\;\mathrm{e}^{j2\pi m\frac{1}{TL}t}, & 0\le n\le L-1,\\ \sum_{k=n-L+1}^{L-1}\sum_{m=0}^{L-1} h_{k,m}\,\mathbf{z}\,[n-k]\;\mathrm{e}^{j2\pi m\frac{1}{TL}t}, & L\le n\le 2L-2, \end{cases}$$

of $\psi_\epsilon(t-nT)$ from $\tilde{y}(t)$. However, we note that

$$\begin{aligned} y(t) &:= \Big(\tilde{y}(t)+\tilde{y}(t+TL)\Big)\cdot \mathbf{1}_{[0,LT)}(t)\\ &= \sum_{n=0}^{L-1}\sum_{k=0}^{L-1}\Big(\sum_{m=0}^{L-1} h_{k,m}\,\mathbf{z}\,[n-k]\;\mathrm{e}^{j2\pi m\frac{1}{TL}t}\Big)\,\psi_\epsilon(t-nT) \end{aligned}$$

exhibits a simpler and symmetric expression, where the coefficient of $\psi_\epsilon(t-nT)$ is given by

$$y_n = \tilde{y}_n + \tilde{y}_{n+L} = \sum_{k=0}^{L-1}\sum_{m=0}^{L-1} h_{k,m}\,\mathbf{z}\,[n-k]\;\mathrm{e}^{j2\pi\frac{mn}{L}}\,,$$

for every $0\le n\le L-1$. Using the notation from Section 4.2.3, this can be expressed as

$$\mathbf{y} = \sum_{k=0}^{L-1}\sum_{m=0}^{L-1} h_{k,m}\,\mathbf{M}^m\mathbf{T}^k\mathbf{z} = \mathbf{H}\,\mathbf{z}, \tag{8.19}$$

where $\mathbf{y}=\{y_n\}_{n=0}^{L-1}\in\mathbb{C}^L$, $\mathbf{z}=\{z_\ell\}_{\ell=0}^{L-1}\in\mathbb{C}^L$ and $\mathbf{H}=\sum_{k=0}^{L-1}\sum_{m=0}^{L-1} h_{k,m}\,\mathbf{M}^m\mathbf{T}^k$. In summary, the channel output of H in (8.16) with respect to an input signal of the form (8.18) can be reformulated as the output of a finite dimensional channel $\mathbf{H}\in\mathbb{C}^{L\times L}$ with respect to $\mathbf{z}\in\mathbb{C}^L$.
The coefficients $\{h_{k,m}\}$ in (8.19) correspond to the values of the spreading function in (8.13). In the following sections when we will consider channels we will mainly referring to the finite dimensional problem as stated in (8.13) and (8.19).

9. Data Transmission over LTV Channels: Single Antenna Receiver

In this section our aim is to transmit data over a LTV channel without prior information on the channel. We assume that the receiver is equipped with a single antenna. This section has been partially published in [31, 34].

9.1. Finite chirp signals

Chirp signals posses almost optimal coherence. Therefore, in [60] they were used to construct deterministic measurement matrices.
Let $L \geq 3$ be a prime number, the chirp signal $\mathbf{c}_{mL+r} \in \mathbb{C}^L$ with *base frequency* $m \in \mathbb{Z}_L$ and *chirp rate* $r \in \mathbb{Z}_L$ is defined by

$$\mathbf{c}_{mL+r}[x] = \tfrac{1}{\sqrt{L}} \, \mathrm{e}^{j\frac{2\pi}{L}r} \mathrm{e}^{j\frac{2\pi}{L}mx} \mathrm{e}^{j\frac{2\pi}{L}rx^2} \;, \quad x \in \mathbb{Z}_L \;. \tag{9.1}$$

Note that there are exactly L^2 chirp vectors in $\mathbb{C}^L$. Their inner products satisfy [60, 132] the relation

$$|\langle \mathbf{c}_{mL+r_1}, \mathbf{c}_{nL+r_2} \rangle| = \begin{cases} 1 & \text{if } \; n = m \;\; \text{and} \;\; r_1 = r_2 \\ 0 & \text{if } \; n \neq m \;\; \text{and} \;\; r_1 = r_2 \\ \frac{1}{\sqrt{L}} & \text{otherwise} \end{cases} . \tag{9.2}$$

It can be seen from (9.2) that the set of all L^2 chirp signals in $\mathbb{C}^L$ can be partitioned into L sets, each of which contains exactly L chirp signals with the same chirp rate and which are mutually orthogonal.
The eigenvectors of time-frequency shifts are chirp signals, in fact, the following lemma shows that the chirp rate of a chirp signal is invariant under any time-frequency shift. This lemma will play a crucial role in the design of our transmit signals for the single antenna receiver.

Lemma 9.1.1. *Let L be a prime number and let $\omega = \mathrm{e}^{j\frac{2\pi}{L}}$. Then*

$$\mathbf{M}^{\nu}\mathbf{T}^{\tau}\mathbf{c}_{mL+r} = \omega^{\kappa}\mathbf{c}_{nL+r} \qquad \text{for all } \nu, \tau \in \mathbb{Z}_L$$

with $\kappa = (r\tau - m)\,\tau$ and $n = m + \nu - 2r\tau$.
Moreover if $\nu - 2r\tau \equiv 0 \mod L$ then $\mathbf{c}_{mL+r}$ is an eigenvector of $\mathbf{M}^{\nu}\mathbf{T}^{\tau}$ with the eigenvalue ω^{κ}.

Proof. First observe that

$$\left(\mathbf{M}^{\nu}\mathbf{T}^{\tau}\mathbf{c}_{mL+r}\right)[x] = \frac{1}{\sqrt{L}}\omega^{\nu x}\omega^{r}\omega^{m(x-\tau)}\omega^{r(x-\tau)^2}\ . \tag{9.3}$$

For $\nu - r2\tau = 0$, we get in (9.3)

$$\frac{1}{\sqrt{L}}\omega^{r+mx+rx^2}\omega^{(\nu-r2\tau)x}\omega^{(r\tau-m)\tau} = \omega^{\kappa}\mathbf{c}_{mL+r}\,(x)$$

where we set $\kappa = (r\tau - m)\,\tau$. For $\nu - r2\tau \neq 0$ we get for (9.3)

$$\frac{1}{\sqrt{L}}\omega^{r+(m+\nu-2r\tau)x+rx^2}\omega^{\kappa} = \omega^{\kappa}\mathbf{c}_{nL+r}\,[x]$$

with $\kappa = (r\tau - m)\,\tau$ and $n = m + \nu - 2r\tau$. □

Remark 9.1.1. *A different method to obtain chirp signals is by simply computing the eigenvalue decompositions of $\mathbf{M}\mathbf{T}^{k}$ for $k = 1, \ldots, L-1$ where L is the dimension and is an odd prime number. Then each eigenvector corresponds to a chirp signal. This stems from the fact that the set of all time-frequency shift can be split into $L+1$ commutative groups of L elements when the dimension is a prime number [83].*

Define the matrix $\mathbf{C}_r = [\,\mathbf{c}_r,\ \ldots,\ \mathbf{c}_{(L-1)L+r}\,] \in \mathbb{C}^{L\times L}$ for $r = 0, \ldots L-1$, whose columns consist of all chirp signals with chirp rate r. Because of (9.2), the columns of $\mathbf{C}_r$ are mutually orthogonal and so $\mathbf{C}_r$ is orthogonal. Therewith, let $\mathbf{U} = [\mathbf{C}_0, \mathbf{C}_1, ..., \mathbf{C}_{L-1}] \in \mathbb{C}^{L\times L^2}$ be the concatenation of all matrices $\mathbf{C}_r$, so that it consists of all L^2 chirp signals in dimension L as columns.

9.2. Gabor & chirp measurement matrix

The Gabor matrix generated with the Alltop vector (5.1) and the matrix $\mathbf{U}$ whose columns are formed by chirp signals are known to be suitable measurement

matrices for CS [27, 70, 86]. Consider the measurement matrix $\mathbf{\Phi} = [\mathbf{G}_{\boldsymbol{\alpha}}|\mathbf{U}]$, which is the concatenation of the Gabor matrix $\mathbf{G}_{\boldsymbol{\alpha}}$ generated by the Alltop window $\boldsymbol{\alpha}$ and the matrix $\mathbf{U}$ described above. We will show that $\mathbf{\Phi}$ possesses small coherence and is therefore a good measurement matrix. For this we need the following lemma, which is a special case of Theorem 5.38 in [94].

Lemma 9.2.1. *Let $L \geq 5$ be a prime number, set $\omega = \mathrm{e}^{j\frac{2\pi}{L}}$ and let $f(x) = c_3x^3 + c_2x^2 + c_1x + c_0$ be a polynomial of 3rd degree with $c_i \in \{0, ..., L-1\}$ for all i and with $c_3 \neq 0$. Then*

$$\left| \sum_{x \in \mathbb{Z}_L} \omega^{f(x)} \right| \leq 2\sqrt{L}\,. \tag{9.4}$$

Based on this lemma, we can derive an upper bound for the coherence of the measurement matrix $\mathbf{\Phi}$.

Lemma 9.2.2. *Consider the matrix $\mathbf{\Phi} = [\mathbf{G}_{\boldsymbol{\alpha}}|\mathbf{U}]$ then its coherence is upper bounded by $\mu(\mathbf{\Phi}) \leq \frac{2}{\sqrt{L}}$.*

Proof. Since $\mu(\mathbf{U}) = \frac{1}{\sqrt{L}}$ and $\mu(\mathbf{G}_{\boldsymbol{\alpha}}) = \frac{1}{\sqrt{L}}$, we only need to consider the inner products $|\langle \mathbf{z}, \mathbf{x} \rangle|$ where $\mathbf{z}$ is a column from $\mathbf{G}_{\boldsymbol{\alpha}}$ and $\mathbf{x}$ is a column of $\mathbf{U}$. To this end we set $\omega = \mathrm{e}^{j\frac{2\pi}{L}}$ and observe

$$\begin{aligned} &|\langle \mathbf{M}^{\nu}\mathbf{T}^{\tau}\boldsymbol{\alpha}, \mathbf{c}_{mL+r} \rangle| = |\langle \boldsymbol{\psi}, \mathbf{c}_{nL+r} \rangle| \\ &= \frac{1}{L}\Big| \sum_{t \in \mathbb{Z}_L} \omega^{-r}\omega^{-(rt^2+nt)}\omega^{t^3} \Big| = \frac{1}{L}\Big| \sum_{t \in \mathbb{Z}_L} \omega^{t^3 - rt^2 - nt - r} \Big| \leq \frac{2}{\sqrt{L}} \end{aligned}$$

where we used Lemmas 9.1.1 and 9.2.1. □

It is known that the Welch bound in (2.1) is achieved if and only if the vectors $\varphi_0, \ldots, \varphi_{N-1}$ form an equiangular tight frame for $\mathbb{C}^L$, in which case we have $N \leq L^2$ [37, 65]. Note that the Welch bound with $N = L^2$ is $\frac{1}{\sqrt{L+1}}$. Since $\mathbf{\Phi}$ has $2L^2$ column vectors in $\mathbb{C}^L$, Lemma 9.2.2 indicates that the coherence of $\mathbf{\Phi}$ must be almost optimal up to a constant factor.

9.3. Message transmission

Now we are finally ready to state the main result of this section.

Theorem 9.3.1. *Given a LTV channel* H *with $\mathbf{H} \in \mathbb{C}^{L \times L}$ as deduced in* (8.17) *with support* $\mathrm{supp}(\mathbf{H})$ *in the delay-Doppler domain, where $L \geq 5$ is a prime*

number. If $\mathbf{m} = (m_1, \ldots, m_Q)^T$ *is an arbitrary message whose length satisfies*

$$(1+Q)\,|\,\mathrm{supp}(\mathbf{H})| < \sqrt{\frac{L}{32}} + \frac{1}{2}\,, \tag{9.5}$$

then there exists a signal $f(t)$ *of the form* (8.18)*, which allows us to identify the channel* $\mathbf{H}$ *and to recover the message* $\mathbf{m}$ *from the channel output* $(\mathrm{H}f)\,(t)$*.*

Note that $\mathrm{supp}\,(\mathbf{H})$ is with respect to the representation of $\mathbf{H}$ in the basis of time-frequency shift matrices, i.e. $\mathrm{supp}(\mathbf{H})$ are the indices of all non-negative coefficients obtained by

$$\mathbf{h}\,[\lambda] = \langle \mathbf{A}, \pi\,(\lambda)\rangle_{\mathrm{HS}}\ , \quad \text{for}\ \ \lambda \in \mathbb{Z}_L \times \mathbb{Z}_L\,.$$

We see that the product of message length and channel support is upper bounded by a constant. So there is a trade-off between the message length and the support size of the channel. If the channel support $|\,\mathrm{supp}(\mathbf{H})|$ is sufficiently small, an additional message can be transmitted over the channel, alongside the pilot for channel estimation, and reconstructed at the receiver. The bound in (9.5) is not optimal. The numerical results will show that this bound is rather pessimistic. Indeed, this observation is not surprising since coherence based methods from CS are applied to deduce (9.5). These methods are limited by the Welch bound, which is also known as the quadratic bottle neck in the CS literature. Simulation results indicate that $Q\,|\,\mathrm{supp}(\mathbf{H})| \sim L/\log(...)$, i.e. the product of message length and channel support size is proportional to L up to some logarithmic factor.
The following proof of Theorem 9.3.1 is constructive.

Proof of Theorem 9.3.1. Our first aim is to define $\mathbf{z}$ in the transmit signal f given by

$$f(t) = \sum_{\ell=0}^{L-1} z_\ell\, \psi_\epsilon(t - \ell T)\,. \quad t \in \mathbb{R}\,,$$

First, we choose a set $\mathcal{Q} = \{i_1, \ldots, i_Q\} \subset \{0, \ldots, L-1\}$ of $Q < L$ arbitrary distinct indices. Then we define the vector $\mathbf{z} \in \mathbb{C}^L$ for the transmit signal by

$$\mathbf{z} = \boldsymbol{\alpha} + \sum_{q=1}^{Q} m_q\, \mathbf{c}_{i_q} = \boldsymbol{\alpha} + \sum_{r=0}^{L-1} \widetilde{m}_r\, \mathbf{c}_r, \tag{9.6}$$

where $\boldsymbol{\psi} \in \mathbb{C}^L$ is the Alltop vector, defined in (5.1), $\mathbf{c}_r \in \mathbb{C}^L$ are chirp signals (with base frequency zero) as defined in (9.1), and where we set $\widetilde{m}_r = 0$ if

$r \notin \mathcal{Q}$ to get the right hand side expression. Now, let $\mathbf{y}$ be the channel output obtained as in (8.19) where the input signal is constructed using (9.6). Then by Lemma 9.1.1 we have

$$\begin{aligned}\mathbf{y} = \mathbf{H}\mathbf{z} &= \sum_{k,\ell=0}^{L-1} h_{k,l}\mathbf{M}^{\ell}\mathbf{T}^{k}\left(\boldsymbol{\alpha} + \sum_{r=0}^{L-1}\widetilde{m}_r\,\mathbf{c}_r\right)\\ &= \sum_{k,\ell=0}^{L-1} h_{k,\ell}\,\mathbf{M}^{\ell}\mathbf{T}^{k}\boldsymbol{\alpha} + \sum_{r=0}^{L-1}\sum_{k,\ell=0}^{L-1} h_{k,\ell}\,m_r\,\mathrm{e}^{j\frac{2\pi}{L}k^2 r}\,\mathbf{c}_{(\ell-2kr)L+r}\\ &= \mathbf{G}_{\boldsymbol{\alpha}}\mathbf{h} + \sum_{r=0}^{L-1}\mathbf{C}_r\widetilde{\mathbf{m}}_r = \mathbf{G}_{\boldsymbol{\alpha}}\mathbf{h} + \mathbf{U}\widetilde{\mathbf{m}} = \boldsymbol{\Phi}\begin{bmatrix}\mathbf{h}\\ \widetilde{\mathbf{m}}\end{bmatrix}, \end{aligned} \tag{9.7}$$

where $\mathbf{h} = \{h\}_{k,\ell=0}^{L-1} \in \mathbb{C}^{L^2}$, $\boldsymbol{\Phi} = [\mathbf{G}_{\boldsymbol{\alpha}}|\mathbf{U}]$ is the measurement matrix constructed in Section 9.2 and where the vectors $\widetilde{\mathbf{m}}_r \in \mathbb{C}^L$ are defined by

$$\widetilde{\mathbf{m}}_r[p] = \widetilde{m}_r \sum_{\substack{k,\ell=0\\ \ell-2kr=p \mod L}}^{L-1} h_{k,\ell}\,\mathrm{e}^{j\frac{2\pi}{L}k^2 r}, \quad p = 0, 1, \ldots, L-1, \tag{9.8}$$

and where $\widetilde{\mathbf{m}} \in \mathbb{C}^{L^2}$ is given by $\widetilde{\mathbf{m}} = \left[\widetilde{\mathbf{m}}_0^{\mathrm{T}}, \ldots, \widetilde{\mathbf{m}}_{L-1}^{T}\right]^{\mathrm{T}}$. Note that in particular $\widetilde{\mathbf{m}}$ is block-sparse in the sense that all vectors $\widetilde{\mathbf{m}}_r$ with $r \notin \mathcal{Q}$ are identical to zero. In fact, we have

$$|\operatorname{supp}(\widetilde{\mathbf{m}})| \leq Q \cdot |\operatorname{supp}(\mathbf{h})|,$$

and therefore

$$\left|\operatorname{supp}\left(\left[\mathbf{h}^{\mathrm{T}}\,|\,\widetilde{\mathbf{m}}^{\mathrm{T}}\right]^{T}\right)\right| \leq (1+Q)\cdot|\operatorname{supp}(\mathbf{h})| =: k.$$

So (9.7) is an overdetermined system of linear equations for the s-sparse vector $[\mathbf{h}^{\mathrm{T}}|\widetilde{\mathbf{m}}^{\mathrm{T}}]^{\mathrm{T}}$. According to Lemma 9.2.2 and Theorem 2.4.2, the $2k$-th RIC of measurement matrix $\boldsymbol{\Phi}$ satisfies

$$\begin{aligned}\delta_{2k} &\leq (2k-1)\,\mu(\boldsymbol{\Phi}) < (2k-1)\frac{2}{\sqrt{L}}\\ &\leq [2(1+Q)\cdot|\operatorname{supp}(\mathbf{h})| - 1]\,\frac{2}{\sqrt{L}}\,.\end{aligned}$$

If the message length Q satisfies (9.5), one obtains $\delta_{2k} < 1/\sqrt{2}$ and Theorem 2.3.1 shows that (9.7) has a unique solution $\left[\mathbf{h}^{\mathrm{T}}\,|\,\widetilde{\mathbf{m}}^{\mathrm{T}}\right]^{T}$. Then the message

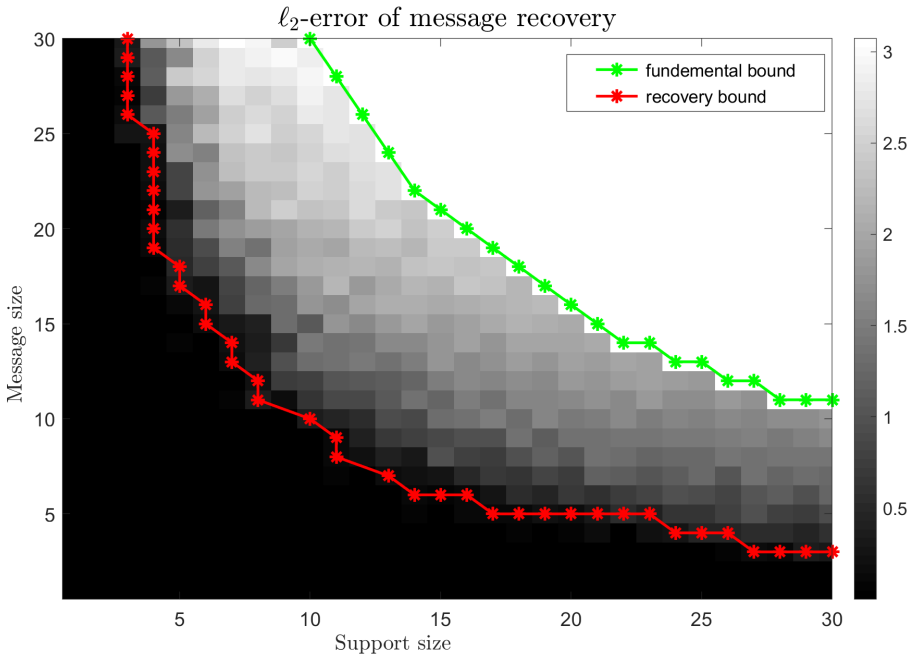

Figure 9.1.: Message recovery rate for $L = 307$.

$\mathbf{m} = (m_1, \ldots, m_Q)^{\mathrm{T}}$ is obtained by solving (9.8). □

9.4. Numerical experiments

Figure 9.1 illustrates the trade-off between the support size of the channel and the number of transmitted messages as given by (9.5). The underlying transmission scheme is discussed in Section 9.3. The dimension L was set to $L = 307$. The support of the channel coefficients were chosen uniformly random over the index set $\{0, \ldots L^2 - 1\}$ and the values were chosen as Gaussian random variables with mean 0 and variance 1. For the messages a set of Q distinct integers were chosen uniformly random over $\{0 \ldots L - 1\}$ and each message entry was also chosen as a complex Gaussian random variable with mean 0 and variance 1. For each channel realization, we have generated four different messages and averaged the ℓ_2-error of the recovered messages. Each data point in Figure 9.1 is averaged over 108 channel realizations hence a total of 432 different messages. Orthogonal Matching Pursuit (OMP) was chosen as the recovery algorithm to solve (9.7). OMP has recovery guarantees similar to the optimization problem given in (P_1) [37].

In Figure 9.1, black represents a small recovery error and white represents a large recovery error. The green line marks a fundamental bound where the message recovery cannot be achieved as the degrees of freedom involved with the size of the message and the channel support exceeds the ambient dimension $L = 307$. The red line marks the bound where an average error of 1% is achieved.

Figure 9.1 shows that the actual recovery rate is far better than the theoretical guarantee obtained in Theorem 9.3.1. Theorem 9.3.1 only guarantees the recovery of a message with a size of at most $2 = \lfloor(\sqrt{307} - \sqrt{32})/\sqrt{32}\rfloor$, which is much smaller than the experimental result.

The message size increases with decreasing support size. An intuitive explanation of this observation is that the sparsity in (9.7) increases multiplicatively with the message size. Hence, each message adds additional support with the same cardinality of the channel support to the signal which has to be recovered.

10. Data Transmission over LTV Channels: Multi-Antenna Receiver

In this section our aim is to transmit data over a LTV channel without prior information on the channel. The transmit signal contains both, the message and a pilot signal to estimate the wireless channel (similar to Section 9. We assume that the receiver is equipped with multiple antennas. That allows the receiver in a first step to decompose the received signal into its propagation paths by beamforming. In a second step the receiver then reconstructs the delay-Doppler shifts and the corresponding path gain for each direction and decodes the message with the channel knowledge.
This section has been partially published in [32, 34].

10.1. Multi-antenna receiver design

In the multi-antenna case we consider the channel model given by

$$y_k(t) = \sum_{n=1}^{N_P} a_k(\theta_n) \iint_{\mathbb{R}\times\mathbb{R}} h_n(\tau,\nu)\, \mathrm{e}^{j2\pi\nu(t-\tau)} f(t-\tau) d\tau d\nu \tag{10.1}$$

where $a_k(\theta)$ is the kth element in the array steering vector, i.e., $\mathbf{a}(\theta) = [a_1(\theta), \ldots, a_{N_R}(\theta)]^T \in \mathbb{C}^{N_R}$, and h_n is the spreading function of the channel associated with the n-th path. We use an input signal similar to (8.8). To this end, we define, analogously to Section 9, the message symbols $\mathbf{m} = [m_1, \ldots, m_Q]^{\mathrm{T}}$ with $Q \in \mathbb{N}$. Then the transmit signal has the form

$$\begin{aligned} s_{\mathbf{m}}(t) &= \sum_{r=0}^{L-1} \mathbf{z}[r] \sum_{k\in\mathbb{Z}} \delta(t - kLT + rT) \\ \text{with} \qquad \mathbf{z} &= \mathbf{p} + \textstyle\sum_{q=1}^{Q} m_q\, \mathbf{a}_q \ . \end{aligned} \tag{10.2}$$

Therein $\mathbf{p} \in \mathbb{C}^L$ is a pilot signal and $\mathbf{a}_q \in \mathbb{C}^L$ are auxiliary signals, which serve as carrier signals for the message symbols. Both, $\mathbf{p}$ and $\{\mathbf{a}_q\}_{q=1}^Q$ are also known at the receiver side. So similarly as in the channel identification problem discussed in Section 8.1, we use a weighted delta train, but now we modulate not only a pilot signal onto this delta train but also our message symbols. Moreover, our primarily goal is not to identify the channel but to recover the symbols $\mathbf{m}$ from the received signal.

Then the question is: How should we choose $\mathbf{z}$ in order to transmit a message and recover it stably? This time we want to utilize the fact that we have multiple-antennas in our encoding scheme.

The model in (10.1) expresses the fact that the signal received at the k-th antenna is a linear combination of the transmitted signal f sent through N_P different paths, where each path, considered as a communication channel, is represented by the spreading function h_n. Each path in (10.1) is modeled by an LTV-system in the sense of (8.1). At this point we want to recall the reason why (8.1) is employed to describe communication channels in the first place. The time-variance of a communication channel stems from the fact that the received signal is a superposition of multiple propagation paths where each scatterer contributes an attenuated and time-frequency shifted version of the input to the received signal [3, 8, 9]. From a physical point of view, this motivates the assumption that each h_n is contained in a small compact area on the time-frequency domain and that each path has a different direction of arrival, provided that the angular resolution of the receiver is high enough. Formally, we make the following assumption on the channel model (10.1):

$$\begin{aligned} &(\mathbf{i})\ \theta_n \neq \theta_m \text{ if } m \neq n. \\ &(\mathbf{ii})\ \text{For each } n, \text{ there exist } k, \ell \in \{0, \dots, L-1\} \text{ with} \\ &\qquad \operatorname{supp}(h_n) \subset \Big[kT, (k+1)T\Big) \times \Big[\ell\Omega, (\ell+1)\Omega\Big). \end{aligned} \tag{10.3}$$

Assumption (**i**) indicates that no two propagation paths have the same direction of arrival, while Assumption (**ii**) states that the spreading function h_n, associated with the n-th path, is supported in only one of the L^2 rectangles $[kT, (k+1)T) \times [\ell\Omega, (\ell+1)\Omega)$, $k, \ell = 0, \dots, L-1$. Figure 10.1 illustrates an example with $N_P = 8$, where (**i**) each of the N_P paths is associated with a distinct DoA angle θ_n and (**ii**) the support of h_n, $n = 1, \dots, N_P$ are displayed in different colors and each of them is contained in one of the small rectangles.

Note that Condition (**ii**) allows the support of the spreading functions h_m and h_n, associated to different paths, to lie in the same rectangle (k, l) of the time-frequency plane. This is illustrated in Figure 10.1. There the supports

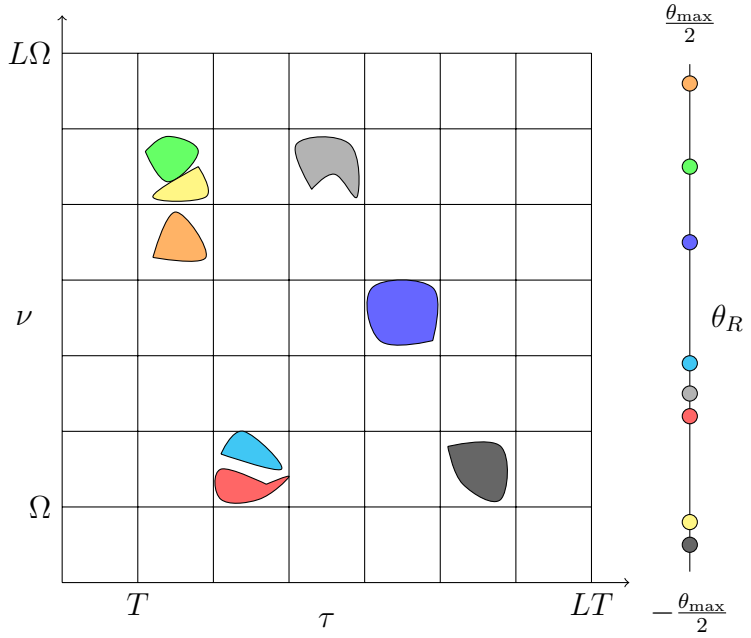

Figure 10.1.: Here $L = 7$ is chosen. The contribution of each path to the support of the spreading function is illustrated in a different color. On the right, the DoAs of each propagation path is shown with the corresponding colors.

corresponding to the paths colored by green and yellow share the same rectangle hence the same pair of indices (k, l). The same also holds true for the paths colored by cyan and red. We will see in a moment that Assumption (**i**) plays a key role since it guarantees the separability of different paths.

Figure 10.2 illustrates the full transmission and recovery cycle for one message $\mathbf{m} \in \mathbb{C}^Q$. The underlying idea is to split the received signal in the angular domain and then the signals from different directions are analyzed separately. Basically, we treat each propagation path as a channel of its own and choose the best path.

10.1.1. Encoder

Given $\mathbf{m} \in \mathbb{C}^Q$, we choose a prime number $L > Q$, an integer K so that $0 \leq K \leq L - Q$, and a unimodular vector $\mathbf{c} \in \mathbb{C}^L$ satisfying condition of

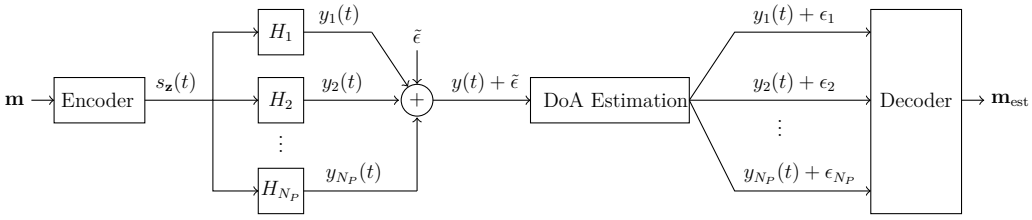

Figure 10.2.: Full transmission and recovery cycle for one message $\mathbf{m}$ and its estimation $\mathbf{m}_{\text{est}}$. The encoder generates from the message $\mathbf{m} \in \mathbb{C}^Q$ the vector $\mathbf{z} \in \mathbb{C}^L$, which is then transmitted through the multipath LTV channel. The received signal $y(t) + \tilde{\epsilon}$ is the superposition of the signals from all propagation paths, where $\tilde{\epsilon}$ is an additive noise term and H_n describes the channel corresponding to the nth path.

Theorem 4.2.13. Therewith, we define the vector $\mathbf{z} \in \mathbb{C}^L$ by

$$\mathbf{z} = \sum_{\ell=0}^{K-1} \mathbf{M}^\ell \mathbf{c} + \sum_{\ell=K}^{Q+K-1} m_{\ell-K}\, \mathbf{M}^\ell \mathbf{c}\,, \tag{10.4}$$

which encodes the message $\mathbf{m}$ by linearly combining the coefficient m_ℓ with the vector $\mathbf{M}^\ell \mathbf{c}$. As discussed in Section 4.2, the finite Gabor frame with a unimodular window $\mathbf{c} \in \mathbb{C}^L$ can be partitioned into L orthogonal bases, namely, $\{\mathbf{M}^\nu \mathbf{T}^\tau \mathbf{c}\}_{\nu=0}^{L-1}$ for $\tau = 0, \ldots, L-1$. For convenience, we define the orthogonal matrix

$$\mathbf{\Psi}_\tau = \operatorname{diag}(\mathbf{T}^\tau \mathbf{c})\mathbf{F} = \left[\mathbf{T}^\tau \mathbf{c}, \mathbf{M}^1 \mathbf{T}^\tau \mathbf{c}, \cdots, \mathbf{M}^{L-1} \mathbf{T}^\tau \mathbf{c}\right]$$

for $\tau = 0, \ldots, L-1$ and the vector

$$\mathbf{v}_K := (\underbrace{1, \ldots, 1}_{K-\text{times}}, m_0, \cdots, m_{Q-1}, 0, \cdots, 0)^T \in \mathbb{C}^L, \tag{10.5}$$

where the first K entries are ones, the next Q entries are the message symbols and the remaining entries are zeros. Therewith, (10.4) can be written as $\mathbf{z} = \mathbf{\Psi}_0 \mathbf{v}_K$ and by the orthogonality of $\mathbf{\Psi}_0$, one has

$$\mathbf{\Psi}_0^* \mathbf{z} \;= L \mathbf{v}_K \tag{10.6}$$

and in particular $|\operatorname{supp}(\mathbf{\Psi}_0^* \mathbf{z})| = |\operatorname{supp}(\mathbf{v}_K)| = Q + K$. So the input signal fed into the channel has the general form in (10.2) but with $\mathbf{z} \in \mathbb{C}^L$ as given in (10.4).

10.1.2. Decoder

The received signal at each antenna is given in (10.1) and the vector $\mathbf{y}(t) \in \mathbb{C}^{N_R}$ containing the received signals at all antennas can be written as

$$\mathbf{y}(t) = (\mathrm{H}f)(t) = \sum_{n=1}^{N_P} \mathbf{a}(\theta_n)\, r_n(t)\,, \qquad t \in \mathbb{R}\,, \tag{10.7}$$

with r_n given by

$$r_n(t) = \iint_{\mathbb{R}^2} h_n(\tau,\nu)\, \mathrm{e}^{j2\pi\nu(t-\tau)} f(t-\tau) d\tau d\nu. \tag{10.8}$$

Equation (10.7) has the usual form of a direction-of-arrival (DoA) estimation problem. Hence, we can apply common methods from DoA estimation to determine the directions θ_n from the received signal $\mathbf{y}$ [133–136]. After a successful DoA estimation applied to (10.7), we obtain for each direction of arrival, θ_n, a received signal (10.8). The first condition in our assumption (10.3) guarantees that each received signal $r_n(t)$ corresponds to a different propagation path. So up to this point, we have separated the received signal into its scattering components each corresponding to a different path. The main idea now is to analyze each path separately. We apply Lemma 8.1.1 to he received signal r_n along each path to get a finite dimensional problem:

$$\mathbf{r}_n = \mathcal{Z}_L^{LT} r_n = \mathbf{G}_{\mathbf{z}} \mathbf{h}_n = \Big(\sum_{k,\ell=0}^{L-1} h_{n,(k,\ell)} \mathbf{M}^\ell \mathbf{T}^k \Big) \mathbf{z}\,. \tag{10.9}$$

To simplify notation, we omitted again the variables (τ,ν) and simply write h_n and we do not differentiate between the spreading function of a single path h_n, and the corresponding $(LT, L\Omega)$-quasi-periodic spreading function $h_n^{(LT,L\Omega)}$. Assumption (**ii**) in (10.3) states that the time-frequency spread of a propagation path is small in the sense that the support of each spreading function h_n is contained in a box length T and width Ω in the time-frequency plane (cf. Figure 10.1). Hence, for each h_n there exists a unique pair (k_n, ℓ_n) such that

$$\mathrm{supp}(h_n) \subset (k_n T, (k_n+1)T] \times (\ell_n \Omega, (\ell_n+1)\Omega] \tag{10.10}$$

and so, (10.9) simplifies to

$$\mathbf{r}_n = \mathcal{Z}_L^{LT} r_n = h_{n,(k_n,\ell_n)} \mathbf{M}^{\ell_n} \mathbf{T}^{k_n} \mathbf{z}\,. \tag{10.11}$$

Next, our aim is to estimate for each $n = 1, \ldots, N_P$ the complex path gain

$h_{n,(k_n,\ell_n)} \in \mathbb{C}$, the pair (k_n, ℓ_n), which corresponds to the time-frequency support of the nth propagation path, and the message $\mathbf{m}$ from the vector $\mathbf{r}_n = \mathcal{Z}_L^{LT} r_n$.

In order to build an intuition why this problem is solvable in the first place, we begin our analysis by considering exact measurements without noise. The case where the signals are corrupted by noise is analyzed in detail in Section 10.2. The following Theorem shows that if the message $\mathbf{m}$ encoded in the signal (10.4) is sufficiently short then it can be recovered at the receiver.

Theorem 10.1.1. *Let* H *be a LTV channel given as in* (10.7) *and* (10.8)*, and let* $s_{\mathbf{m}}(t)$ *be a transmit signal of the form* (10.2) *with* $\mathbf{z}$ *given in* (10.4)*, with* $K \geq 1$*, and with a message* $\mathbf{m} \in \mathbb{C}^Q$*. If* $Q \leq \frac{L-1}{2} - K$ *then* $\mathbf{m}$ *can be recovered from the received signal* $\mathbf{y} = \mathrm{H} s_{\mathbf{m}}$*.*

Remark 10.1.1. *To maximize the message length* Q *one has to choose* $K = 1$*.*

Theorem 10.1.1 assumes that the employed antenna array at the receiver can partition the received signal into its multi-path components. Hence, instead of considering one channel with multiple-paths we consider N_P channels with only one path. Therefore, the condition on the message length is independent of the support of the delay-Doppler shifts.

The following proof of Theorem 10.1.1 is constructive in the sense that it shows how the message $\mathbf{m}$ can be reconstructed from the received signal.

Proof. Again for ease of notation we write h_n instead of $h_{n,(k_n,\ell_n)}$. As described above, we first apply standard DOA estimation algorithm to the received signal $\mathbf{y}$ to obtain the signal r_n, $n = 1, \ldots, N_p$ from the different directions. Out of this N_P signals, we choose one (e.g. the signal with the largest energy) and apply the vectorized Zak transform to obtain the vector $\mathbf{r}_n$. Based on $\mathbf{r}_n$ we determine now the pair (k_n, ℓ_n), describing the location of the channel support, the channel coefficient h_n, $\mathbf{v}_K$, and finally the message $\mathbf{m}$.

First we show that

$$k_n = \operatorname*{arg\,min}_{k=0,1,\ldots,L-1} |\mathrm{supp}(\mathbf{\Psi}_k^* \mathbf{r}_n)| \ . \tag{10.12}$$

Indeed, multiplying the received signal (10.11) from path n with the orthogonal

matrix $\mathbf{\Psi}_{k_n}^*$ yields

$$\begin{aligned}\left(\mathbf{\Psi}_{k_n}^*\mathbf{r}_n\right)[x] &= \left\langle \mathbf{M}^x\mathbf{T}^{k_n}\mathbf{c}, \mathbf{r}_n\right\rangle \\ &= h_n\left\langle \mathbf{M}^x\mathbf{T}^{k_n}\mathbf{c}, \mathbf{M}^{\ell_n}\mathbf{T}^{k_n}\mathbf{z}\right\rangle \\ &= h_n \mathrm{e}^{j\frac{2\pi}{L}k_n(\ell_n-x)}\left\langle \mathbf{M}^{x-\ell_n}\mathbf{c}, \mathbf{z}\right\rangle \\ &= h_n \mathrm{e}^{j\frac{2\pi}{L}k_n(\ell_n-x)}\left(\mathbf{\Psi}_0^*\mathbf{z}\right)[x-\ell_n] \\ &= Lh_n \mathrm{e}^{j\frac{2\pi}{L}k_n(\ell_n-x)}\mathbf{T}^{\ell_n}\mathbf{v}_K[x]\ . \end{aligned} \tag{10.13}$$

for $x = 0, \ldots, L-1$, and where we used (10.6) to obtain the last line. So together with (10.6), it follows that

$$\left|\mathrm{supp}(\mathbf{\Psi}_{k_n}^*\mathbf{r}_n)\right| = |\mathrm{supp}(\mathbf{v}_K)| = Q+K = |\mathrm{supp}(\mathbf{\Psi}_0^*\mathbf{z})|\ ,$$

i.e., if $\mathbf{z} \in \mathbb{C}^L$ has a sparse representation with respect to the basis $\mathbf{\Psi}_0$ then $\mathbf{r}_n$ has a sparse representation with respect to the basis $\mathbf{\Psi}_{k_n}$. In particular, $\mathbf{\Psi}_{k_n}^*\mathbf{r}_n \in \mathbb{C}^L$ has $L-(Q+K)$ non-zero entries. So if we choose $Q+K \leq (L-1)/2$ then then $\mathbf{\Psi}_{k_n}^*\mathbf{r}_n$ has at least $L - \frac{L-1}{2} = \frac{L+1}{2}$ zero entries.

Now we note that the L^2 entries of the collection $\mathbf{\Psi}_0^*\mathbf{r}_n, \ldots, \mathbf{\Psi}_{L-1}^*\mathbf{r}_n$ coincides (up to ordering) with the entries of the short-time Fourier transform $\mathrm{V}_{\mathbf{c}}\mathbf{r}_n$ (cf. Section 4.2.4) and that Theorem 4.2.13 implies that $\mathrm{V}_{\mathbf{c}}\mathbf{r}_n$ has at most $L-1$ zero entries. So since $\mathbf{\Psi}_{k_n}^*\mathbf{r}_n$ has at least $(L+1)/2$ zero entries, it follows that any other vector $\mathbf{\Psi}_k^*\mathbf{r}_n$, with $k \neq k_n$, can have at most $(L-1) - \frac{L+1}{2} = \frac{L-3}{2}$ zero entries. This observation allows us to uniquely determine $k_n \in \{0, \ldots, L-1\}$ by counting the number of zero entries in $\mathbf{\Psi}_0^*\mathbf{r}_n, \ldots, \mathbf{\Psi}_{L-1}^*\mathbf{r}_n$. In other words, we showed (10.12) provided that $Q+K \leq \frac{L-1}{2}$. Next, we observe that according to (10.13), $\mathbf{\Psi}_{k_n}^*\mathbf{r}_n$ is (up to a factor) equal to $\mathbf{T}^{\ell_n}\mathbf{v}_K$. Since $\mathbf{v}_K$ is given by (10.5), it follows that $\ell_n \in \{0, \ldots, L-1\}$ can be determined by finding the first nonzero entry appearing in $\mathbf{\Psi}_{k_n}^*\mathbf{r}_n$. Since $K \geq 1$, the first entry of $\mathbf{v}_K$ is equal to 1, and so by (10.13), the channel coefficient h_n is obtained by

$$h_n = \tfrac{1}{L}\left(\mathbf{\Psi}_{k_n}^*\mathbf{r}_n\right)[\ell_n]\ .$$

Finally, knowing k_n, ℓ_n and h_n, $\mathbf{v}_K$ and the message vector $\mathbf{m} = (m_0, ..., m_{Q-1})^T$ can be determined from (10.13). □

Algorithm 2 Multi-Antenna Decoder

Input: $\mathbf{r}_n$, $K, Q \in \mathbb{N}$

1) $(k_n, \ell_n) = \arg\min_{(k,\ell)} \xi_n(k,\ell)$

2) $\widetilde{\mathbf{r}}_n = \mathbf{\Psi}_0^* \mathbf{T}^{-k_n} \mathbf{M}^{-\ell_n} \mathbf{r}_n$

3) $\hat{h}_n = \frac{1}{K} \sum_{s=0}^{K-1} \mathrm{e}^{-2\pi j \frac{k_n}{L} s} \widetilde{\mathbf{r}}_n[s]$

4) $\hat{m}_i = \frac{\widetilde{\mathbf{r}}_n[i+K]}{\hat{h}_n}$, for $i = 0, \ldots, Q-1$

Output: $\hat{\mathbf{m}}_n$

10.2. Multi-Antenna receiver for noisy measurements

The previous section discussed multi-antenna receivers for LTV channels and showed how a joint channel estimation and data transmission can be implemented, in principle. Nevertheless, the discussion did not consider that the received signal might be corrupted by additional noise. But with additional noise, the sparsity assumptions made in the proof of Theorem 10.1.1 no longer hold. So the described channel estimation and signal recovery procedure might no longer work. Especially (10.12), will fail now.

Therefore, this section discusses a strategy how the receiver can recover messages from noisy measurements.

We assume that the received signal $\mathbf{y}$ is corrupted by noise. So estimating $\mathbf{r}_n$ from (10.7) yields a corrupted version of $\mathbf{r}_n$. Instead of (10.11) the received signal from the n-th path is now given by

$$\mathbf{r}_n := h_n \mathbf{M}^{\ell_n} \mathbf{T}^{k_n} \mathbf{z} + \boldsymbol{\epsilon}\,, \qquad n = 1, \ldots, N_p \tag{10.14}$$

where $\boldsymbol{\epsilon}$ characterizes the error due to measurement noise. For simplicity of the following derivation, we assume that $\boldsymbol{\epsilon}$ is i.i.d zero mean Gaussian and variance σ^2. Now we want to use a similar method as in the noiseless case to recover the message from $\mathbf{r}_n$. However, because of the measurement noise, there might not exist a basis $\mathbf{\Psi}_{k_n}$ under which the representation of $\mathbf{r}_n$ is sparse. Therefore, we apply a slightly different strategy. To this end we define for $n = 1, ..., N_P$ the function ξ_n:

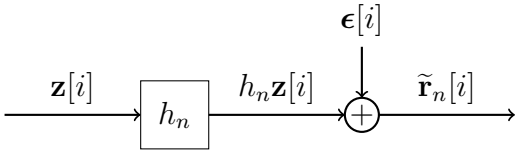

Figure 10.3.: Standard transmission scenario over a Gaussian channel.

$$\begin{aligned}\xi_n(k,\ell) &= \|\boldsymbol{\Psi}_k\mathbf{P}_{\mathbb{1}+\ell}\boldsymbol{\Psi}_k^*\mathbf{r}_n - \mathbf{r}_n\|_2 \\ &= \sqrt{L}\,\|\mathbf{P}_{\mathbb{1}+\ell}\boldsymbol{\Psi}_k^*\mathbf{r}_n - \boldsymbol{\Psi}_k^*\mathbf{r}_n\|_2\ ,\end{aligned}$$

with the projection $\mathbf{P}_\Omega$ defined for $\Omega \subset \{0,\dots,L-1\}$ by

$$\mathbf{P}_\Omega\mathbf{g}[k] = \begin{cases}\mathbf{g}[k] & \text{if } k \in \Omega \\ 0 & \text{if } k \notin \Omega\end{cases}$$

and with $\mathbb{1}+\ell = \{\ell, \ell+1, \dots, \ell+Q+K-1\}$. In the noiseless case, we would have that $\xi_n(k_n,\ell_n) = 0$. In the noisy case, we determine the pair (k_n,ℓ_n), which characterizes the support of of the channel spreading function h_n (10.10) by

$$(k_n,\ell_n) = \arg\min_{(k,\ell)} \xi_n(k,\ell)\ .$$

An overview of the full recovery algorithm in the noisy case is given in Algorithm 2. The algorithm needs as input the parameters K, Q and the measured signal (10.14) from the n-th path. Its output is the estimated message vector $\hat{\mathbf{m}}_n$ from the n-th path. The first two steps of the algorithm recover the time-frequency effects of the channel, the third step computes the maximum-likelihood estimate for the channel attenuation factor and finally the fourth step estimates the messages.

In principle, due to the additional noise, Step 1 of the algorithm may not find the correct support and the following subsection will analyze in some detail the probability for a wrong detection of the support. For the moment, we assume that the first steps make no error, i.e. we assume that the minimizer of $\xi_n(k,\ell)$ is indeed the correct support, i.e. $\arg\min_{k,\ell}\xi_n(k,\ell) = (k_n,\ell_n)$. Recall that the received signal from the nth path is given by (10.14) with $\boldsymbol{\epsilon} \sim \mathcal{N}_{\mathbb{C}}(0,\sigma^2\mathbf{I})$. Since we know the correct support we can simply compute

$$\widetilde{\mathbf{r}}_n = \boldsymbol{\Psi}_0\mathbf{T}^{-k_n}\mathbf{M}^{-\ell_n}\mathbf{r}_n = h_n\mathbf{v}_K + \boldsymbol{\Psi}_0\mathbf{T}^{-k_n}\mathbf{M}^{-\ell_n}\boldsymbol{\epsilon}$$

$$= h_n \mathbf{v}_K + \widetilde{\boldsymbol{\epsilon}}\,,$$

with $\widetilde{\boldsymbol{\epsilon}} \sim \mathcal{N}_{\mathbb{C}}\left(0, \sigma^2 \mathbf{I}_L\right)$, since $\mathbf{T}^{-k_n}\mathbf{M}^{-\ell_n}$ is a unitary matrix. The first K elements in $\widetilde{\mathbf{r}}_n$ contain h_n disturbed by noise and the elements indexed by K up to $K+Q-1$ contain the messages multiplied by h_n and disturbed by additive noise. This is a standard scenario in communication theory [137] and its block diagram is illustrated in Figure 10.3. In this scenario a two step transmission scheme is applied. First, the channel is estimated for instance with a *maximum-likelihood* (ML) estimator, then in the second step data is transmitted over the channel and is recovered at the receiver using the previously estimated information. Hence, we have reduced the problem to a LTI system identification and message estimation setup. The ML estimate $\hat{h}_n$ of h_n is given by

$$\hat{h}_n = \frac{1}{K} \sum_{s=0}^{K-1} \mathrm{e}^{-2\pi j \frac{k_n}{L} s}\, \widetilde{\mathbf{r}}_n[s]\,.$$

Additionally, we have $\mathrm{Var}\left[h_n - \hat{h}_n\right] = \sigma^2/K$. Let $i \in \{K, \ldots, K+Q-1\}$ and denote $\kappa = h_n - \hat{h}_n \sim \mathcal{N}_{\mathbb{C}}\left(0, \frac{\sigma^2}{K}\right)$, now following [138] we describe the received signal as

$$\widetilde{\mathbf{r}}_n[i] = \hat{h}_n m_{i-K} + \kappa m_{i-K} + \boldsymbol{\epsilon}[i]\,.$$

So we estimate the message m_{i-K} by

$$\hat{m}_{i-K} = \frac{\widetilde{\mathbf{r}}_n[i]}{\hat{h}_n}\,. \tag{10.15}$$

Then the estimate $\hat{m}_{i-K}$ in (10.15) is a Gaussian random variable with $\hat{m}_{i-K} \sim \mathcal{N}_{\mathbb{C}}\left(m_{i-K}, (1+|m_{i-K}|^2)\,\frac{\sigma^2}{|\hat{h}_n|^2}\right)$.

10.3. Error probability of Step 1

In general, Step 1 of Algorithm 2 might not give the correct support set. Then our recovery algorithm will fail. In this subsection, we want to analyze the probability of such a misdetection of the support set. To this end, let $(k_*, \ell_*) = \arg\min_{k,\ell} \xi_n(k,\ell)$ be the estimated indices of the channel support. In the following we want to show that $(k_*, \ell_*) = (k_n, \ell_n)$ holds with high probability even in the noisy case. To this end, we note that there are two types of errors:

1. Detection of the wrong basis, i.e. $k_* \neq k_n$

2. Detection of the wrong support in the correct basis, i.e. $\ell_* \neq \ell_n | k_* = k_n$.

For ease of presentation, we assume $K = 0$ and $h_n = 1$. The generalization to $h_n \neq 1$ is straight forward.

10.3.1. Detection of the wrong basis

We begin our analysis with the first error type. Notice that if an error occurs then, for some ℓ, we must have

$$\|\mathbf{P}_{\mathbb{1}+\ell}\mathbf{\Psi}_{k_*}\mathbf{r}_n - \mathbf{\Psi}_{k_*}\mathbf{r}_n\|_2^2 < \|\mathbf{P}_{\mathbb{1}+\ell_n}\mathbf{\Psi}_{k_n}\mathbf{r}_n - \mathbf{\Psi}_{k_n}\mathbf{r}_n\|_2^2 . \tag{10.16}$$

Setting $\Omega_* = (\mathbb{1}+\ell)^C$, $\Omega_n = (\mathbb{1}+\ell_n)^C$ and $\widetilde{\mathbf{r}}_{\Omega_*} = \mathbf{P}_{\Omega_*}\mathbf{\Psi}_{k_*}(\mathbf{r}_n - \boldsymbol{\epsilon})$ we have

$$\|\mathbf{P}_{\mathbb{1}+\ell}\mathbf{\Psi}_{k_*}\mathbf{r}_n - \mathbf{\Psi}_{k_*}\mathbf{r}_n\|_2^2 = \|\widetilde{\mathbf{r}}_{\Omega_*} + \boldsymbol{\epsilon}_{\Omega_*}\|_2^2 , \tag{10.17}$$

and

$$\|\mathbf{P}_{\mathbb{1}+\ell_n}\mathbf{\Psi}_{k_n}\mathbf{r}_n - \mathbf{\Psi}_{k_n}\mathbf{r}_n\|_2^2 = \|\boldsymbol{\epsilon}_{\Omega_n}\|_2^2 . \tag{10.18}$$

So $\gamma := \|\widetilde{\mathbf{r}}_{\Omega_*}\|_2^2$ measure the signal energy contained in the complement of the detected set $\Omega_* = (\mathbb{1}+\ell)^C$. For $k_* = k_n$, i.e., if we would have detected the correct basis, γ would be zero. Therefore, γ might be considered as the signal energy lost by a wrong detection of k_n and for some ℓ. Now, let $\boldsymbol{\epsilon}_{\Omega_*}[i] = \widetilde{\alpha}_i + j\widetilde{\beta}_i$ and $\boldsymbol{\epsilon}_{\Omega_n}[i] = \alpha_i + j\beta_i$ where $\alpha_i, \beta_i, \widetilde{\alpha}_i, \widetilde{\beta}_i$ are i.i.d. Gaussian random variables with $\alpha_i, \beta_i, \widetilde{\alpha}_i, \widetilde{\beta}_i \sim \mathcal{N}(0, \frac{\sigma^2}{2})$. Next, reformulating (10.16) using (10.17) and (10.18) delivers

$$\begin{aligned} 0 &\geq \|\widetilde{\mathbf{r}}_{\Omega_*} + \boldsymbol{\epsilon}_{\Omega_*}\|_2^2 - \|\boldsymbol{\epsilon}_{\Omega_n}\|_2^2 \\ &= \|\widetilde{\mathbf{r}}_{\Omega_*}\|_2^2 + \sum_{i\in\Omega_*} 2\Re\{\widetilde{\mathbf{r}}_{\Omega_*}[i]\}\widetilde{\alpha}_i + 2\Im\{\widetilde{\mathbf{r}}_{\Omega_*}[i]\}\widetilde{\beta}_i \\ &\quad + \sum_{i\in\Omega_*} \widetilde{\alpha}_i^2 + \widetilde{\beta}_i^2 - \sum_{i\in\Omega_n} \alpha_i^2 + \beta_i^2 = G^1 + E_1^1 - E_2^1 , \end{aligned}$$

where G^1, E_1^1 and E_2^1 are random variables with $G^1 \sim \mathcal{N}(\gamma, 2\gamma\sigma^2)$ and noticing $|\Omega_*| = |\Omega_n| = L - Q$, we have $E_1^1, E_2^1 \sim Erl(L-Q, \sigma^2)$. For $a \in \mathbb{N}$ and $b > 0$, $Erl(a,b)$ denotes the Erlang distribution [139] with probability density function given by

$$f_{Erl}(x; a, b) = \frac{1}{b^a\,(a-1)!}\mathrm{e}^{-\frac{x}{b}}x^{a-1} . \tag{10.19}$$

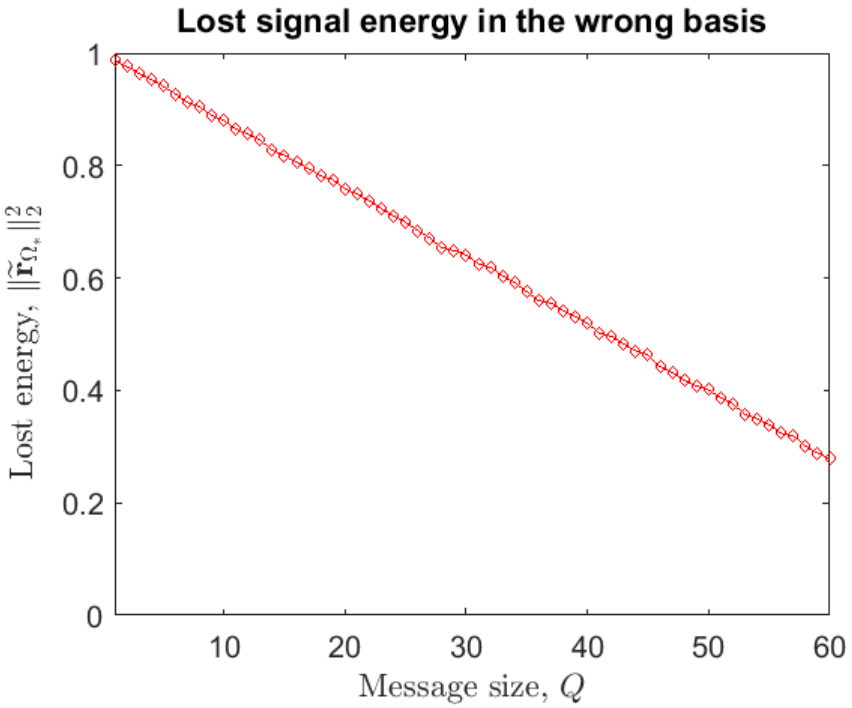

Figure 10.4.: Empirical validation of signal energy loss in the wrong basis relative to sparsity Q. Here $K = 0$ and $L = 83$ is chosen while Q varies.

Accordingly, the error probability can be given by

$$\begin{aligned}\mathbb{P}\{k_* \neq k_n\} &= \mathbb{P}\{G^1 + E_1^1 - E_2^1 \leq 0\} \\ &= \int_{-\infty}^{0} f_{G^1}(u) * f_{E_1^1}(u) * f_{-E_2^1}(u)\mathrm{d}u\,, \end{aligned} \tag{10.20}$$

where $f_{G^1}, f_{E_1^1}$ and $f_{-E_2^1}$ denotes the probability density functions of G^1, E_1^1 and $-E_2^1$ and $*$ denotes the convolution.
As seen in (10.20), the error probability depends on the probability density function of G^1 and therefore on the value $\gamma = \|\widetilde{\mathbf{r}}_{\Omega_*}\|_2^2$. Moreover, with increasing γ the error probability decreases and because of Theorem 4.2.13, $\|\widetilde{\mathbf{r}}_{\Omega_*}\|_2^2$ cannot be zero. In fact, when assuming a unit normed signal, estimating the value of $\|\widetilde{\mathbf{r}}_{\Omega_*}\|_2^2$ corresponds to an estimate of the biggest singular value of $\mathbf{P}_{\Omega_*}\mathbf{\Psi}_{k_*}$. Indeed for any $\mathbf{g} \in \mathbb{C}^L$ we have $\|\mathbf{P}_{\Omega_*}\mathbf{\Psi}_{k_*}\mathbf{g}\|_2 = \|\mathbf{F}_{\Omega_*}\|_{2\to 2}\|\mathbf{g}\|_2$, where $\mathbf{F}_{\Omega_*}$ denotes the partial Fourier matrix, which only contains the rows indexed by Ω_* and $\|\cdot\|_{2\to 2}$ denotes the maximum singular value. Hence, it is enough

to estimate $\|\mathbf{F}_{\Omega_*}\|_{2\to 2}$. In fact, for random sets Ω_* it is well known that the singular values of partial Fourier matrices are well controlled (see, e.g., [140, Theorem 3.1].
Indeed this theoretical observation is empirically justified in Figure 10.4. There $L = 83$ and $K = 0$ are chosen and $1 \leq Q \leq 60$. For each experiment a unit normed pilot signal $\mathbf{c} \in \mathbb{C}^L$ is fixed and $\mathbf{m} \in \mathbb{C}^Q$ is chosen uniformly random over the Q dimensional complex unit ball. Then the channel input vector $\mathbf{z} \in \mathbb{C}^L$ is constructed as described in (10.4). Since the choice of $\mathbf{M}^k\mathbf{T}^\ell$ for $k, \ell \in \{0, \dots L-1\}$ does not change the absolute value, we set $\mathbf{r}_n - \boldsymbol{\epsilon} = \mathbf{z}$ for the experiments. Next, a pair (k, ℓ) with $k, \ell \in \{0, \dots L-1\}$ was chosen uniformly random and $\|\widetilde{\mathbf{r}}_{\Omega_*}\|_2^2 = \|\mathbf{P}_{\Omega_*}\boldsymbol{\Psi}_k^*\mathbf{z}\|_2^2$ was computed. Each data point is obtained by taking the mean of 400 experiments. We observe that the lost signal power in the wrong basis linearly decreases with increasing Q. In the wrong basis the signal energy is "smeared" equally over the basis elements. Hence, for correct support detection a small message size is beneficial.

10.3.2. Detection of the wrong support in the correct basis

A similar expression to (10.20) is obtained when analyzing the second error probability, i.e. $\mathbb{P}\{\ell_* \neq \ell_n | k_* = k_n\}$. Assume that the intersection of the wrongly detected support with the original support has size S, i.e. $|\mathbb{1} + \ell_* \cap \mathbb{1} + \ell_n| = S$. Further, notice

$$\|\mathbf{P}_{\mathbb{1}+\ell_*}\boldsymbol{\Psi}_{k_n}\mathbf{r}_n - \boldsymbol{\Psi}_{k_n}\mathbf{r}_n\|_2^2 = \sum_{i\in\mathcal{A}} |\widetilde{\mathbf{r}}_{\Omega_*}[i] + \epsilon[i]|^2 + \sum_{i\in\mathcal{B}} |\epsilon[i]|_2^2 \,, \tag{10.21}$$

where $\Omega_* = (\mathbb{1} + \ell_*)^C$, $\mathcal{A} = (\mathbb{1} + \ell_n)\backslash\{\mathbb{1}+\ell_*\cap\mathbb{1}+\ell_n\}$ and $\mathcal{B} = (\mathbb{1} + \ell_* \cap \mathbb{1} + \ell_n)^C$, hence we have $|\mathcal{A}| = Q - S$ and $|\mathcal{B}| = L - 2Q + S$. Additionally, we have $|\Omega_n \setminus \mathcal{B}| = Q - S$ since $\mathcal{B} \subset \Omega_n$. Following the previous argument, using (10.18) and (10.21) we get

$$\begin{aligned} 0 &> \|\mathbf{P}_{\mathbb{1}+\ell_*}\boldsymbol{\Psi}_{k_n}\mathbf{r}_n - \boldsymbol{\Psi}_{k_n}\mathbf{r}_n\|_2^2 - \|\mathbf{P}_{\mathbb{1}+\ell_n}\boldsymbol{\Psi}_{k_n}\mathbf{r}_n - \boldsymbol{\Psi}_{k_n}\mathbf{r}_n\|_2^2 \\ &= \sum_{i\in\mathcal{A}} |\widetilde{\mathbf{r}}_{\Omega_*}[i] + \epsilon[i]|^2 + \sum_{i\in\mathcal{B}} |\epsilon[i]|^2 - \sum_{i\in\Omega_n} |\epsilon[i]|^2 \\ &= G^2 + E_1^2 - E_2^2 \,, \end{aligned}$$

with $G^2 \sim \mathcal{N}\left(\|\widetilde{\mathbf{r}}_{\mathcal{A}}\|_2^2, 2\|\widetilde{\mathbf{r}}_{\mathcal{A}}\|_2^2\sigma^2\right)$ and $E_1^2, E_2^2 \sim Erl(Q-S, \sigma^2)$, where $\|\widetilde{\mathbf{r}}_{\mathcal{A}}\|_2^2 = \sum_{i\in\mathcal{A}} |\widetilde{\mathbf{r}}_{\Omega_*}[i]|_2^2$. Finally, we can give the probability of the second error type by

$$\mathbb{P}\{\ell_* \neq \ell_n | k_* = k_n\} = \mathbb{P}\{G^2 + E_1^2 - E_2^2 \leq 0\}$$

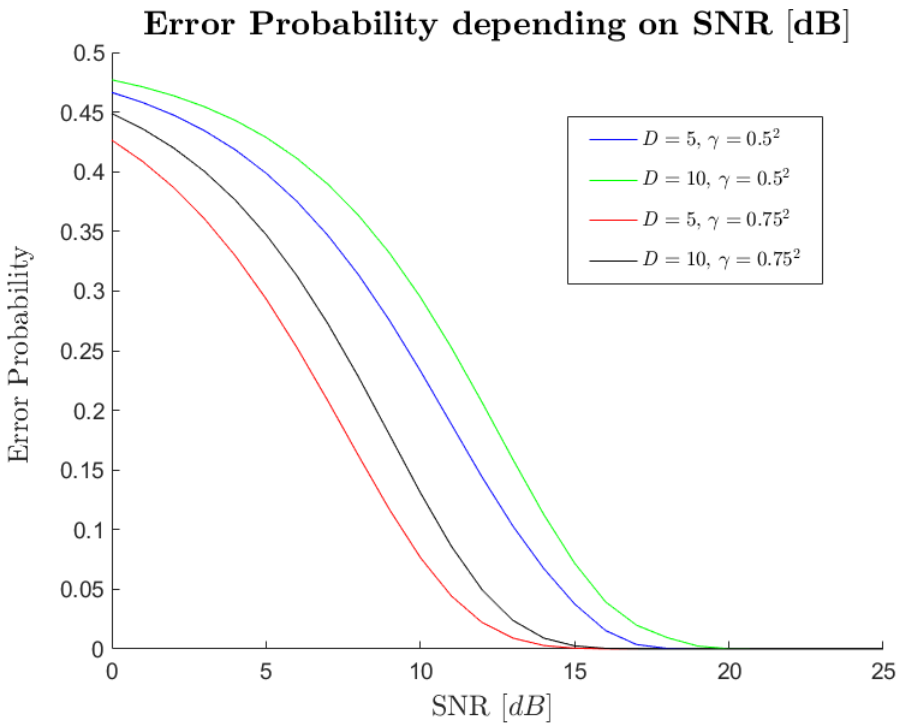

Figure 10.5.: The error probabilities described by $\mathcal{P}(\sigma^2; \gamma, D)$, for different parameters $D = 5, 10$ and $\gamma = 0.5^2, 0.75^2$, depending on the SNR in dB is illustrated where a unit norm signal is assumed. The SNR value in dB of a noise power σ^2 is computed by the formula $SNR[dB] = 10\log_{10}(1/\sigma^2)$.

$$= \int_{-\infty}^{0} f_{G^2}(u) * f_{E_1^2}(u) * f_{-E_2^2}(u)\mathrm{d}u\,. \tag{10.22}$$

It is apparent that (10.20) and (10.22) have the same form and both depend on three parameters, namely:

- D : degrees of freedom, which is given by $L - Q$ in the first error type and by $Q - S$ in the second case
- γ : lost signal power by detection, which is described by $\|\widetilde{\mathbf{r}}_{\Omega_*}\|_2^2$ in the first case and by $\|\widetilde{\mathbf{r}}_{\mathcal{A}}\|_2^2$ in the second error case
- σ^2 : the variance of the white Gaussian noise, i.e. the noise power

We denote the expressions in (10.20) and (10.22) by $\mathcal{P}(\sigma^2; \gamma, D)$. An illustration of $\mathcal{P}(\sigma^2; \gamma, D)$ for a unit normed signal and different SNR values is shown in Figure 10.5. It is apparent that a lower degree of freedom, D, leads to a lower error probability. However, it is obvious that γ has a higher influence on the error probability. This is intuitively expected, since γ describes, in both scenarios, the missing signal power.

Although it is not our main focus, these observations already hints that the *phase shift keying* (PSK) modulation [141] is favourable for this setup. Basically, our aim is to achieve a small D and that small deviations in the support detection leads to high γ. In the first case, $D = L - Q$ depends on the design and how big Q is chosen. The more essential parameter γ is determined by the singular value of the partial Fourier matrix, which are known to be well controlled. Therefore, we can't influence the error probability for the first error type with modulation methods. However, in the second scenario γ increases with increasing $D = Q - S$. In this case we can control the growth of γ depending on the size of D. For instance if the signal power is concentrated on a few entries then although D increases γ won't increase significantly, which then leads to a worse SNR to error probability ratio. If, however, the signal power is uniformly distributed over its entries, then γ will linearly increase with increasing D, which then leads to a better SNR to error probability ratio. This type of modulation can be obtained by using PSK.

10.4. Numerical experiments

In Figure 10.6 the full transmission cycle for the multi-antenna receiver as illustrated in Figure 10.2 is simulated. All numerical experiments were performed for a receiver equipped with a uniform linear array with 40 antennas ($N_R = 40$). The positions of the antennas were chosen uniformly random over all possible sets. The array steering vector was sampled such that the angular resolution for the DoA estimation was roughly around 1°. For the DoA estimation OMP was used to solve the occurring under-determined system of linear equations. In all simulations $L = 59$ was chosen. In Figure 10.6 for each channel, first the set direction of arrivals (arriving angle of the signal) where drawn uniformly random from the set $\left\{\frac{k}{\pi} \,:\, k = 0, \ldots, 179\right\}$. Out of convenience we assumed that the spreading function for each path has the form $h_n(\tau, \nu) = h_n \delta(\tau - \tau_n)\delta(\nu - \nu_n)$, where the pair (τ_n, ν_n) corresponds to the time-frequency shift of the n-th path and h_n is the attenuation factor, which was chosen as a complex Gaussian random variable with mean 0 and variance 1.

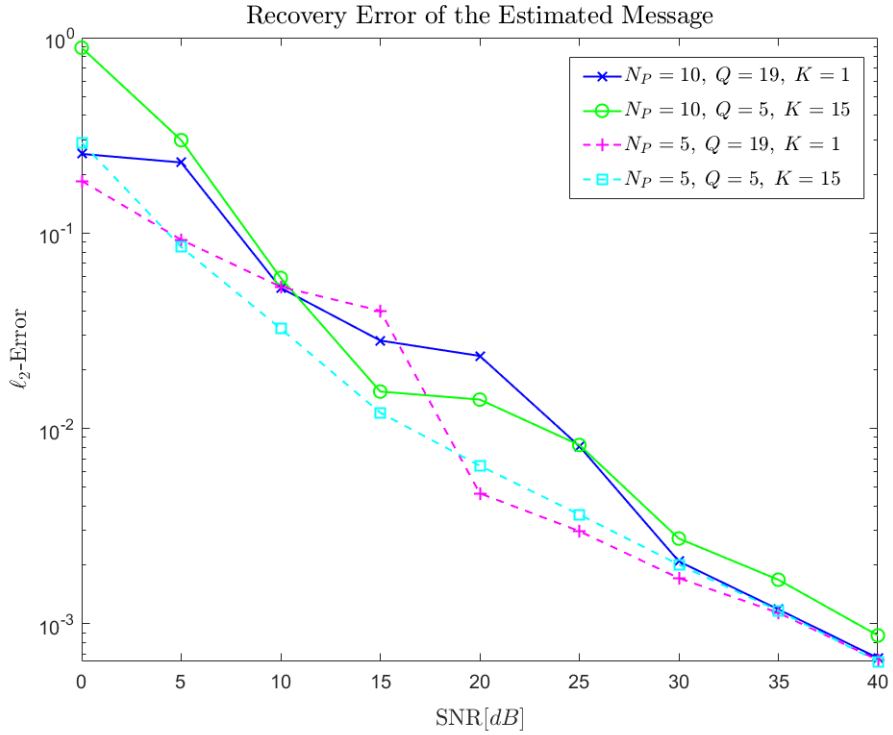

Figure 10.6.: The ℓ_2-error given by $\|\mathbf{m} - \mathbf{m}_{est}\|_2$, relative to $SNR[dB]$ for different number of paths N_P, different number of message length Q and different number of training symbols K.

Hence, for each path the pairs (τ_n, ν_n) where chosen uniformly random over the $L \times L$ discretized time-frequency grid. For each channel realization 5 different messages where generated randomly as complex Gaussian vectors with mean 0 and variance 1. Each data point in Figure 10.6 was averaged over 100 channel realizations, hence over 500 different messages.
In Figure 10.6 the ℓ_2-error of the estimated messages to the original messages (i.e. $\|\mathbf{m} - \mathbf{m}_{est}\|_2$) is illustrated after a full transmission cycle for different number of paths $N_P \in \{5, 10\}$, different sizes of messages $Q \in \{5, 19\}$ and different sizes of training symbols $K \in \{1, 15\}$, recall that $K + Q \leq \frac{L-1}{2}$.
The performance metric used in Figure 10.6 might be unusual since using symbol to error-rate rather than ℓ_2-error is more common. However, in order to use symbol to error-rate metric one needs to a priori fix a modulation method, for example PSK or *quadrature amplitude modulation* (QAM). If we consider each entry in the message $\mathbf{m} \in \mathbb{C}^Q$ to be a symbol from a previously fixed set of symbols, then in Figure 10.6, each message generation assumes a new

random set of symbols on the complex plane. This approach illustrates the performance of the suggested scheme independent of a modulation method and with worst-case assumption since ℓ_2-norm dominates the ℓ_∞-norm on euclidean spaces.
Figure 10.6 clearly illustrates that in cases with high noise the number of training symbols are not able to compensate for the noisy measurements and that the number of paths is the dominating factor for the recovery error. Therefore, having less training symbols and more message symbols is favourable.
For high SNR values (i.e. low noise values) comparing the curves with the same amount of paths we observe that less training symbols outperform those having more training symbols. Which was expected since the training symbols compensate the effect of noise for the cost of messages, therefore in low noise areas those simulations with less training symbols perform better.
Number of training symbols compensate the noise for the simulations in the low SNR values. However, from the simulations it becomes apparent that the number of paths influences the recovery error far more then the number of training symbols.

11. Message Transmission and Estimation over LTV Channels with MUSIC Algorithm

The work in this section is mainly inspired by [12] and is published in [33]. In [12] *multiple signal classification* (MUSIC) algorithm [142] is applied to estimate time delays and Doppler shifts in a received pilot signal. Similar to previous Sections 9 and 10 the transmitted signal contains a pilot and a coded version of the message data. The main difference of the scheme described in this section to the ones in the previous sections is that we do not require the delay-Doppler shifts to lie on a grid on the time-frequency plane.
We consider a narrowband time-varying multipath wireless channel with a multi-antenna receiver equipped with N_A antennas. The communication channel can be modelled as a LTV system given by (8.1), i.e. has the form

$$\mathbf{y}\,(t) = \int_{\mathbb{R}} \int_{\mathbb{R}} \eta\,(\tau,\nu)\, s_T(t-\tau) \mathrm{e}^{j2\pi\nu t} d\tau d\nu\,,\ t \in \mathbb{R}\,, \tag{11.1}$$

where $s_T \in L^2(\mathbb{R})$ is the input signal. The index T is used to emphasize that s_T describes the signal in the time domain. Assuming point scatterers, η for a multipath LTV channel can be described by

$$\eta\,(\tau,\nu) = \sum_{k=1}^{D} \left(\sum_{\ell=1}^{P_k} h_{k,\ell}\, \mathbf{a}\,(\theta_\ell) \right) \delta\,(t-\tau_k)\, \delta\,(\nu-\nu_k)\ ,$$

where $(\tau_k, \nu_k) \in \mathbb{R}^2$ and D denotes mutually distinct delay-Doppler shifts and their number respectively, P_k is the number of all paths sharing the same delay-Doppler shift, $\mathbf{a}(\theta_\ell)$ the array response vector in the θ_ℓ-th direction and $h_{k,\ell}$ the complex attenuation factor of the k-th delay-Doppler shift contributing to the scatterer received from the direction θ_ℓ [3, 12]. Next setting

$$\mathbf{a}_k = \sum_{\ell=1}^{P_k} h_{k,\ell}\, \mathbf{a}\,(\theta_\ell)\ ,\ k = 1, \ldots, D\,,$$

we can describe $\eta(\tau, \nu)$ by

$$\eta(\tau, \nu) = \sum_{k=1}^{D} \mathbf{a}_k \, \delta(t - \tau_k) \, \delta(\nu - \nu_k) \; . \tag{11.2}$$

Note that for spreading functions having the form (11.2) the output y in (11.1) will be contained in $L^2(\mathbb{R})$ since $s_T \in L^2(\mathbb{R})$. Plugging (11.2) into (11.1) we obtain for the received signal the form

$$\mathbf{y}(t) = \sum_{k=1}^{D} \mathbf{a}_k \, s_T(t - \tau_k) \mathrm{e}^{j\nu_k t} \, , \; t \in \mathbb{R}. \tag{11.3}$$

In practical applications the signal has finite support in time and should have a good localization in the frequency domain, i.e. an essential part of its energy should be contained in a finite interval in frequency. Hence, for practical applications it makes sense to consider a frequency domain model of (11.3), which is given by

$$\hat{\mathbf{y}}(\omega) = \sum_{k=1}^{D} \mathbf{a}_k \hat{s}_T(\omega - \nu_k) \mathrm{e}^{-j\omega\tau_k} + \mathbf{n}(\omega) \, , \tag{11.4}$$

where $\hat{\mathbf{y}}$ and $\hat{s}_T$ are the Fourier transforms of $\mathbf{y}$ and s_T, respectively, and $\mathbf{n}$ is the additive noise vector.

11.1. Message coding

Denote by $m_\ell \in \mathbb{R}$ for $\ell = 1, \dots M$ the message symbols which we want to transmit and define the real vector $\mathbf{m} \in \mathbb{R}^M$ as $\mathbf{m} = [m_1, \dots, m_M]^T$ and set $s(\omega) = \hat{s}_T(\omega)$. The main idea here is to build s as a superposition of a pilot signal and carrier signals weighted by the message symbols (entries of the message vector). Denote by $p(\omega)$ a pilot signal and let $\{c_\ell(\omega)\}_{\ell=1}^{M}$ be the set of carrier signals. Then the transmit signal $s(\omega)$ is constructed by

$$s(\omega) = p(\omega) + \sum_{\ell=1}^{M} m_\ell \, c_\ell(\omega) \, . \tag{11.5}$$

Now, using (11.5) we can extend the noiseless case of (11.4) to

$$\hat{\mathbf{y}}(\omega) = \sum_{k=1}^{D} \mathbf{a}_k \, p(\omega - \nu_k) \mathrm{e}^{-j\tau_k\omega} \; + \; \sum_{\ell=1}^{M} m_\ell \sum_{k=1}^{D} \mathbf{a}_k \, c_\ell(\omega - \nu_k) \mathrm{e}^{-j\tau_k\omega} \, .$$

Following [12] we sample $\mathbf{y}(\omega)$ at $\omega = \omega_1, \ldots, \omega_S$ collecting S snapshots from the array. The obtained data can be arranged in matrix form as

$$\begin{aligned}
\mathbf{Y} &= \begin{bmatrix} \hat{\mathbf{y}}^T(\omega_1) \\ \vdots \\ \hat{\mathbf{y}}^T(\omega_S) \end{bmatrix} \\
&= \left[\widetilde{\mathbf{P}}(\boldsymbol{\nu}) \odot \mathbf{V}_{\boldsymbol{\tau}}\right] \mathbf{A} + \sum_{\ell=1}^{M} \left[m_\ell \widetilde{\mathbf{C}}_\ell(\boldsymbol{\nu}) \odot \mathbf{V}_{\boldsymbol{\tau}}\right] \mathbf{A} \\
&= \left[\mathbf{Q}_p(\boldsymbol{\tau}, \boldsymbol{\nu}) \; m_1 \mathbf{Q}_{c_1}(\boldsymbol{\tau}, \boldsymbol{\nu}) \; \ldots \; m_M \mathbf{Q}_{c_M}(\boldsymbol{\tau}, \boldsymbol{\nu})\right] \begin{bmatrix} \mathbf{A} \\ \vdots \\ \mathbf{A} \end{bmatrix} \\
&= \widetilde{\mathbf{Q}}(\boldsymbol{\tau}, \boldsymbol{\nu}, \mathbf{m}) \, \widetilde{\mathbf{A}} \,, \qquad (11.6)
\end{aligned}$$

where $\odot$ denotes the Schur-Hadamard product (pointwise matrix product), and

$$\begin{aligned}
\boldsymbol{\tau} &= \left[\tau_1 \; \ldots \tau_D\right]^T \in \mathbb{R}^D \,, \\
\boldsymbol{\nu} &= \left[\nu_1 \; \ldots \nu_D\right]^T \in \mathbb{R}^D \,, \\
\mathbf{A} &= \left[\mathbf{a}_1^T \ldots \mathbf{a}_D^T\right]^T \in \mathbb{C}^{D \times N_A} \,, \\
\widetilde{\mathbf{P}}(\boldsymbol{\nu}) &= \left[\widetilde{\mathbf{p}}(\nu_1) \; \ldots \; \widetilde{\mathbf{p}}(\nu_D)\right] \in \mathbb{R}^{S \times D} \,, \\
\widetilde{\mathbf{p}}(\nu) &= \left[p(\omega_1 - \nu) \; \ldots \; p(\omega_S - \nu)\right]^T \in \mathbb{R}^S \,, \\
\widetilde{\mathbf{C}}_\ell &= \left[\widetilde{\mathbf{c}}_\ell(\nu_1) \; \ldots \; \widetilde{\mathbf{c}}_\ell(\nu_D)\right] \in \mathbb{R}^{S \times D} \,, \\
\widetilde{\mathbf{c}}_\ell(\nu) &= \left[c_\ell(\omega_1 - \nu) \; \ldots \; c_\ell(\omega_S - \nu)\right]^T \in \mathbb{R}^S \,, \\
\mathbf{V}_{\boldsymbol{\tau}} &= \left[\mathbf{v}(\tau_1) \; \ldots \; \mathbf{v}(\tau_D)\right] \in \mathbb{C}^{S \times D} \,, \\
\mathbf{v}(\tau) &= \left[\mathrm{e}^{-j\tau\omega_1} \; \ldots \; \mathrm{e}^{-j\tau\omega_S}\right]^T \in \mathbb{C}^S \,.
\end{aligned}$$

11.2. Identifiability

The delay-Doppler shift pairs $\{(\tau_k, \nu_k)\}_{k=1}^D$ and the message symbols $\{m_k\}_{k=1}^M$ in (11.6) are said to be identifiable [12] if

$$\widetilde{\mathbf{Q}}(\boldsymbol{\tau}, \boldsymbol{\nu}, \mathbf{m})\widetilde{\mathbf{A}} \neq \widetilde{\mathbf{Q}}(\boldsymbol{\tau}', \boldsymbol{\nu}', \mathbf{m}')\widetilde{\mathbf{A}}'$$

holds for $\boldsymbol{\tau} \neq \boldsymbol{\tau}'$, $\boldsymbol{\nu} \neq \boldsymbol{\nu}'$, $\mathbf{m} \neq \mathbf{m}'$ and for any $\widetilde{\mathbf{A}}'$. Similar to [12] we employ in our model an S dimensional signal manifold given by $\tilde{\mathbf{q}}(\tau, \nu, m)$ describing the columns of $\widetilde{\mathbf{Q}}(\boldsymbol{\tau}, \boldsymbol{\nu}, \mathbf{m})$. The signal manifold is said to be *unambiguous* if every collection of S distinct vectors $\tilde{\mathbf{q}}(\tau, \nu, m)$ is linearly independent. We can now formulate the following identifiability result.

Theorem 11.2.1. *Given the signal model as described in* (11.6) *and assuming that the signal manifold is unambiguous. Then the messages* $\{m_\ell\}_{\ell=1}^{M}$ *are uniquely identifiable, if*

$$M < \frac{S + r_{\mathbf{A}}}{2D} \tag{11.7}$$

holds, where $r_{\mathbf{A}} = \operatorname{rank}(\mathbf{A})$.

Proof. We have to show that provided (11.7), we always have

$$\widetilde{\mathbf{Q}}(\boldsymbol{\tau}, \boldsymbol{\nu}, \mathbf{m})\widetilde{\mathbf{A}} \neq \widetilde{\mathbf{Q}}(\boldsymbol{\tau}', \boldsymbol{\nu}', \mathbf{m}')\widetilde{\mathbf{A}}'$$

for $\boldsymbol{\tau} \neq \boldsymbol{\tau}'$, $\boldsymbol{\nu} \neq \boldsymbol{\nu}'$, $\mathbf{m} \neq \mathbf{m}'$ and for any $\widetilde{\mathbf{A}}'$. The proof mainly follows the steps in [143]. In order to prove the statement we distinguish two cases.
Case 1: Set $\widetilde{\mathbf{Q}}' = \widetilde{\mathbf{Q}}(\boldsymbol{\tau}', \boldsymbol{\nu}', \mathbf{m}')$ and assume that $\widetilde{\mathbf{Q}}$ and $\widetilde{\mathbf{Q}}'$ have d columns in common and we have $0 \leq d \leq 2DM - S$, then

$$\begin{bmatrix} \widetilde{\mathbf{Q}} & \widetilde{\mathbf{Q}}' \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{A}} \\ -\widetilde{\mathbf{A}}' \end{bmatrix} \neq 0$$

holds, if and only if

$$\dim\left(\ker\left(\begin{bmatrix} \widetilde{\mathbf{Q}} & \widetilde{\mathbf{Q}}' \end{bmatrix}\right)\right) < \operatorname{rank}\left(\begin{bmatrix} \widetilde{\mathbf{A}} & -\widetilde{\mathbf{A}}' \end{bmatrix}^T\right).$$

Obviously, we have

$$r_{\mathbf{A}} < \operatorname{rank}\left(\begin{bmatrix} \widetilde{\mathbf{A}} & -\widetilde{\mathbf{A}}' \end{bmatrix}^T\right) \text{ and } \operatorname{rank}\left(\begin{bmatrix} \widetilde{\mathbf{Q}} & \widetilde{\mathbf{Q}}' \end{bmatrix}\right) = 2DM - d\,,$$

hence we directly obtain $2DM - S < r_{\mathbf{A}}$.
Case 2: Suppose that for the common column number holds $DM > d > 2DM - S$. We have

$$\begin{bmatrix} \widetilde{\mathbf{Q}} & \widehat{\mathbf{Q}}' \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{A}} \\ -\widehat{\mathbf{A}}' \end{bmatrix} \neq 0\,, \tag{11.8}$$

where $\widehat{\mathbf{Q}}'$ is the $S \times (DM - d)$ matrix obtained from $\widetilde{\mathbf{Q}}'$ by deleting the d

identical columns of $\widetilde{\mathbf{Q}}$, $\widehat{\mathbf{A}}'$ is the $(DM-d)\times N_A$ matrix obtained by deleting the d rows of $\widehat{\mathbf{A}}'$ corresponding to the deleted columns in $\widetilde{\mathbf{Q}}$ and $\widehat{\mathbf{A}}$ denotes the $DM\times N_A$ matrix obtained from $\widetilde{\mathbf{A}}$ by subtracting the deleted rows of $\widehat{\mathbf{A}}'$ from the corresponding rows in $\widetilde{\mathbf{A}}$. We have

$$l=\dim\left(\ker\left(\begin{bmatrix}\widetilde{\mathbf{Q}} & \widehat{\mathbf{Q}}'\end{bmatrix}\right)\right)=(2DM-d)-\operatorname{rank}\left(\begin{bmatrix}\widetilde{\mathbf{Q}} & \widehat{\mathbf{Q}}'\end{bmatrix}\right).$$

Hence, (11.8) holds if and only if

$$l<\operatorname{rank}\left(\begin{bmatrix}\widehat{\mathbf{A}} & -\widehat{\mathbf{A}}'\end{bmatrix}^T\right). \tag{11.9}$$

By the unambiguity assumption and with $S>2DM-d$ we deduce $l=0$ and since $\begin{bmatrix}\widehat{\mathbf{A}} & -\widehat{\mathbf{A}}'\end{bmatrix}^T\neq 0$, (11.9) holds. □

Note that by suitable choice of pilot and carrier functions in (11.5) the condition of unambiguity can be fulfilled.

11.3. Data estimation with MUSIC

Theorem 11.2.1 guarantees uniqueness of the message given the condition (11.7) holds. However, it does not deliver a reconstruction procedure for message estimation. In this section we present a recovery method based on the MUSIC algorithm [142]. Later in Section 11.4 the performance of the suggested scheme is heuristically evaluated and it becomes apparent that the suggested method achieves the bound in Theorem 11.2.1 for reasonable noise power levels.

Assuming the Doppler shifts to be small it is possible to simplify (11.6) [12]. To this end we consider the Taylor series expansion of $s(\omega)$ at the sample points $\{\omega_i\}_{i=1}^S$ and neglect the higher order terms. We then have,

$$s(\omega_i-\nu_k)\approx s(\omega_i)-\nu_k s'(\omega_i),$$

where $s'(\omega)$ denotes the derivate of $s(\omega)$. We can now approximate the measurement $\mathbf{Y}$ by

$$\mathbf{Y}\approx\left[\mathbf{P}\mathbf{V}_{\boldsymbol{\tau}}-\mathbf{P}'\mathbf{V}_{\boldsymbol{\tau}}\mathbf{D}_{\boldsymbol{\nu}}\right]\mathbf{A}+\sum_{\ell=1}^{M}m_\ell\left[\mathbf{C}_\ell\mathbf{V}_{\boldsymbol{\tau}}-\mathbf{C}'_\ell\mathbf{V}_{\boldsymbol{\tau}}\mathbf{D}_{\boldsymbol{\nu}}\right]\mathbf{A}+\mathbf{N}$$

$$= \begin{bmatrix} \boldsymbol{\Gamma}_{\mathbf{P}}(\boldsymbol{\tau},\boldsymbol{\nu}) & m_1\boldsymbol{\Gamma}_{\mathbf{C}_1}(\boldsymbol{\tau},\boldsymbol{\nu}) & \dots & m_M\boldsymbol{\Gamma}_{\mathbf{C}_M}(\boldsymbol{\tau},\boldsymbol{\nu}) \end{bmatrix} \begin{bmatrix} \mathbf{A} \\ \vdots \\ \mathbf{A} \end{bmatrix} + \mathbf{N}$$

$$= \widetilde{\boldsymbol{\Gamma}}_{(\boldsymbol{\tau},\boldsymbol{\nu},\mathbf{m})}\widetilde{\mathbf{A}} + \mathbf{N} \tag{11.10}$$

where $\mathbf{N}$ is the measurement noise, and

$$\begin{aligned} \mathbf{P} &= \operatorname{diag}\left(p(\omega_1) \ \dots \ p(\omega_S)\right) \in \mathbb{R}^{S\times S} \\ \mathbf{P}' &= \operatorname{diag}\left(p'(\omega_1) \ \dots \ p'(\omega_S)\right) \in \mathbb{R}^{S\times S} \\ \mathbf{C}_\ell &= \operatorname{diag}\left(c_\ell(\omega_1) \ \dots \ c_\ell(\omega_S)\right) \in \mathbb{R}^{S\times S} \\ \mathbf{C}'_\ell &= \operatorname{diag}\left(c'_\ell(\omega_1) \ \dots \ c'_\ell(\omega_S)\right) \in \mathbb{R}^{S\times S} \\ \mathbf{D}_{\boldsymbol{\nu}} &= \operatorname{diag}\left(\nu_1 \ \dots \ \nu_D\right) \in \mathbb{R}^{D\times D}, \end{aligned}$$

with $p'(\omega) = \frac{d}{d\omega}p(\omega)$ and $c'_\ell(\omega) = \frac{d}{d\omega}c_\ell(\omega)$. Two Doppler shifts sharing the same delay can't be distinguished in (11.10), however, this has no effect on the suggested message estimation method.

Denote by $\mathbf{E}_s$ the left singular vectors of $\mathbf{Y}$ corresponding to the D largest singular values. Let $\mathbf{E}_n$ be a $S \times (S-D)$ matrix whose columns are orthogonal to those of $\mathbf{E}_s$. The columns of $\mathbf{E}_n$ span the noise subspace.

Next, note that the submatrix $\boldsymbol{\gamma}_{\tau,\nu}$ of $\widetilde{\boldsymbol{\Gamma}}_{(\boldsymbol{\tau},\boldsymbol{\nu},\mathbf{m})}$ containing all the columns depending on the delay-Doppler shift (τ,ν) is given by

$$\begin{aligned} \boldsymbol{\gamma}_{\tau,\nu} = \Big[&\Big(\mathbf{P}\mathbf{v}(\tau) - \nu\mathbf{P}'\mathbf{v}(\tau)\Big) \quad m_1\Big(\mathbf{C}_1\mathbf{v}(\tau) - \nu\mathbf{C}'_1\mathbf{v}(\tau)\Big) \\ &\dots \quad m_M\Big(\mathbf{C}_M\mathbf{v}(\tau) - \nu\mathbf{C}'_M\mathbf{v}(\tau)\Big)\Big] = \mathbf{G}_\tau\mathbf{g}_{\nu,\mathbf{m}}, \end{aligned}$$

where $\mathbf{g}_{\nu,\mathbf{m}} = [1 \ \nu \ m_1 \ \nu m_1 \ \dots \ m_M \ \nu m_M]$ and

$$\begin{aligned} \mathbf{G}_\tau = \Big[&\mathbf{P}\mathbf{v}(\tau) \ \ -\mathbf{P}'\mathbf{v}(\tau) \ \ \mathbf{C}_1\mathbf{v}(\tau) \ \ -\mathbf{C}'_1\mathbf{v}(\tau) \\ &\dots \ \ \mathbf{C}_M\mathbf{v}(\tau) \ \ -\mathbf{C}'_M\mathbf{v}(\tau)\Big]. \end{aligned}$$

The MUSIC loss function for the present scenario is described as

$$\mathrm{L}(\tau,\nu) = \frac{\mathbf{g}^*_{\nu,\mathbf{m}}[\operatorname{Re}\left(\mathbf{G}^*_\tau\mathbf{E}_n\mathbf{E}^*_n\mathbf{G}_\tau\right)]\mathbf{g}_{\nu,\mathbf{m}}}{\mathbf{g}^*_{\nu,\mathbf{m}}[\operatorname{Re}\left(\mathbf{G}^*_\tau\mathbf{G}_\tau\right)]\mathbf{g}_{\nu,\mathbf{m}}}.$$

Minimizing $\mathrm{L}(\tau,\nu)$ with respect to $\mathbf{g}_{\nu,\mathbf{m}}$ is equivalent to finding the minimum

eigenvalue and the corresponding eigenvector of the following $(2M+2)\times(2M+2)$ matrices

$$\mathrm{Re}\left(\mathbf{G}_\tau^*\mathbf{E}_n\mathbf{E}_n^*\mathbf{G}_\tau\right)\xi_{\min} = \lambda_{\min}\mathrm{Re}\left(\mathbf{G}_\tau^*\mathbf{G}_\tau\right)\xi_{\min}$$

The D deepest minima of $\lambda_{\min}$ as a function of τ yield the time delays, corresponding Doppler shifts can be found by

$$\hat{\nu}_k = \frac{\xi_{\min,2}(\hat{\tau}_k)}{\xi_{\min,1}(\hat{\tau}_k)}$$

where $\hat{\tau}_k$ is the estimate of τ_k and $\xi_{\min,i}$ is the ith element of $\xi_{\min}$ [12]. Finally, fixing an estimate $\hat{\tau}$, the ℓth message symbol is obtained by

$$m_\ell = \frac{\xi_{\min,2\ell+1}(\hat{\tau})}{\xi_{\min,1}(\hat{\tau})}\,.$$

Note that computations can be speeded-up by separately computing $\mathrm{Re}\left(\mathbf{G}_\tau^*\mathbf{G}_\tau\right)$, since it has no dependence on τ, ν or $\mathbf{m}$. Hence, computations can be speeded up by computing the matrix $\mathrm{Re}\left(\mathbf{G}_\tau^*\mathbf{G}_\tau\right)$ separately.

11.4. Numerical results

In order to demonstrate the performance of the suggested scheme multiple numerical experiments were carried out.

In all experiments the spatial signatures $\mathbf{a}_k$ were constructed randomly. Each entry in the array response vector had absolute value one and a uniformly distributed random phase factor. The attenuation factors of different paths where chosen as Gaussian random variables $\sim\mathcal{N}(0,1)$. The delay-Doppler shift pairs were chosen uniformly random s.t. $(\tau,\nu)\in[0,2\pi]\times[-\frac{\pi}{8},\frac{\pi}{8}]$. The noise was assumed to be zero mean Gaussian. Each message vector $\mathbf{m}$ was chosen uniformly random over the M dimensional unit ball. A message estimate, $\hat{\mathbf{m}}$, was accepted as a correct recovery if the euclidean distance between the actual message and its estimate was smaller than 0.1, i.e. $\|\mathbf{m}-\hat{\mathbf{m}}\|_2 < 0.1$.

In both Figures 11.1 and 11.2, for each channel, 10 messages were generated randomly. The recovery rates in all figures are computed from 100 channel realizations. Hence, for each data point, 1000 experiments were performed and the number of correct recoveries were counted. The final recovery rate is obtained by dividing the number of correctly recovered messages by 1000. Consequently, a recovery rate of 1 implies that every message was estimated correctly.

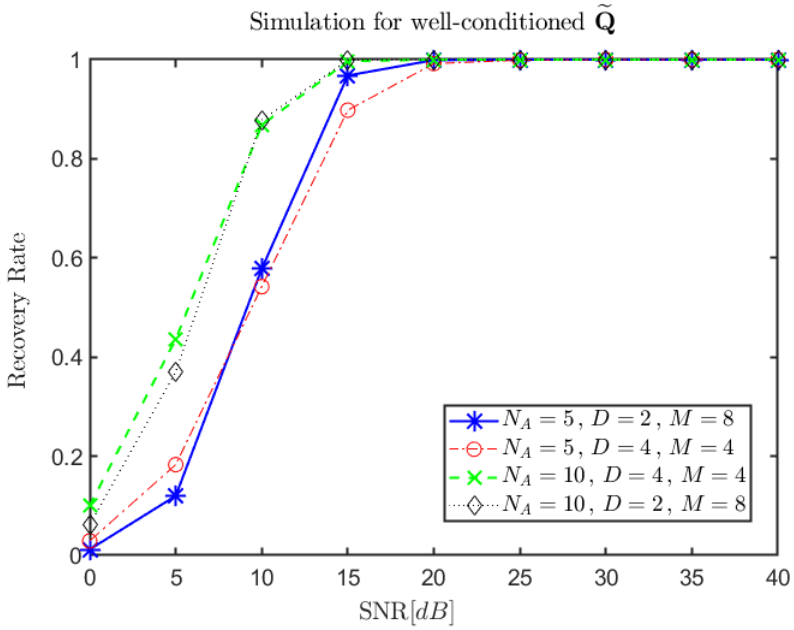

Figure 11.1.: Recovery rate for fixed number of samples $S = 32$, varying numbers of delay-Doppler shifts $D = 2, 4$, message lengths $M = 4, 8$ and antennas $N_A = 5, 10$. The pilot and carrier functions are chosen such that $\widetilde{\mathbf{Q}}$ is well-conditioned.

Figure 11.1 depicts the recovery rate to *signal to noise ratio* (SNR) in dB for small antenna array setups with $N_A = 5, 10$ antennas and $S = 32$ samples. Pilot and carrier signals were chosen as

$$p(\omega) = \sin(\omega) \quad \text{and} \quad c_\ell(\omega) = \sin((\ell + 1) \cdot \omega)\,,$$

this choice of functions results in a well-conditioned $\widetilde{\mathbf{Q}}$ in (11.10). Following the uniqueness condition in Theorem 11.2.1, varying sizes of messages $M = 4, 8$ and delay-Doppler shifts $D = 2, 4$ were chosen. In Figure 11.1, an apparent result is that the suggested scheme achieves the identifiability bound obtained in Theorem 11.2.1 for reasonable SNR values.

Figure 11.2 illustrates the recovery rate to SNR in dB for a large antenna array setup with $N_A = 50$ antennas. The number of samples was fixed to $S = 128$. Illustrated is the recovery rates for varying numbers of delay-Doppler shifts $D = 2, 4$ and different message lengths $M = 2, 4$. The pilot and carrier signals

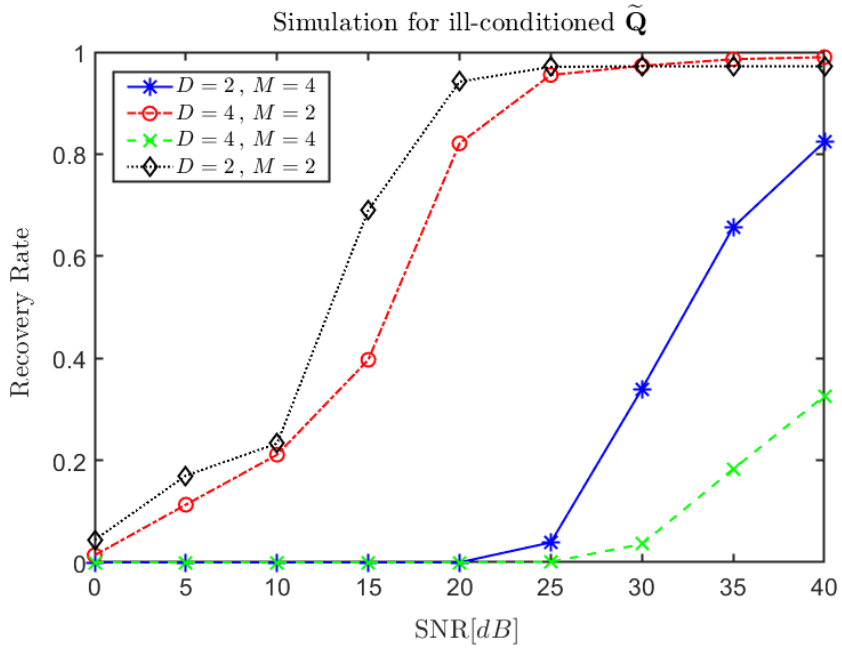

Figure 11.2.: Recovery rate for fixed number of samples $S = 128$ and antennas $N_A = 50$, varying numbers of delay-Doppler shifts $D = 2, 4$ and message lengths $M = 2, 4$. The pilot and carrier functions are chosen such that $\widetilde{\mathbf{Q}}$ is ill-conditioned.

were chosen as

$$p(\omega) = \mathrm{sinc}(\omega) \quad \text{and} \quad c_\ell(\omega) = \mathrm{sinc}^{(\ell+1)}(\omega)\,,$$

with $\mathrm{sinc}(\omega) = \frac{\sin(\pi\omega)}{\pi\omega}$. This choice results in an ill-conditioned $\widetilde{\mathbf{Q}}$.
In Figure 11.2 the achievable bounds given by Theorem 11.2.1 would be, $M < 16$ and $M < 32$ for $D = 4$ and $D = 2$, respectively. Additionally, by the large number of antennas the setup in Figure 11.2 is more robust to noise. However, already for relatively small sizes of messages the second setup performs very poorly compared to the receiver constellation in Figure 11.1. The reason for it can be traced back to the choice of the pilot and carrier functions. In the second figure, this choice results in an ill-conditioned $\widetilde{\mathbf{Q}}$, amplifying the sensitivity to noise, hence, drastically reducing the message length which can be transmitted. Note that if $\widetilde{\mathbf{Q}}$ is either ill-conditioned or well-conditioned solely depends on the choice of $p(\omega)$ and $c_\ell(\omega)$. Therefore, a suitable choice of pilot and carrier functions guarantees the performance of the presented method.

Bibliography

[1] Lang Tong, Brian M. Sadler, and Min Dong. "Pilot-assisted wireless transmissions: general model, design criteria, and signal processing". In: *IEEE Signal Process. Mag.* vol. 21, no. 6, pp. 12–25 (2004).

[2] Bernard Sklar. "Rayleigh fading channels in mobile digital communication systems. I. Characterization". In: *IEEE Comm. Mag.* vol. 35, no. 7, pp. 90–100 (1997).

[3] Franz Hlawatsch and Gerald Matz. *Wireless Communications over Rapidly Time-Varying Channels*. Academic Press, Oxford, 2011.

[4] Trym H. Eggen, Arthur B. Baggeroer, and James C. Preisig. "Communication over Doppler spread channels. Part I: Channel and receiver presentation". In: *IEEE Journal of Oceanic Engineering* vol. 25, no. 1, pp. 62–71 (2000).

[5] Mark Johnson, Lee Freitag, and Milica Stojanovic. "Improved Doppler tracking and correction for underwater acoustic communications". In: *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Munich, Germany*. 1997.

[6] Michail K. Tsatsanis and Georgios B. Giannakis. "Modelling and equalization of rapidly fading channels". In: *International Journal of Adaptive Control and Signal Process.* vol. 10, pp. 159–176 (1996).

[7] Kien T. Truong and Robert W. Heath. "Effects of channel aging in massive MIMO systems". In: *IEEE Journal of Comm. and Networks* vol. 15, no. 4, pp. 338–351 (2013).

[8] Waheed U. Bajwa et al. "Compressed channel sensing: A new approach to estimating sparse multipath channels". In: *Proc. IEEE* vol. 98, no. 6, pp. 1058–1076 (2010).

[9] Akbar M. Sayeed. "Deconstructing multiantenna fading channels". In: *IEEE Trans. Signal Process.* vol. 50, no. 10, pp. 2563–2579 (2002).

[10] David Tse and Pramod Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, Cambridge, 2005.

[11] Thomas Kailath. “Measurements on time-variant communication channels”. In: *IRE Trans. Inf. Theory* vol. 8, no. 5, pp. 229–236 (1962).

[12] Andreas Jakobsson, A. Lee Swindlehurst, and Petre Stoica. “Subspace-based estimation of time delays and Doppler shifts”. In: *IEEE Trans. Signal Process.* vol. 46, no. 9, pp. 2472–2483 (1998).

[13] Götz E. Pfander and David F. Walnut. “Measurement of time-variant linear channels”. In: *IEEE Trans. Inf. Theory* vol. 52, no. 11, pp. 4808–4820 (2006).

[14] Volker Pohl et al. “Continuous flat-fading MIMO channels: Achievable rate and optimal length of the training and data phases”. In: *IEEE Trans. Wireless Commun.* vol. 4, no. 4, pp. 1889–1900 (2005).

[15] Dennis Gabor. “Theory of communication. Part 1: The analysis of information”. In: *Journal Inst. Electr. Eng.* vol. 93, no. 26, pp. 429–441 (1946).

[16] Lotfi Asker Zadeh. “Frequency analysis of variable networks”. In: *Proc. IRE* vol. 38, no. 3, pp. 291–299 (1950).

[17] Philip Bello. “Characterization of randomly time-variant linear channels”. In: *IEEE Trans. Commun. Syst.* vol. 11, no. 4, pp. 360–393 (1963).

[18] Robert W. Chang. “Synthesis of band-limited orthogonal signals for multichannel data transmission”. In: *Bell Syst. Tech. Journal* vol. 45, no. 10, pp. 1775–1796 (1966).

[19] Götz E. Pfander. “Gabor Frames in Finite Dimensions”. In: *Finite Frames*. Ed. by Peter G. Casazza and Gitta Kutyniok. Springer, Birkhäuser Boston, 2012, pp. 193–239.

[20] Badri Narayan Bhaskar, Gongguo Tang, and Benjamin Recht. “Atomic norm denoising with applications to line spectral estimation”. In: *IEEE Trans. Signal Process.* vol. 61, no. 23, pp. 5987–5999 (2013).

[21] Gerlind Plonka and Manfred Tasche. “Prony methods for recovery of structured functions”. In: *GAMM-Mitteilungen* vol. 37, no. 2, pp. 239–258 (2014).

[22] Tomas Sauer. “Prony’s method: An old trick for new problems”. In: *Snapshots Modern Math., Oberwolfach* (2018).

[23] Yingbo Hua and Tapan K. Sarkar. “Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise”. In: *IEEE Trans. on Acoust. Speech and Signal Process.* vol. 38, no. 5, pp. 814–824 (1990).

[24] Stephen B. Weinstein and Paul M. Ebert. "Data transmission by frequency-division multiplexing using the discrete Fourier transform". In: *IEEE Trans. on Commun. Technology* vol. 19, no. 5, pp. 628–634 (1971).

[25] Andrea Goldsmith. *Wireless communications*. Cambridge Univ. Press, Cambridge, UK, 2005.

[26] John R. Klauder et al. "The theory and design of chirp radars". In: *The Bell Systems Tech. Journal* vol. 39, no. 4, pp. 745–808 (1960).

[27] Alihan **Kaplan**, Volker Pohl, and Dae Gwan Lee. "The Statistical Restricted Isometry Property for Gabor Systems". In: *in Proc. IEEE Statistical Signal Processing Workshop (SSP), Freiburg, Germany*. 2018.

[28] Alihan **Kaplan**, Volker Pohl, and Holger Boche. "Deterministic Matrices with a Restricted Isometry Property for Partially Structured Sparse Signals". In: *in Proc. 13th International Conference on Sampling Theory and Applications (SampTA), Bordeaux, France*. 2019.

[29] Alihan **Kaplan**, Volker Pohl, and Holger Boche. "A New Family of Low Coherence Finite Gabor Frames and Applications in Compressive Sampling". In: *in Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Cannes, France*. 2019.

[30] Alihan **Kaplan**, Volker Pohl, and Dae Gwan Lee. "On Compressive Sensing of Sparse Covariance Matrices using Deterministic Sensing Matrices". In: *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, Canada*. 2018.

[31] Alihan **Kaplan**, Dae Gwan Lee, and Volker Pohl. "Message Transmission ThroughUnderspread Time-Varying Linear Channels". In: *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain*. 2020.

[32] Alihan **Kaplan**, Dae Gwan Lee, and Volker Pohl. "A message transmission scheme for linear time-varying multipath channels". In: *in Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Atlanta, USA*. 2020.

[33] Alihan **Kaplan** and Volker Pohl. "Message Transmission Over Rapidly Time-Varying Channels". In: *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, Canada*. 2021.

[34] Alihan **Kaplan**, Volker Pohl, and Dae Gwan Lee. "Data Transmission Over Linear Time-Varying Channels". In: *IEEE Trans. Signal Process.* vol. 70, pp. 3357–3370 (2022).

[35] Alihan **Kaplan** et al. “Sparse Deterministic and Stochastic Channels: Identification of Spreading Functions and Covariances”. In: *Compressed Sensing in Information Processing*. Ed. by Gitta Kutyniok, Holger Rauhut, and Robert J. Kunsch. Springer, Birkhäuser Cham, 2022, pp. 105–144.

[36] Dae Gwan Lee, Alihan **Kaplan**, and Volker Pohl. “Permissible Support Patterns for Identifying the Spreading Function of Time-Varying Channels”. In: *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, Canada*. 2018.

[37] Simon Foucard and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhäuser, New York, 2013.

[38] Emmanuel J. Candès and Michael B. Wakin. “An Introduction to Compressive Sampling”. In: *IEEE Signal Process. Mag.* vol. 25, no. 2, pp. 21–30 (2008).

[39] David L. Donoho. “Compressed Sensing”. In: *IEEE Trans. Inf. Theory* vol. 52, no. 4, pp. 1289–1306 (2006).

[40] Simon Foucart. “Flavors of Compressive Sensing”. In: *Approximation Theory XV: San Antonio 2016*. Ed. by Gregory E. Fasshauer and Larry L. Schumaker. Springer International Publishing, Cham, 2017, pp. 61–104.

[41] Gerhard Wunder et al. “Sparse Signal Processing Concepts for Efficient 5G System Design”. In: *IEEE Access* vol. 3, pp. 195–208 (2015).

[42] Zhu Han, Husheng Li, and Wotao Yin. *Compressive Sensing for Wireless Networks*. Cambridge Univ. Press, Cambridge, 2013.

[43] Matthew A. Herman and Thomas Strohmer. “High-resolution Radar via Compressed Sensing”. In: *IEEE Trans. Signal Process.* vol. 57, no. 6, pp. 2275–2284 (2009).

[44] Michael Lustig et al. “Compressed Sensing MRI”. In: *IEEE Signal Process. Mag.* vol. 25, no. 2, pp. 72–82 (2008).

[45] Holger Boche et al. *Compressed Sensing and its Applications*. Birkhäuser, Cham, 2015.

[46] Angshul Majumdar. *Compressed Sensing for Engineers*. CRC Press, Boca Raton, 2018.

[47] Balas K. Natarajan. “Sparse Approximate Solutions to Linear Systems”. In: *SIAM Journal on Computing* vol. 24, no. 2, pp. 227–234 (1995).

[48] Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. "Compressed Sensing and best k-term Approximation". In: *Journal of the American Mathematical Society* vol. 22, no. 1, pp. 211–231 (2009).

[49] David L. Donoho and Michael Elad. "Optimally Sparse Representation in General (nonorthogonal) Dictionaries via ℓ_1-minimization". In: *Proceedings of the National Academy of Sciences* vol. 100, no. 5, pp. 2197–2202 (2003).

[50] Emmanuel J. Candes and Terence Tao. "Decoding by Linear Programming". In: *IEEE Trans. on Inf. Theory* vol. 51, no. 12, pp. 4203–4215 (2005).

[51] Emmanuel J. Candes, Justin K. Romberg, and Terence Tao. "Stable Signal Recovery from Incomplete and Inaccurate Measurements". In: *Commun. Pure Appl. Math.* vol. 59, no. 8, pp. 1207–1223 (2006).

[52] T. Tony Cai and Anru Zhang. "Sparse Representation of a Polytope and Recovery of Sparse Signals and Low-Rank Matrices". In: *IEEE Trans. on Inf. Theory* vol. 60, no. 1, pp. 122–132 (2013).

[53] Richard G. Baraniuk et al. "A Simple Proof of the Restricted Isometry Property for Random Matrices". In: *Constructive Approximation* (2007).

[54] Mark Rudelson and Roman Vershynin. "On Sparse Reconstruction from Fourier and Gaussian Measurement". In: *Commun. Pure Appl. Math.* vol. 61, no. 8, pp. 1025–1045 (2013).

[55] Andreas M. Tillmann and Marc E. Pfetsch. "The Computational Complexity of the Restrichted Isometry Property, the Nullspace Property, and Related Concetps in Compressed Sensing". In: *IEEE Trans. on Inf. Theory* vol. 60, no. 2, pp. 1248–1259 (2013).

[56] Afonso S. Bandeira et al. "Certifying the Restricted Isometry Property is Hard". In: *IEEE Trans. on Inf. Theory* vol. 59, no. 6, pp. 3448–3450 (2013).

[57] Ronald A. DeVore. "Deterministic Constructions of Compressed Sensing Matrices". In: *Journal of Complexity* vol. 23, pp. 918–925 (2007).

[58] Guangwu Xu and Zhiqiang Xu. "Compressed Sensing Matrices from Fourier Matrices". In: *IEEE Trans. on Inf. Theory* vol. 61, no. 1, pp. 469–478 (2014).

[59] Nam Yul Yu and Ying Li. "Deterministic Construction of Fourier-based Compressed Sensing Matrices using an Almost Difference Set". In: *EURASIP Journal on Advances in Signal Process.* vol. 2013, no. 1, pp. 1–14 (2013).

[60] Lorne Applebaum et al. "Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery". In: *Appl. Comput. Harmon. Anal.* vol. 26, no. 2, pp. 283–290 (2009).

[61] Afonso S. Bandeira et al. "The Road to Deterministic Matrices with the Restricted Isometry Property". In: *Journal of Fourier Analysis and Applications* vol. 19, no. 6, pp. 1123–1149 (2013).

[62] Jean Bourgain et al. "Explicit Constructions of RIP Matrices and Related Problems". In: *Duke Math. Journal* vol. 159, no. 1, pp. 145–185 (2011).

[63] David L. Donoho and Xiaoming Huo. "Uncertainty Principles and Ideal Atomic Decomposition". In: *IEEE Trans. on Inf. Theory* vol. 47, no. 7, pp. 2845–2862 (2001).

[64] Lloyd Welch. "Lower Bounds on the Maximum Cross Correlation of Signals". In: *IEEE Trans. on Inf. Theory* vol. 20, no. 3, pp. 397–399 (1974).

[65] Thomas Strohmer and Robert W. Heath Jr. "Grassmannian Frames with Applications to Coding and Communication". In: *Appl. Comput. Harmon. Anal.* vol. 14, no. 3, pp. 257–275 (2003).

[66] Roger A Horn and Charles R Johnson. *Matrix Analysis*. Cambridge University Press, New York, 2012.

[67] Waheed U. Bajwa, Robert Calderbank, and Dustin G. Mixon. "Two are better than one: Fundamental parameters of frame coherence". In: *Appl. Comput. Harmon. Anal.* vol. 33, no. 1, pp. 58–78 (2012).

[68] Shuxing Li and Gennian Ge. "Deterministic Construction of Sparse Sensing Matrices via Finite Geometry". In: *IEEE Trans. on Signal Process.* vol. 62, no. 11, pp. 2850–2859 (2014).

[69] Zhi Gu et al. "Deterministic Compressed Sensing Matrices from Sequences with Optimal Correlation". In: *IEEE Access* vol. 7, pp. 16704–16710 (2019).

[70] Robert Calderbank, Stephen Howard, and Sina Jafarpour. "Construction of a Large Class of Deterministic Sensing Matrices that Satisfy a Statistical Isometry Property". In: *IEEE Journal of Selected Topics in Signal Process.* vol. 4, no. 2, pp. 358–374 (2010).

[71] Waheed U. Bajwa, Robert Calderbank, and Sina Jafarpour. "Why Gabor frames? Two fundamental measures of coherence and their role in model selection". In: *IEEE Journal of Communications and Networks* vol. 12, no. 4, pp. 289–307 (2010).

[72] Karin Schnass. "Average performance of orthogonal matching pursuit (OMP) for sparse approximation". In: *IEEE Signal Process. Letters* vol. 25, no. 12, pp. 1865–1869 (2018).

[73] Shamgar Gurevich and Ronny Hadani. "Incoherent dictionaries and the statistical restricted isometry property". In: *arXiv preprint arXiv:0809.1687* (2008).

[74] Shamgar Gurevich and Ronny Hadani. "The statistical restricted isometry property and the Wigner semicircle distribution of incoherent dictionaries". In: *arXiv preprint arXiv:0812.2602* (2008).

[75] Karlheinz Gröchenig. "Uncertainty principles for time-frequency representations". In: *Advances in Gabor analysis* pp. 11–30 (2003).

[76] Ole Christensen. *An introduction to frames and Riesz bases*. Springer, Birkhäuser Boston, 2003.

[77] Richard J. Duffin and Albert C. Schaeffer. "A class of nonharmonic Fourier series". In: *Trans. Am. Math. Soc.* vol. 72, no. 2, pp. 341–366 (1952).

[78] Peter G. Casazza, Gitta Kutyniok, and Friedrich Philipp. "Introduction to Finite Frame Theory". In: *Finite Frames*. Ed. by Peter G. Casazza and Gitta Kutyniok. Springer, Birkhäuser Boston, 2012, pp. 1–51.

[79] Hans G. Feichtinger and Karlheinz Gröchenig. "Gabor frames and time-frequency analysis of distributions". In: *Journal Funct. Anal.* vol. 146, no. 2, pp. 464–495 (1997).

[80] Stephen Boyd, Stephen P. Boyd, and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[81] Hans Georg Feichtinger and Karlheinz Gröchenig. "Acceleration of the Frame Algorithm". In: *IEEE Trans. Signal Process.* vol. 41, no. 12, pp. 3331–3340 (1993).

[82] Hans G. Feichtinger, Werner Kozek, and Franz Luef. "Gabor Analysis over Finite Abelian Groups". In: *Appl. Comput. Harmon. Anal.* vol. 26, no. 2, pp. 230–248 (2009).

[83] Stephen D. Howard, Robert Calderbank, and William Moran. "The finite Heisenberg-Weyl groups in radar and communications". In: *EURASIP Journal on Adv. in Signal Process.* (2006), pp. 1–12.

[84] Jim Lawrence, Götz E. Pfander, and David Walnut. "Linear independence of Gabor systems in finite dimensional vector spaces". In: *Journal of Fourier Analysis and Applications* vol. 11, pp. 715–726 (2005).

[85] Romanos-Diogenes Malikiosis. "A note on Gabor frames in finite dimensions". In: *Appl. Comput. Harmon. Anal.* vol. 38, no. 2, pp. 318–330 (2015).

[86] Götz E. Pfander, Holger Rauhut, and Jared Tanner. "Identification of matrices having a sparse representation". In: *IEEE Trans. Signal Process.* vol. 56, no. 11, pp. 5376–5388 (2008).

[87] Felix Krahmer, Götz E. Pfander, and Peter Rashkov. "Uncertainty in time–frequency representations on finite Abelian groups and applications". In: *Appl. Comput. Harmon. Analy.* vol. 25, no. 2, pp. 209–225 (2008).

[88] Götz E. Pfander and Holger Rauhut. "Sparsity in time-frequency representations". In: *Journal of Fourier Analysis and Applications* vol. 16, no. 2, pp. 233–260 (2010).

[89] Felix Krahmer, Shahar Mendelson, and Holger Rauhut. "Suprema of chaos processes and the restricted isometry property". In: *Comm. on Pure and Appl. Mathematics* vol. 67, no. 11, pp. 1877–1904 (2014).

[90] Shamgar Gurevich, Ronny Hadani, and Nir Sochen. "The finite harmonic oscillator and its applications to sequences, communication, and radar". In: *IEEE Trans. on Inf. Theory* vol. 54, no. 9, pp. 4239–4253 (2008).

[91] W. Alltop. "Complex sequences with low periodic correlations (corresp.)" In: *IEEE Trans. on Inf. Theory* vol. 26, no. 3, pp. 350–354 (1980).

[92] François Auger et al. *Time-frequency toolbox*. CNRS France and Rice University, 1996.

[93] Ole Christensen, Hans G. Feichtinger, and Stephan Paukner. "Gabor analysis for imaging". In: *Handbook of Mathematical Methods in Imaging*. Ed. by Otmar Scherzer. Springer, New York, 2015, pp. 1717–1757.

[94] Rudolf Lidl and Harald Niederreiter. *Finite Fields*. Cambridge University Press, 1997.

[95] Bao Luong. *Fourier analysis on finite Abelian groups*. Springer, Birkhäuser Boston, 2009.

[96] Robert M. Gray. "Toeplitz and circulant matrices: A review". In: *Foundations and Trends in Communications and Information Theory* vol. 2, no. 3, pp. 155–239 (2006).

[97] Franz Hlawatsch and Gloria Faye Boudreaux-Bartels. "Linear and quadratic time-frequency signal representations". In: *IEEE Signal Process. Magazine* vol. 9, no. 2, pp. 21–67 (1992).

[98] Daniel Griffin and Jae Lim. "Signal estimation from modified short-time Fourier transform". In: *IEEE Trans. on Acoustics, Speech and Signal Process.* vol. 32, no. 2, pp. 236–243 (1984).

[99] Jont B. Allen and Lawrence R. Rabiner. "A unified approach to short-time Fourier analysis and synthesis". In: *Proceedings of the IEEE* vol. 65, no. 11, pp. 1558–1564 (1977).

[100] David L. Donoho and Philip B. Stark. "Uncertainty principles and signal recovery". In: *SIAM Journal on Appl. Math.* vol. 49, no. 3, pp. 906–931 (1989).

[101] Gerald B. Folland and Alladi Sitaram. "The uncertainty principle: a mathematical survey". In: *Journal of Fourier Analy. and Appl.* vol. 3 no. 3, pp. 207–238 (1997).

[102] Terence Tao. "An uncertainty principle for cyclic groups of prime order". In: *Math. Res. Lett.* vol. 12, no. 1, pp. 121–127 (2005).

[103] Thomas Strohmer and Benjamin Friedlander. "Analysis of sparse MIMO radar". In: *Appl. Comput. Harmon. Analy.* vol. 37 no. 3, pp. 361–388 (2014).

[104] Götz E. Pfander and David F. Walnut. "Measurement of time-variant linear channels". In: *IEEE Trans. Inf. Theory* vol. 52, no. 11 pp. 4808–4820 (2006).

[105] Götz E. Pfander, Holger Rauhut, and Joel A. Tropp. "The restricted isometry property for time–frequency structured random matrices". In: *Probability Theory and Related Fields* vol. 156, pp. 707–737 (2013).

[106] Dominik Dorsch and Holger Rauhut. "Refined analysis of sparse MIMO radar". In: *Journal of Fourier Analy. and Appl.* vol. 23, pp. 485–529 (2017).

[107] Reinhard Heckel and Helmut Bölcskei. "Identification of sparse linear operators". In: *IEEE Trans. Inf. Theory* vol. 59, no. 12 pp. 7985–8000 (2013).

[108] Alexander Fish et al. "Delay-Doppler channel estimation in almost linear complexity". In: *IEEE Trans. Inf. Theory* vol. 59, no. 11, pp. 7632–7644 (2013).

[109] Afonso S. Bandeira, Nikolaus Doppelbauer, and Dmitriy Kunisky. "Dual bounds for the positive definite functions approach to mutually unbiased bases". In: *Sampling Theory, Signal Processing, and Data Analysis* vol. 20, no. 2, pp. 18 (2022).

[110] Somshubhro Bandyopadhyay et al. "A new proof for the existence of mutually unbiased bases". In: *Algorithmica* vol. 34, no. 4, pp. 512–528 (2002).

[111] Philippe Delsarte, J.M. Goethals, and J.J. Seidel. "Bounds for systems of lines, and Jacobi polynomials". In: *Geometry and Combinatorics*. Ed. by D.G. Corneil and R. Mathon. Academic Press, Elsevier, 1991, pp. 193–207.

[112] John Stillwell. *Elements of number theory*. Springer Science & Business Media, 2003.

[113] Graham Everest and Thomas Ward. *An introduction to number theory*. Springer Science & Business Media, 2005.

[114] Colin McDiarmid. "On the method of bounded differences". In: *Surveys in combinatorics* vol. 141, no. 1, pp. 148–188 (1989).

[115] Joel A. Tropp. "On the conditioning of random subdictionaries". In: *Appl. Comput. Harmon. Analy.* vol. 25, no. 1, pp. 1–24 (2008).

[116] John J. Benedetto, Robert L. Benedetto, and Joseph T. Woodworth. "Optimal ambiguity functions and Weil's exponential sum bound". In: *Journal of Fourier Analy. and Appl.* vol. 18, no. 3, pp. 471–487 (2012).

[117] Florence Jessie MacWilliams and Neil James Alexander Sloane. *The theory of error-correcting codes*. North-Holland, Amsterdam, 1977.

[118] Douglas Robert Stinson. *Combinatorial designs: constructions and analysis*. Springer, New York, 2004.

[119] Leonard D. Baumert. *Cyclic Difference Sets*. Springer, Heidelberg, 1971.

[120] James Singer. "A theorem in finite projective geometry and some applications to number theory". In: *Trans. Amer. Math. Soc.* vol. 43, no. 3, pp. 377–385 (1938).

[121] Michael Elad. *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer, 2010.

[122] Joel A. Tropp. "Greed is good: Algorithmic results for sparse approximation". In: *IEEE Trans. Inf. Theory* vol. 50, no. 10, pp. 2231–2242 (2004).

[123] Björn Ottersten, Peter Stoica, and Richard Roy. "Covariance matching estimation techniques for array signal processing applications". In: *Digital Signal Processing* vol. 8, no. 3, pp. 185–210 (1998).

[124] Daniel Romero et al. "Compressive covariance sensing: Structure-based compressive sensing beyond sparsity". In: *IEEE Signal Process. Mag.* vol. 33, no. 1, pp. 78–93 (2015).

[125] Gautam Dasarathy et al. "Sketching sparse matrices, covariances, and graphs via tensor products". In: *IEEE Trans. Inf. Theory* vol. 61, no. 3, pp. 1373–1388 (2015).

[126] Shayne Waldron. "Group frames". In: *Finite Frames*. Ed. by Peter G Casazza and Gitta Kutyniok. Springer, Birkhäuser Boston, 2012, pp. 209–243.

[127] Gerald Matz, Helmut Bolcskei, and Franz Hlawatsch. "Time-frequency foundations of communications: Concepts and tools". In: *IEEE Signal Process. Magazine* vol. 30, no. 6, pp. 87–96 (2013).

[128] David Walnut, Götz E. Pfander, and Thomas Kailath. "Cornerstones of Sampling of Operator Theory". In: *Excursions in Harmonic Analysis* vol. 4, pp. 291–332 (2015).

[129] Karlheinz Gröchenig. *Foundations of Time-Frequency Analysis*. Birkhäuser, Boston, 2001.

[130] Dae Gwan Lee, Götz E. Pfander, and Volker Pohl. "Sampling and reconstruction of multiple-input multiple-output channels". In: *IEEE Trans. Signal Process.* vol. 67, no. 4, pp. 961–976 (2018).

[131] Gerald B. Folland. *Real analysis: modern techniques and their applications*. Wiley, New York, 1999.

[132] Peter G. Casazza and Matthew Fickus. "Fourier transforms of finite chirps". In: *EURASIP Journal on Adv. in Signal Process.* vol. 2006, pp. 1–7 (2006).

[133] Petre Stoica and Randolph L. Moses. *Spectral analysis of signals*. Prentice Hall, Upper Saddle River, New Jersey, 2005.

[134] Hamid Krim and Mats Viberg. "Two decades of array signal processing research: the parametric approach". In: *IEEE Signal Process. Mag.* vol. 13, no. 4, pp. 67–94 (1996).

[135] Zai Yang et al. "Sparse methods for direction-of-arrival estimation". In: *Academic Press Library in Signal Processin*. Ed. by Rama Chellappa and Sergios Theodoridis. Vol. 7. Academic Press, 2018, pp. 509–581.

[136] S. Unnikrishna Pillai. *Array signal processing*. Springer Science & Business Media, 1989.

[137] Heinrich Meyr, Marc Moeneclaey, and Stefan A. Fechtel. *Digital communication receivers: synchronization, channel estimation, and signal processing*. Wiley, New York, 1998.

[138] Babak Hassibi and Bertrand M. Hochwald. "How much training is needed in multiple-antenna wireless links?" In: *IEEE Trans. Inf. Theory* vol. 49, no. 4, pp. 951–963 (2003).

[139] Merran Evans et al. *Statistical distributions*. Wiley, New Jersey, 2011.

[140] Emmanuel J. Candes and Justin Romberg. "Quantitative robust uncertainty principles and optimally sparse decompositions". In: *Found. Comput. Math.* vol. 6, no. 2, pp. 227–254 (2006).

[141] Marcelo S. Alencar and Valdemar C. da Rocha. *Communication systems*. Springer, Cham, Switzerland, 2020.

[142] Ralph Schmidt. "Multiple emitter location and signal parameter estimation". In: *IEEE Trans. on Antennas and Propagation* vol. 34, no. 3, pp. 276–280 (1986).

[143] Mati Wax and Ilan Ziskind. "On unique localization of multiple sources by passive sensor arrays". In: *IEEE Trans. on Acoust. Speech and Signal Process.* vol. 37, no. 7, pp. 996–1000 (1986).