TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Theoretische Informationstechnik

Analysis of Convolutions with Support Restrictions

Philipp Walk

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. Sandra Hirche
Prüfer der Dissertation:
1. Univ.-Prof. Dr.-Ing. Dr. rer. nat. Holger Boche
2. Univ.-Prof. Dr. rer. nat. Massimo Fornasier (schriftliche Beurteilung)
   Univ.-Prof. Dr. sc. techn. Gerhard Kramer (mündliche Prüfung)

Die Dissertation wurde am 31.10.2013 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 06.05.2014 angenommen.

Abstract

For several communication models, the dispersive part of a communication channel is described by a bilinear map $\mathcal{B}$ between the possible sets of input signals and channel parameters. The received channel output has then to be identified from the image set $\mathcal{B}(X, Y)$ of the input signal set X and the channel state (parameter) set Y. The main result in this dissertation characterize the compressibility of the image set with respect to an ambient dimension n. It can be shown for time-discrete signals, that a restricted norm multiplicativity property (RNMP) of $\mathcal{B}$ on the set of all s resp. f sparse signals $\Sigma_s^n \times \Sigma_f^n$ with bounds $\alpha, \beta > 0$, independent of n, is sufficient to establish the restricted isometry property (RIP) with exponentially high probability from $m = \mathcal{O}((s + f) \log n)$ random sub-Gaussian measurements. In practical use, this means, that whenever the channel parameter $\mathbf{y} \in Y$ is f−sparse and fixed, any s−sparse data signal $\mathbf{x} \in X$ can be stably reconstructed from $\mathcal{O}((2s + f) \log n)$ measurements. Thus, the number of degrees of freedom of each output grows only additively instead of multiplicatively with the input dimensions s and f. This is a substantial improvement in the output compressibility and suggests a substantially reduced rate in compressed sampling algorithms. The best known example of a bilinear map is the discrete convolution. If the input signals (sequences) having finite support length, the RNMP with bounds dependent only on the support lengths will be established. This result has several implications for blind system and signal detection, noncoherent communication of sporadic and short–message type user data and strategies for its compressive reception. Moreover, relating the RNMP to the phase retrieval problem, it can be shown that a stable reconstruction up to a global sign from $4n - 1$ symmetrized magnitude Fourier measurements of all signals in $\mathbb{C}^n$ is possible. For complex-valued signals where the first coefficient is real, a stability result holds also for $4n - 3$ symmetrized magnitude Fourier measurements. Here the symmetrization is a non-linear operation and generates a linear-phase filter having a real-valued spectrum. Hence, the stability result implies injectivity of the measurements up to a global sign.

In the last part of this work, we study semi-discrete convolutions and design a spectrally efficient time-limited pulse, which allows an overlapping pulse position modulation scheme for ultra-wideband communication. For this we investigate an orthogonalization method, which was developed in 1950 by Per-Olov Löwdin [Löw50; Löw70]. Our objective is to obtain a set of n orthogonal (Löwdin) pulses, which remain time-limited and spectrally efficient for UWB systems, from a set of n equidistant translates of a time-limited optimal spectral designed UWB pulse. We derive an approximate Löwdin orthogonalization (ALO) by using circulant approximations for the Gram matrix to obtain a practical filter implementation, which can be seen as a convolution with a finite sequence and a compact function. We show that the centered ALO and Löwdin pulses converge pointwise to the same Nyquist pulse as n tends to infinity. The

set of translates of the Nyquist pulse forms an orthonormal basis for the shift-invariant space generated by the initial spectral optimal pulse. The ALO transformation provides a closed-form approximation of the Löwdin transform, which can be implemented in an analog fashion without the need of analog to digital conversions. Furthermore, we investigate the interplay between the optimization and the orthogonalization procedure by using methods from the theory of shift-invariant spaces. Finally, a connection between the results in this thesis and wavelet and frame theory is developed.

Acknowledgments

Coming from a mathematical physics background I got the opportunity to investigate signal processing and communication theory in great detail. I am very grateful to my advisor Prof. Holger Boche for giving me the opportunity to work under a great scientific freedom in these fields. My gratitude goes to my colleague and friend Peter Jung for supporting me with an endless amount of time, inspiration and motivation. During my time at the Technical University Berlin our discussions developed to an intensive fruitful collaboration, which inspired and stimulated my scientific work. I would like to express my gratitude to Prof. Massimo Fornasier for serving as the second referee of this dissertation. The discussions in his seminars, his knowledge and expertise around Compressed Sensing helped me a lot to formulate my results more precisely. I thank Prof. Sandra Hirche for acting as the chairman of my thesis committee and Prof. Gerhard Kramer for serving as an additional objective referee. I would also like to thank my mentor Igor Bjelaković for his limitless patience in teaching me faithful scientific working.

My heartfelt gratitude goes to my family, my mother and my friends, who stayed around me during all these years and gave me so much support. Many thanks go to my colleagues Jochen, Sander, Sid, Janis, Moritz, Gisbert and Volker for their ears always open to my mathematical problems. Moreover, I want to thank David Gross and Götz Pfander for their support outside my University, expanding my horizon in many ways.

Contents

List of Figures

Notation

$\mathbb{K}$	Field of complex $\mathbb{C}$ or real numbers $\mathbb{R}$.
$a, b, c, \dots, z$	Field elements, integers or functions.
$\alpha, \beta, \dots, \epsilon$	Positive real numbers.
$e, \imath$	Euler and Imaginary number.
s, f	Natural numbers, exclusively used for the support length of sequences.
$\mathbf{a}, \mathbf{b}, \mathbf{c}, \dots$	Vectors $\mathbf{a} = (a_0, \dots, a_{n-1})^T$ or sequences $\mathbf{a} = (\dots, a_{-n}, \dots, a_n \dots)$ over $\mathbb{K}$.
$\mathbf{A}, \mathbf{B}, \mathbf{C}, \dots$	Matrices over $\mathbb{K}$.
$\mathbb{1}_n$	Identity matrix in $\mathbb{K}^{n \times n}$.
$\mathbf{S}_n$	Right shift matrix in $\mathbb{K}^{n \times n}$.
$\mathbf{\Gamma}_n$	Time-reversal matrix in $\mathbb{K}^{n \times n}$, see (1.59).
$\mathbf{F}_n$	Fourier matrix in $\mathbb{C}^{n \times n}$.
$\mathbf{\Phi}$	Subgaussian random matrix.
$\mathbf{a}^T, \mathbf{A}^T$	Transpose of vector or matrix.
$\mathbf{a}^*, \mathbf{A}^*$	Adjoint of a vector or matrix or involution in an algebra.
$\mathrm{A}, \mathrm{B}, \mathrm{C}, \dots$	Linear maps.
S	Frame Operator.
Z	Zak Transform.
L	Laplace Transform.
$\mathscr{H}, \dots, \mathscr{Z}$	Linear spaces over $\mathbb{K}$.
$\mathscr{H}$	Hilbert space.
$\mathscr{V}, \mathscr{S}$	Shift-invariant spaces.
$\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{S}$	Non-linear maps.
$\mathcal{S}$	Symmetrization map on $\mathbb{K}^n$, see (2.133).
$\mathcal{B}$	Bilinear map (multiplication) in the Hilbert space $\mathscr{H}$.
$\mathcal{B}_n, \mathcal{B}_{s,f}$	Bilinear map on $\mathbb{K}^n \times \mathbb{K}^n$ respectively on $\mathbb{K}^s \times \mathbb{K}^f$, see (1.20).
$\mathfrak{A}, \mathfrak{B}$	Algebras.
$\mathscr{N}(\mathrm{A})$	Nullset of the map A.
$[n]$	First n integers: $\{0, 1, \dots, n-1\}$.
$\lfloor a \rfloor, \lceil a \rceil, \oplus, \ominus$	Floor, ceil and additon modulo operation, see Definition 8.
$[a, b]$	All integers $c \in \mathbb{Z}$ or numbers $c \in \mathbb{R}$ with $a \leq c \leq b$.
$A, B, \dots, P$	Arbitrary set of integers.
P, Q	Generalized arithmetic progressions, see (2.67).
$T, U, V, \dots, Z$	Arbitrary sets.
$\mathcal{P}, \mathcal{Q}, \mathcal{R}$	Nets for compact sets.
$N_\epsilon(X)$	ϵ–Covering number of the convex precompact set X.
$\mathbf{x} \otimes \mathbf{y}$	Kronecker product between $\mathbf{x} \in \mathbb{K}^s$ and $\mathbf{y} \in \mathbb{K}^f$: $(\mathbf{x} \otimes \mathbf{y})_{is+j} = x_i y_j$.
$\mathscr{K}$	Hilbert Tensor space with ℓ^2-norm and inner product.

$\mathscr{K}_1$	Set of Kronecker products (rank−1 tensors) $\mathscr{K}_1 := \left\{\mathbf{x}\otimes\mathbf{y} \mid \mathbf{x}\in\mathbb{K}^s, \mathbf{y}\in\mathbb{K}^f\right\}$.
$\mathscr{M}$	Linear space of $s\times f$ matrices with Frobenius norm $\|\mathbf{A}\|_F := \sqrt{\mathrm{tr}(\mathbf{A}\mathbf{A}^*)}$.
$\mathscr{M}_1$	Set of rank−1 matrices in $\mathbb{K}^{s\times f}$.
$\odot$, $\underline{\odot}$	Point product respectively generalized point product, see (1.38).
$\circledast, \circledast$	Circular convolution respectively correlation $\mathbb{K}^n\times\mathbb{K}^n\to\mathbb{K}^n$ (1.51) and (1.52).
$*, \star$	Discrete and continuous convolution respectively correlation, see Definition 5.
$\mathrm{supp}(\mathbf{x})$	Support of the sequence respectively vector $\mathbf{x}$, see (1.2).
Σ_k^n	Set of vectors in $\mathbb{K}^n$ with support cardinality not larger than k, see (1.5).
ℓ_k^2	Set of the infinite sequences with support cardinality less or equal k.
$\langle\mathbf{x},\mathbf{y}\rangle$	Inner Product of $\mathbf{x}$ and $\mathbf{y}$ given by $\langle\mathbf{x},\mathbf{y}\rangle := \sum_i x_i\bar{y}_i$, see (1.4).
$\|\mathbf{x}\|$	Euclidean norm, norm of the vector $\mathbf{x}\in\mathbb{K}^n$ given by $\|\mathbf{x}\| := \sqrt{\langle\mathbf{x},\mathbf{x}\rangle}$.
$\ell^2, \ell^2_{[n]}$	Hilbert space of the square summable sequences (vectors) on $\mathbb{Z}$ respectively $[n]$.
L^2, L^2_T	Hilbert space of the square integrable functions $x:\mathbb{R}\to\mathbb{K}$ respectively $x:T\to\mathbb{K}$.
$\|\cdot\|_\infty$	Maximum norm for vectors, sequences or matrices.
$\|\cdot\|_p$	p-norm for vectors or sequences with $1\le p<\infty$.
$\|\mathbf{A}\|_F, \|\mathbf{A}\|_\infty$	Frobenius norm respectively Supremums (Maximum) norm of the matrix $\mathbf{A}$.
$\|a\|, \|A\|$	Absolute value of $a\in\mathbb{C}$, cardinality of $A\subset\mathbb{Z}$ and measure for $A\subset\mathbb{R}$.
$X^{a,b}$	Shell in X with inner radius a and outer radius b.
$X^1 = X^{0,1}$	Unit ball in X, i.e., all elements $\mathbf{x}\in X$ with $\|\mathbf{x}\|\le 1$.
δ_{ij}	Kronecker symbol, defined as 1 for all $i=j\in\mathbb{Z}$ and 0 else.
$\mathbf{e}_i$	Euclidean unit-vector in $\mathbb{K}^n$ with jth component δ_{ij}.
χ_I	Characteristic function, $\chi_I(i)=1$ if $i\in I$ and 0 elsewhere.
$\mathcal{N}(0,\sigma^2)$	Normal (Gaussian) distribution with mean 0 and variance σ^2
$\Re(a), \Im(a)$	Real and Imaginary part of $a\in\mathbb{C}$.
$\mathbf{1}_n$	The vector $\mathbf{1}=\mathbf{1}_n=(1,1,\dots,1)^T\in\mathbb{K}^n$.
log	Logarithmus to the basis 10.
ln	Natural Logarithmus to the basis e.
ϕ	Freiman homomorphism, see (2.63).
$\sigma(A)$	Doubling constant of the set $A\subset\mathbb{Z}$.
ω, ν	Frequency variable.
$\det(\mathbf{A})$	Determinant of the square matrix $\mathbf{A}$.
$\mathrm{co}(X), \mathrm{span}(X)$	Convex hull of the set X respectively linear span.

1 Introduction

Signals with support restrictions are fundamental ingredients in many signal processing scenarios as for example in image reconstruction, signal communication, signal reconstruction, sampling and many more. In fact, since any man-made operation is limited in time and space it is only possible to obtain deterministic knowledge about a signal on a finite scale, where finite is meant on a continuous or discrete time domain. Nowadays, a lot is known about linear systems with sparse (compressible) or compact (band-limited) signals. The best known examples are *linear time invariant systems* (LTI), represented by a convolution with another *fixed* signal, i.e., the channel. But in many communication scenarios, the channel is unknown or varying, and one is faced with an infinite union of LTI systems, which do not admit anymore a linear structure. The task is then to find universal sampling and coding strategies, which allow a reconstruction and encoding for any unknown but fixed channel state. Generalizing the convolution to an arbitrary multiplication (bilinear mapping), we end up by identifying multiplications of signal sets. The aim here, is to design universal linear measurements, allowing a stable reconstruction of the output signals with the smallest amount of measurements. To achieve this, support restrictions of the signals are necessary. Here, the main focus relies on an efficient signal recovery in sparse convolution systems by using methods from *Compressed Sensing* and *Geometrical Functional Analysis*. To guarantee recovery (deconvolution) for sparse signals, one has to provide *stability* or at least *injectivity*. Due to the bilinear nature of the convolution, injectivity and stability can only hold in a weaker sense, which we will express by the *restricted norm multiplicativity property* (RNMP).

Communication over a channel, e.g., a wireless network or an optical fibre, is usually realized by time-continuous signals with a compact or compressible support in the time domain. Modelling real-world analog systems (time-continuous models) requires heavy functional analysis, which in many applications is not mathematically tractable any more. Hence, in the last decades much effort was done to express the actions of analog signal models by efficient digital models. In the last part of this work we will consider linear continuous-time invariant signal models. An approximation of compactly supported time-continuous signals then leads to sparse sequences and the continuous convolution transform to a discrete convolution. Finally, to obtain finite dimensions, the discrete convolution is further approximated by the circular convolution, resulting in aliasing due to the boundary effects. This enables a pulse position modulation scheme for an almost orthogonal signaling which at the same time efficiently utilizes the power in a given frequency band.

Outline

This dissertation is composed of two main parts: The Löwdin orthogonalization of a sequence of compactly supported pulses given by semi-discrete convolutions in [WJ12a] allowing an efficient orthogonal PPM transmission and the invertibility of convolved sparse pulses in [WJ13a] allowing a recovery of the discrete convolution at sub-Nyquist sampling rates [WJ12b].

The first chapter will introduce the concept of bilinear maps (multiplications) and derive general properties of bilinear maps. Then, an investigation on point products is carried out in more detail in Section 1.3 and the convolution respectively correlation is investigated further in Section 1.3.1.

To understand compressive sampling on sparse multiplications we will first investigate its representation and provide conditions for unicity of multiplications in finite dimensions in Chapter 2. In Section 2.2 we will show the RNMP for discrete convolutions with support restrictions, i.e., the cardinality of the support is bounded. This RNMP is then be used to derive a new result for a stable reconstruction of complex-valued $n-$dimensional signals, with real first coefficient, from $4n - 3$ symmetrized magnitude Fourier measurements in Section 2.3. This relates the autocorrelation to the results derived in the previous Section 2.2.

In Chapter 3 the RNMP property, defined in Section 1.2.5, will be used to obtain low-dimensional stable embedding results in Section 3.1 for sparse multiplications. This gives insight into possible sub-Nyquist sampling on images of bilinear maps. Our framework can then be applied to a compressive signal recovery in Chapter 3.

In the last Chapter 4 an orthogonalization scheme for compactly supported time-continuous signals enabling a spectrally efficient signaling for Ultra-wideband systems will be developed. Here again, support restrictions will be used in a stable and efficient way.

1.1 Signals with Restricted Support

In this work, the focus will be on signals with finite energy (energy signals), i.e., elements of *Hilbert spaces* $\mathscr{H} = (\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathscr{H}}, \|\cdot\|_{\mathscr{H}})$ over the field[1] $\mathbb{K}$, where the *norm* (square-root of the energy) of $x \in \mathscr{H}$ is induced by the scalar product $\langle\cdot,\cdot\rangle_{\mathscr{H}}$ as

$$\|x\| = \|x\|_{\mathscr{H}} := \sqrt{\langle x, x\rangle_{\mathscr{H}}}, \tag{1.1}$$

where the subscript $\mathscr{H}$ in the norm will be omitted for Hilbert spaces. A second key property of a signal is given by its *support* which is defined for continuous functions x on $T \subset \mathbb{R}$ or for a sequence on $T \subset \mathbb{Z}$ by

$$\operatorname{supp}(x) := \overline{\{t \in T \mid x(t) \neq 0\}}. \tag{1.2}$$

Here the bar $\overline{(\cdot)}$ denotes for sets the *closure* and for numbers the *complex conjugate*. If x is not a continuous function, but a measurable function, the definition can be extended to the *essential support* by introducing a measure on $\mathbb{R}$, see e.g. [LL01, p.13]. In this work only the Lebesgue measure and the point-measure (Haar measure) will be considered. Therefore, define for all Lebesgue measurable functions on T, i.e., *time-continuous signals* denoted by normal letters, the *scalar product* as

$$\langle x, y\rangle_{L^2(T)} := \int_{t\in T} x(t)\overline{y(t)}dt. \tag{1.3}$$

The set of equivalence classes of *time-continuous signals* on T with finite energy is denoted by $\mathscr{H} = L^2(T)$. If the signal is defined on $I \subset \mathbb{Z}$, then the point-measure defines a scalar product for sequences in (1.3), i.e., for *time-discrete signals*, denoted by bold letters with coefficients $(\mathbf{x})_k = x(k) = x_k$, as

$$\langle \mathbf{x}, \mathbf{y}\rangle_{\ell^2_I} := \sum_{k\in I} x(k)\overline{y(k)}. \tag{1.4}$$

The space of *square-summable sequences* on I is the Hilbert space $\mathscr{H} = \ell^2_I$. For $I = \mathbb{Z}$ we write $\mathscr{H} = \ell^2$ and for $I = [n] := \{0, 1, \ldots, n-1\}$ we write $\mathscr{H} = \ell^2_{[n]}$, which is an $n-$dimensional Hilbert space over $\mathbb{K}$. Similar, for $T = \mathbb{R}$ we write $L^2 = L^2_{\mathbb{R}}$ and for compact $|T| < \infty$ we refer by $L^2_T := L^2(T)$ to the signals with support contained in T having finite energy[2]. Furthermore, the set of vectors in $\mathbb{K}^n$ with support cardinality less or equal $k \leq n$ is given by

$$\Sigma^n_k := \{\mathbf{x} \in \mathbb{K}^n \mid |\operatorname{supp}(\mathbf{x})| \leq k\} \tag{1.5}$$

and is called the set of $k-$*sparse* signals in $\mathbb{K}^n$. As a generalization we will denote by

$$\ell^2_k := \left\{\mathbf{x} \in \ell^2 \;\middle|\; |\operatorname{supp}(\mathbf{x})| \leq k\right\} = \left\{\mathbf{x} \in \ell^2_I \;\middle|\; I \subset \mathbb{Z}, |I| \leq k\right\} \tag{1.6}$$

[1] We will formulate all results for real and complex-valued signals $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ as long as possible. Only if necessary we will restrict ourselves to real-valued signals.

[2] Although, in most literature the p-norm index is used in the lower position, we will always refer by the lower index to support restrictions, which are dominant in this thesis.

the set of sequences with support cardinality not larger than $k \in \mathbb{N}$. Finally let us introduce the general p*−norm* for $1 \le p < \infty$ as

$$\|x\|_p = \left(\int_{\mathbb{R}} |x(t)|^p dt\right)^{1/p} \quad , \quad \|\mathbf{x}\|_p = \left(\sum_{k\in\mathbb{Z}} |x_k|^p\right)^{1/p} \tag{1.7}$$

and the *supremum norm*

$$\|x\|_\infty = \operatorname*{ess\,sup}_{t\in\mathbb{R}} |x(t)| \quad , \quad \|\mathbf{x}\|_\infty = \sup_{k\in\mathbb{Z}} |x_k|, \tag{1.8}$$

where $\operatorname{ess\,sup}$ is the essential supremum up to sets of measure zero. The closure of the set of all sequences $\mathbf{x}$ respectively measurable functions x having finite norms in (1.7) resp. (1.8) defines the complete normed space (Banach space) ℓ^p respectively L^p. Moreover, for matrices $\mathbf{X}$ we define the ∞−norm by using the multi-index $\mathbf{k} \in \mathbb{Z}^2$ in the supremum. Whenever it is clear from the context, the index in the norm and the scalar product will be omitted, see e.g. [You80; Chr03].

For any subset X of a normed space $\mathscr{X}$ and for any subset $I \subset \mathbb{R}$ the elements $x \in X$ with support in I are denoted by

$$X_I := \{x \in X \mid \operatorname{supp}(x) \subset I\} \tag{1.9}$$

and for any positive numbers $a \le b$ the shell with inner-radius a and outer-radius b is written as

$$X^{a,b} := \{x \in X \mid a \le \|x\|_{\mathscr{X}} \le b\}. \tag{1.10}$$

Hence, for $a = b = 1$ we obtain the sphere $X^{1,1}$ and for $a = 0$ the ball X^b with radius b in X.

1.2 Multiplication of Signals

Multiplication of signals in the engineering community is usually understood as a point-wise multiplication in the Fourier domain by using the well known PARSEVAL relation, see e.g. [OSB99]. In this work, a more general concept of multiplication will be introduced, given by bilinear maps. As a special bilinear map, the point-product in various bases will be investigated.

1.2.1 Bilinear maps

Let $\mathscr{X}, \mathscr{Y}$ and $\mathscr{Z}$ be arbitrary linear spaces (signal spaces) and consider a map

$$\mathcal{B}: \mathscr{X} \times \mathscr{Y} \to \mathscr{Z}. \tag{1.11}$$

The mathematically most tractable non-linear "operation" between the signal space $\mathscr{X}$ and $\mathscr{Y}$ is given by a *bilinear map*, see e.g. [Gre67]:

Definition 1 (Bilinear map)**.** *Let $\mathscr{X}, \mathscr{Y}$ and $\mathscr{Z}$ be linear spaces. Then a map $\mathcal{B}\colon \mathscr{X} \times \mathscr{Y} \to \mathscr{Z}$ is called a* bilinear map *if for every fixed $x \in \mathscr{X}$ and $y \in \mathscr{Y}$ the maps*

$$\begin{aligned} \mathcal{B}(x,\cdot)&\colon \mathscr{Y} \to \mathscr{Z} \\ \mathcal{B}(\cdot,y)&\colon \mathscr{X} \to \mathscr{Z} \end{aligned} \tag{1.12}$$

are linear. For $s, f, n \in \mathbb{N}$ we denote a bilinear map $\mathcal{B} : \mathbb{K}^s \times \mathbb{K}^f \to \mathbb{K}^n$ by $\mathcal{B}_n^{s,f}$ and if $s = f = n$ we write $\mathcal{B}_n$.

Hence, for each fixed element (state) x respectively y the maps in (1.12) define a linear system. Therefore, bilinear maps can be used to describe a linear system with infinitely many states. Let us emphasize at this point that bilinear maps can map different linear spaces $\mathscr{X} \neq \mathscr{Y}$ into another linear space $\mathscr{Z}$, where the image $Z = \mathcal{B}(\mathscr{X}, \mathscr{Y})$ is not necessarily a linear subset of $\mathscr{Z}$. Exactly this non linear mapping of linear structures can increase the complexity of the output signal set Z dramatically. Therefore, the challenge taken on this thesis is to characterize the complexity of the image and to find conditions for $\mathcal{B}, \mathscr{X}, \mathscr{Y}$ to obtain a low-dimensional embedding of the image.

By the definition of bilinear maps it is clear that the null set $\mathscr{N}(\mathcal{B})$ contains at least $(0, \mathscr{Y}) \cup (\mathscr{X}, 0)$. Hence, the linear maps in (1.12) are only invertible if at least $x \neq 0$ respectively $y \neq 0$. Before we start to characterize invertibility in this section, we will study the geometric and algebraic structure of bilinear maps in more detail.

1.2.2 Geometric Structure

A direct consequence of (1.12) is the *homogeneity* and *positive (absolute) homogeneity*:

$$\lambda\mathcal{B}(x,y) = \mathcal{B}(\lambda x, y) = \mathcal{B}(x, \lambda y) \quad \text{for} \quad \lambda \in \mathbb{K} \tag{1.13}$$

$$\lambda\mathcal{B}(x,y) = \mathcal{B}(\lambda x, y) = \mathcal{B}(x, \lambda y) \quad \text{for} \quad \lambda \geq 0. \tag{1.14}$$

These properties define a *linear cone* respectively (positive) *cone*, see e.g. [Lue69].

Definition 2 (Cone)**.** *Let $\mathscr{X}$ be a linear space. A set $X \subset \mathscr{X}$ is a* linear cone*, if for every $x \in X$ it holds:*

$$\lambda x \in X \quad \textit{for all} \quad \lambda \in \mathbb{K} \tag{1.15}$$

and a cone*, if it holds:*

$$\lambda x \in X \quad \textit{for all} \quad \lambda \geq 0. \tag{1.16}$$

Remark. Obviously, every linear cone is a cone, but not every cone is a linear cone. Moreover, our definition of a cone includes positive cones, negative cones, double cones and convex cones.

Since cones are not necessarily convex or linear, we have to define their convex hull and linear hull, see for example [Web94].

Definition 3 (Convex set)**.** *Let $\mathscr{X}$ be a linear space. A set $X \subset \mathscr{X}$ is* convex*, if and only if for all $x, y \in X$ it holds*

$$\forall \lambda \in (0,1)\colon \lambda x + (1-\lambda) y \in X. \tag{1.17}$$

Moreover, we define for any arbitrary subset $X \subset \mathbb{R}^n$ the convex hull *by*

$$\operatorname{co}(X) := \{ \lambda x_1 + (1-\lambda) x_2 \mid x_1, x_2 \in X, \lambda \in (0,1) \} . \tag{1.18}$$

*The convex set X is $k-$*dimensional*, if the* linear hull (span) *of X*

$$\operatorname{span}(X) := \left\{ \sum_{i=0}^{n} \lambda_i x_i \;\middle|\; n \in \mathbb{N}, x_i \in X, \lambda_i \in \mathbb{K} \right\} \tag{1.19}$$

has dimension k.

Hence, we have with (1.13) the following important property for bilinear maps.

Lemma 1 (Linear Cone Invariance)**.** *Let $\mathcal{B}$ be a bilinear map and $X \subset \mathscr{X}$ or $Y \subset \mathscr{Y}$ linear cones, then the image $\mathcal{B}(X,Y)$ is again a linear cone in $\mathscr{Z}$.*

Remark. Note that even if X, Y are convex, then $\mathcal{B}(X,Y)$ is not necessarily convex.

1.2.3 Algebraic Structure: Multiplication and Banach Algebra

In the special case where $\mathscr{X} = \mathscr{Y} = \mathscr{Z}$ are finite dimensional, every bilinear map

$$\mathcal{B} = \mathcal{B}_{\mathscr{X},\mathscr{X}}\colon \mathscr{X} \times \mathscr{X} \to \mathscr{X} \tag{1.20}$$

defines a (product, binary operation) *multiplication* in $\mathscr{X}$ and hence an *algebra* $\mathfrak{B} = (\mathscr{X}, \mathcal{B})$. If $\mathscr{X}$ is a normed space, for example a finite dimensional Banach space, with norm $\|\cdot\|_{\mathscr{X}}$ and there exists $\beta > 0$, such that

$$\|\mathcal{B}(\mathbf{x},\mathbf{y})\|_{\mathscr{X}} \le \beta \|\mathbf{x}\|_{\mathscr{X}} \|\mathbf{y}\|_{\mathscr{X}} \quad \text{for} \quad \mathbf{x},\mathbf{y} \in \mathscr{X}, \tag{1.21}$$

then $\mathfrak{B}$ is isomorphic to a *normed algebra*, see for example [BD73; Pal94]. This is also valid for infinite dimensional normed spaces. Hence we can define:

Definition 4 (Normed Algebra)**.** *Let $\mathscr{X}$ be a normed space with Norm $\|\cdot\|_{\mathscr{X}}$, then $\mathfrak{B} = (\mathscr{X}, \mathcal{B})$ is a* normed algebra*, if for the bilinear map $\mathcal{B}$*

$$\|\mathcal{B}(\mathbf{x},\mathbf{y})\|_{\mathscr{X}} \le \|\mathbf{x}\|_{\mathscr{X}} \|\mathbf{y}\|_{\mathscr{X}} \quad \textit{for} \quad \mathbf{x},\mathbf{y} \in \mathscr{X} \tag{1.22}$$

holds. Then $\|\cdot\|_{\mathscr{X}}$ is called an algebra-norm *and the bilinear map $\mathcal{B}$ is called a product or multiplication in $\mathscr{X}$.*

Indeed if the bilinear map fulfills (1.21), then it is also a continuous (bounded) map, see for example [BD73, Prop.4], and we can always rescale the norm $\|\cdot\|$ to a $\|\cdot\|_0$, such that (1.22) holds [Kan08, Prop.1.1.1]. The property (1.22) is known as *sub-multiplicativity of the norm*, see for example [Pal94; Kan08]. Since $\mathscr{H} = \ell^2_{[n]}$ is with the 2−norm $\|\cdot\|$ a normed space and since every bilinear map on a finite dimensional normed space $\mathscr{X}$ is bounded (continuous), we have a normed algebra. For infinite dimensional normed spaces, the most prominent multiplication is the *convolution* and the *correlation* (sesquilinear map for $\mathbb{K} = \mathbb{C}$), see for example [OSB99]. Here we write for sequences $x_k := x(k)$.

Definition 5 (Convolution and Correlation)**.** *For $p \geq 1$ the* convolution *for time-discrete signals $\mathbf{x} \in \ell^p, \mathbf{y} \in \ell^1$ and time-continuous signals $x \in L^p, y \in L^1$ is defined by*

$$(\mathbf{x} * \mathbf{y})(k) = \sum_{l \in \mathbb{Z}} x_l y_{k-l} \quad \textit{for all} \quad k \in \mathbb{Z}, \tag{1.23}$$

$$(x * y)(t') = \int_{\mathbb{R}} x(t) y(t' - t) dt \quad \textit{for a.e.} \quad t' \in \mathbb{R}. \tag{1.24}$$

The correlation *is defined as*

$$(\mathbf{x} \star \mathbf{y})(k) = \sum_{l \in \mathbb{Z}} x_l \overline{y_{k+l}} \quad \textit{for all} \quad k \in \mathbb{Z}, \tag{1.25}$$

$$(x \star y)(t') = \int_{\mathbb{R}} x(t) \overline{y(t' + t)} dt \quad \textit{for a.e.} \quad t' \in \mathbb{R}. \tag{1.26}$$

Indeed, the convolutions are well defined, since we have in the discrete case (discrete local compact group, see for example [Rud62, Thm.1.1.6] and [Gar07, Thm.9.4.1] for $1 \leq p \leq \infty$

$$\|\mathbf{x} * \mathbf{y}\|_p \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_1 \quad \text{for all} \quad \mathbf{x} \in \ell^p, \mathbf{y} \in \ell^1 \tag{1.27}$$

and in the time-continuous case it follows from the *Minkowski* integral inequalities, see for example [SW71, Thm.1.3.]

$$\|x * y\|_p \leq \|x\|_p \|y\|_1 \quad \text{for all} \quad x \in L^p, y \in L^1. \tag{1.28}$$

The same holds for the correlation, since ℓ^p and L^p are closed against complex conjugation and time-reversal. The above convolution inequalities on local compact abelian groups, here $\mathbb{R}$ or $\mathbb{Z}$, are known as special cases from the well known YOUNG inequalities [You1912,You1913] for $1 < q, p = r < \infty$. Hence, for infinite dimensional spaces we have only for $p = 1$ a normed convolution algebra (Banach algebra), see for example [LL01],[Rud62, Thm.1.1.7], which is in strong contrast to finite dimensions, see next Section 1.3.1.

1.2.4 Inverse Elements

One of the main tasks in signal processing is the reconstruction of a signal from its measurement (observation). For LTI systems (convolution systems) with $\mathscr{X} = \mathscr{Y} = \mathscr{Z}$ this is known as *deconvolution*, i.e., by knowing the channel state y and observing the output $z = x * y$, one has to determine the corresponding transmitted signal x, such that

$$\mathcal{B}(x,y) = z. \tag{1.29}$$

In communication theory this is also known as *equalization*.

In terms of algebra a unique deconvolution corresponds to bijectivity of the left multiplication $\mathrm{L}_x := \mathcal{B}(x,\cdot)$ and right multiplication $\mathrm{R}_y := \mathcal{B}(\cdot,y)$ as defined in (1.12). If bijectivity of L_x and R_y holds for every $x,y \in \mathscr{X}$, then $\mathfrak{B}$ is called a *divison algebra*, see for example [Pal04]. Note that in commutative algebras bijectivity of L_x for every $x \in \mathscr{X}$ implies bijectivity of R_y for every $y \in \mathscr{Y}$ and vice versa. For finite-dimensional algebras with $\mathbb{K} = \mathbb{R}$ it is known, that if $\mathfrak{B}$ is a division algebra, it has dimension either $1, 2, 4$ or 8, i.e. is isomorphic to one of the following algebras: $\mathbb{R}, \mathbb{C}, \mathbb{H}$ (Quaternions) or $\mathbb{O}$ (Octonions). Moreover, if the algebra is normed (Banach), then the norm is *multiplicative* if

$$\|\mathcal{B}(\mathbf{x},\mathbf{y})\|_{\mathscr{X}} = \|\mathbf{x}\|_{\mathscr{X}} \|\mathbf{y}\|_{\mathscr{X}} \quad \text{for} \quad \mathbf{x},\mathbf{y} \in \mathscr{X} \tag{1.30}$$

and $\mathfrak{B}$ is called an *absolute-valued algebra*, which is also a division algebra, see e.g. [Pal04]. This fundamental result from abstract algebra, shows that for dimension $n > 8$, deconvolution can not be possible for every channel $\mathbf{y}$! Hence, for a unique deconvolution (by knowledge of one input), one either needs further constraints to determine the transmitted signal $\mathbf{x}$ or one needs to classify (topological) sets $X, Y \subset \mathscr{X}$, for which L_x respectively R_y are bijective (invertible).

We will consider in the first part of this thesis, Chapter 2 and Chapter 3, the circular convolution $\mathcal{B} = \circledast$ on $\mathbb{K}^n$ or more generally the *Fourier analysis on finite groups*, which is isomorphic to a *finite-dimensional commutative unital Banach algebra* on $\ell^2_{[n]}$ as shown in Lemma 3. For infinite dimensional convolution algebras, i.e., infinite groups, things become more difficult [SS03].

1.2.5 Restricted Norm Multiplicativity

We will now formalize this idea and introduce a property for the triple $(\mathcal{B}, X, Y)$ such that L_x is invertible on Y for every $x \in X$ and R_y is invertible on X for every $y \in Y$, as long as Y respectively X are linear spaces.

Definition 6 (RNMP on $X \times Y$)**.** *Let be $\mathscr{X}, \mathscr{Y}, \mathscr{Z}$ normed spaces, X, Y be cones in $\mathscr{X}$ respectively $\mathscr{Y}$ and $\alpha, \beta > 0$. Then a bilinear map $\mathcal{B}\colon \mathscr{X} \times \mathscr{Y} \to \mathscr{Z}$ has the* restricted norm multiplicativity property (RNMP) *on $X \times Y$ with bounds $0 < \alpha \le \beta < \infty$ if*

$$\alpha \|x\|_{\mathscr{X}} \|y\|_{\mathscr{Y}} \le \|\mathcal{B}(x,y)\|_{\mathscr{Z}} \le \beta \|x\|_{\mathscr{X}} \|y\|_{\mathscr{Y}} \quad \textit{for all} \quad x \in X, y \in Y \tag{1.31}$$

holds.

Remark. If $\alpha = \beta = 1$, then the norm is called *multiplicative* and we *locally* have an absolute-valued property as in an absolute-valued algebra. Since this, as mentioned earlier, is not the case in general on the whole space, but rather on restricted subsets X, Y, the authors called this in [WJ12b] the *restricted norm multiplicativity property* (RNMP) of $\mathcal{B}$ on $X \times Y$. The definition by cones is in fact no restriction, since for any sets X, Y we can extend the RNMP due to the bilinearity of $\mathcal{B}$ to cones, by defining $X' := \{\lambda X \mid \lambda > 0\}$ without changing the bounds α, β. The RNMP refers in a certain form to the complement of the set of *topological divisor of zeros* as used in algebra, see for example [BD73, Def.12, §2], with the distinction that the Cartesian product $X \times Y$ is not required to have any linear or convex structure.

The following questions pop up immediately:

1. Which multiplications $\mathcal{B}$ have the RNMP on some linear subspaces $X, Y \subset \mathscr{X}$?
2. How can we find the "largest" RNMP cones X and Y for fixed $\mathcal{B}$?
3. If $\mathcal{B}$ has the RNMP on $X, Y \subset \mathscr{X}$, does this imply $\mathcal{B}(X, Y) = \mathcal{B}(\operatorname{span}(X), \operatorname{span}(Y))$?

The last can infact generally be answered in the negative. Even if we select the largest possible RNMP pair (X, Y), there could exist pairs $(\mathbf{x}, \mathbf{y}) \notin X \times Y$ with $\mathcal{B}(\mathbf{x}, \mathbf{y}) \in \mathcal{B}(\mathscr{X}, \mathscr{Y})$ but not in $Z = \mathcal{B}(X, Y)$. In other words, the pre-image set $O \subset X \times Y$ does not have to be a Cartesian product of two sets. On the other hand, we can project O on the Cartesian product $P_X O \times P_Y O$ by defining the projection operators $P_X : X \times Y \to X$ respectively P_Y. Since the Cartesian product is not a normed space we define $O^{1,1} := \{\mathbf{o} \in O \mid \|P_X \mathbf{o}\| = 1 = \|P_Y \mathbf{o}\|\}$. Therefore, we can define a *general RNMP* as in [WJ12b], which maximizes the possible range of cones X, Y.

Definition 7 (General RNMP on $X \times Y$)**.** *Let be $\mathscr{X}, \mathscr{Y}, \mathscr{Z}$ normed spaces, X, Y cones in $\mathscr{X}$ respecitvely $\mathscr{Y}$ and $\alpha, \beta > 0$. Then a bilinear map $\mathcal{B}\colon \mathscr{X} \times \mathscr{Y} \to \mathscr{Z}$ has the* general RNMP *on $X \times Y$ with bounds $0 < \alpha \le \beta < \infty$ if*

$$0 < \alpha := \sup_{\substack{O \subset X \times Y \\ \mathcal{B}(O) = \mathcal{B}(X,Y) \setminus \{\mathbf{0}\}}} \inf_{(x,y) \in O^{1,1}} \|\mathcal{B}(x, y)\|_{\mathscr{Z}} \tag{1.32}$$

and

$$\beta := \sup_{(x \times y) \in X^{1,1} \times Y^{1,1}} \|\mathcal{B}(x, y)\|_{\mathscr{Z}} \tag{1.33}$$

holds.

Essentially, the implicit definition by such a set O would remove the redundancy in representing $\mathcal{B}(X, Y)$, i.e., it would remove unnecessary direction pairs in $X \times Y$. But the exact determination of the set O is maybe impossible and would depend on $\mathcal{B}$ as well as on the cones X and Y.

In finite dimensions the bilinear map is always bounded, i.e., a bound $\beta < \infty$ exists. However, if the bilinear map is bounded, the infimum in (1.32) is attained for some $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ due to the compactness of $X^{1,1} \times Y^{1,1}$. Then, whenever X, Y are convex (uncountably many elements) an element on the nullset can be arbitrarily closely approximated by a sequence $\{(\mathbf{x}_n, \mathbf{y}_n)\} \subset X \times Y$ and a bound $\alpha > 0$ would never exist. Let us investigate therefore an example for cones X, Y with finitely many directions, in order to illustrate and justify this general definition.

Example. Assume $X = \{\lambda \mathbf{x} \mid \lambda \in \mathbb{K}, \mathbf{x} \in \{\mathbf{x}_1, \mathbf{x}_2\}\} \subset \mathbb{R}^2$ and $Y = \{\lambda \mathbf{y} \mid \lambda \in \mathbb{K}, \mathbf{y} \in \{\mathbf{y}_1, \mathbf{y}_2\}\} \subset \mathbb{R}^2$. Hence X, Y are linear cones in $\mathbb{R}^2$ with exactly two directions. Let us define the bilinear map $\mathcal{B} : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^4$ by

$$\mathcal{B}(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} x_0 & 0 & x_1 & 0 \\ 0 & x_0 & 0 & x_1 \\ x_2 & 0 & x_0 & 0 \\ 0 & x_1 & 0 & x_0 \end{pmatrix} \begin{pmatrix} y_0 \\ 0 \\ y_1 \\ 0 \end{pmatrix} \in \mathbb{R}^4. \tag{1.34}$$

Clearly, $\mathcal{B}$ is left linear, since $\mathbf{x}$ comprise a matrix (linear map) and right linear, since we have for fixed $\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y} \in \mathbb{R}^2$

$$\mathcal{B}(\mathbf{x} + \tilde{\mathbf{x}}, \mathbf{y}) = \begin{pmatrix} x_0 + \tilde{x}_0 & 0 & x_1 + \tilde{x}_1 & 0 \\ 0 & x_0 + \tilde{x}_0 & 0 & x_1 + \tilde{x}_1 \\ x_1 + \tilde{x}_1 & 0 & x_0 + \tilde{x}_0 & 0 \\ 0 & x_1 + \tilde{x}_1 & 0 & x_0 + \tilde{x}_0 \end{pmatrix} \begin{pmatrix} y_0 \\ 0 \\ y_1 \\ 0 \end{pmatrix} \tag{1.35}$$

$$= \begin{pmatrix} x_0 & 0 & x_1 & 0 \\ 0 & x_0 & 0 & x_1 \\ x_1 & 0 & x_0 & 0 \\ 0 & x_1 & 0 & x_0 \end{pmatrix} \mathbf{y} + \begin{pmatrix} \tilde{x}_0 & 0 & \tilde{x}_1 & 0 \\ 0 & \tilde{x}_0 & 0 & \tilde{x}_1 \\ \tilde{x}_1 & 0 & \tilde{x}_0 & 0 \\ 0 & \tilde{x}_1 & 0 & \tilde{x}_0 \end{pmatrix} \mathbf{y} = \mathcal{B}(\mathbf{x}, \mathbf{y}) + \mathcal{B}(\tilde{\mathbf{x}}, \mathbf{y}). \tag{1.36}$$

Let $\mathbf{x}_1 = (1,1)^T, \mathbf{x}_2 = (-2,1)^T, \mathbf{y}_1 = (0.5, -0.5)^T, \mathbf{y}_2 = (2,1)^T$, then we get the following image points

$$\mathcal{B}(\mathbf{x}_1, \mathbf{y}_1) = \mathbf{0}, \mathbf{z}_1 = \mathcal{B}(\mathbf{x}_1, \mathbf{y}_2) = \begin{pmatrix} 3 \\ 0 \\ 3 \\ 0 \end{pmatrix}, \mathbf{z}_2 = \mathcal{B}(\mathbf{x}_2, \mathbf{y}_1) = \begin{pmatrix} -\frac{3}{2} \\ 0 \\ \frac{3}{2} \\ 0 \end{pmatrix}, \mathbf{z}_3 = \mathcal{B}(\mathbf{x}_2, \mathbf{y}_2) = \begin{pmatrix} -3 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{1.37}$$

where $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3$ are three different directions (distinct elements in the Grassmanian $G_{1,4}$), see Fig. 1.1. Hence, we need to represent three directions by X and Y, but they cannot be written as a Cartesian product of smaller cones in X and Y.

As mentioned before, the non-Cartesian set $O \subset X \times Y$ can be always projected to O_X and O_Y. Hence O_X, O_Y are the minimal cones to represent the image, i.e. $\mathcal{B}(O_X, O_Y) = \mathcal{B}(X, Y)$. For our example we have $O = \{(\mathbf{x}_1, \mathbf{y}_2), (\mathbf{x}_2, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2)\}$, see Fig. 1.1. We will use this construction later in the proof of Lemma 5.

Moreover, such a set O in general lacks linear or convex properties. For practical use it is often easier to derive the RNMP for convex cones or linear subspaces X, Y.

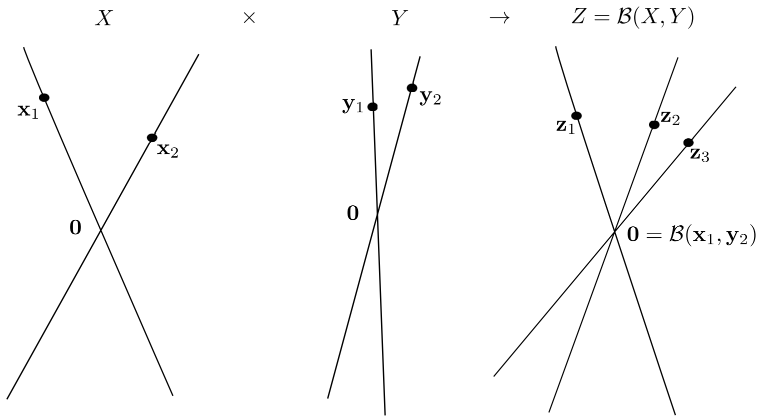


Figure 1.1: Example for the RNMP on a non-Cartesian product.

Therefore, the main task in this thesis is to find a fixed multiplication and hence algebra, the largest possible family of invertible linear systems. Unfortunately, since invertibility of linear maps is usually defined on linear sets, non-linear restrictions seem to be pathological. On the other hand, from a geometric point of view, the RNMP on convex cones can be used to obtain sharp covering results for non-linear images, as derived in Chapter 3.

1.3 Finite Point Products

We will start with the simplest non-trivial multiplication, the *point product* in finite dimension:

$$\begin{aligned} \odot\colon \mathbb{K}^n \times \mathbb{K}^n &\to \mathbb{K}^n \\ (\mathbf{x}, \mathbf{y}) &\mapsto \mathbf{x} \odot \mathbf{y} = (x_1 y_1, \dots, x_n y_n)^T \end{aligned} \tag{1.38}$$

which defines a *commutative multiplication* in $\mathbb{K}^n$. Moreover, the vector $\mathbf{1} = \mathbf{1}_n := (1, 1, \dots, 1)^T$ is the unit element in $\mathfrak{B}_n^{\odot} = (\mathbb{K}^n, \|\cdot\|, \odot)$, defining a *commutative normed algebra with unit*. Changing to different bases $\{\mathbf{u}_k\}_{k=0}^{n-1}$ and $\{\mathbf{w}_k\}_{k=0}^{n-1}$ we loose commutativity for the unitary basis transformations $\mathbf{U}^* = (\mathbf{u}_0 \dots \mathbf{u}_{n-1})$ and $\mathbf{W}^* = (\mathbf{w}_0 \dots \mathbf{w}_{n-1})$ with

$$\mathbf{x} \underset{\sim}{\odot} \mathbf{y} = \mathbf{U}\mathbf{x} \odot \mathbf{W}\mathbf{y}, \tag{1.39}$$

since $\mathbf{U}\mathbf{x} \neq \mathbf{W}\mathbf{x}$ for some $\mathbf{x} \in \mathbb{K}^n$. On the other hand, for $\mathbf{U} = \mathbf{W}$ commutativity is preserved. Hence, we define the *(normalized) generalized commutative point product* by unitary $(n \times n)$−matrices $\mathbf{U}$ and $\mathbf{V}$ as

$$\mathbf{x} \underline{\odot} \mathbf{y} := \frac{1}{\sqrt{n}\,\|\mathbf{U}\|_\infty} \mathbf{V}(\mathbf{U}\mathbf{x} \odot \mathbf{U}\mathbf{y}), \tag{1.40}$$

with the maximum[3] norm $\|\mathbf{U}\|_\infty := \max_{i,j} |u_{ij}|$. Commutativity holds by $\mathbf{U}\mathbf{x} \odot \mathbf{U}\mathbf{y} = \mathbf{U}\mathbf{y} \odot \mathbf{U}\mathbf{x}$ (commutativity of the point product). Before we prove general ℓ^2−norm inequalities, we

[3]The supremum can be substituted on finite sets by the maximum.

introduce the counterpart of the identity matrix, the *(discrete) Fourier matrix* $\mathbf{F} := \mathbf{F}_n \in \mathbb{C}^{n\times n}$ with elements $(\mathbf{F})_{lk} = \frac{1}{\sqrt{n}} e^{-2\pi\imath\frac{lk}{n}}$ for $l,k \in [n]$.

Lemma 2 (RNMP for Sparse Point Products)**.** *Let s, f, n be natural numbers with $s \le f \le n$, and $\underline{\odot}$ as in* (1.40) *for some unitary matrices* $\mathbf{U}$ *and* $\mathbf{V}$*, then*

$$\frac{1}{n\|\mathbf{U}\|_\infty}\|\mathbf{x}\|\|\mathbf{x}\| \le \|\mathbf{x}\,\underline{\odot}\,\mathbf{x}\| \quad \textit{for all} \quad \mathbf{x}\in\Sigma_s, \tag{1.41}$$

$$\sqrt{\frac{s}{n}}\|\mathbf{x}\|\|\mathbf{y}\| \ge \|\mathbf{x}\,\underline{\odot}\,\mathbf{y}\| \quad \textit{for all} \quad \mathbf{x}\in\Sigma_s, \mathbf{y}\in\Sigma_f. \tag{1.42}$$

Remark. The lower bounds are sharp for some $\mathbf{y} = \mathbf{x} \in \Sigma_s$ and $\mathbf{U} = \mathbf{F}$ respectively $\mathbf{U} = \mathbb{1}_n$. The upper bound is attained for $\mathbf{U} = \mathbf{F}$ and $\mathbf{U} = \mathbb{1}_n$ with some $\operatorname{supp}(\mathbf{x}) = I, \operatorname{supp}(\mathbf{y}) = J$ only in the special cases with $|I| = 1, |J| \le n$ and $|I| \le n, |J| = n$.

Proof. Using (1.40) and unitarity of $\mathbf{V}$ we get for the ℓ^2-norm $\|\cdot\|$ of the product

$$\|\mathbf{x}\,\underline{\odot}\,\mathbf{y}\|^2 = \frac{1}{n\|\mathbf{U}\|_\infty^2}\|\mathbf{Ux}\odot\mathbf{Uy}\|^2 = \frac{1}{n\|\mathbf{U}\|_\infty^2}\sum_{i=0}^{n-1}|(\mathbf{Ux})_i|^2|(\mathbf{Uy})_i|^2. \tag{1.43}$$

Using the HÖLDER inequality, see for example [Web94, Cor. 5.2.6], with $p = 1$ and $q = \infty$ we get $|(\mathbf{Ux})_i|^2 = |\langle\mathbf{x},\mathbf{u}_i\rangle|^2 \le \|\mathbf{x}\|_1^2\|\mathbf{u}_i\|_\infty^2$ for each $i\in[n]$. This yields the upper bound

$$\|\mathbf{x}\,\underline{\odot}\,\mathbf{y}\|^2 \le \frac{\|\mathbf{x}\|_1^2}{n\|\mathbf{U}\|_\infty^2}\sum_i\|\mathbf{u}_i\|_\infty^2|(\mathbf{Uy})_i|^2 \tag{1.44}$$

$$\le \frac{1}{n}\|\mathbf{x}\|_1^2\|\mathbf{y}\|^2. \tag{1.45}$$

The upper bound (1.45) follows from the sparsity of $\mathbf{x}$ by

$$\|\mathbf{x}\,\underline{\odot}\,\mathbf{y}\|^2 \le \frac{s}{n}\|\mathbf{x}\|^2\|\mathbf{y}\|^2, \tag{1.46}$$

where equality is obtained for $\mathbf{U} = \mathbf{F}$ if $\mathbf{x}\in\Sigma_1$ and $\operatorname{supp}(\mathbf{y}) = [n]$ or if $|\operatorname{supp}(\mathbf{x})| = f$. To see this, we first note, that (1.44) becomes an equality for $\mathbf{U} = \mathbf{F}$. The last inequality is sharp only for those $\mathbf{x}$ which are constant on their support of length s. Now, the first inequality (1.44), involves $\mathbf{U}$ and $\mathbf{x}$ and yields equalities only if $\mathbf{x}\in\Sigma_1$. The other case, follows directly from

$$\|\mathbf{Fx}\odot\mathbf{Fy}\|^2 = \sum_i|(\mathbf{Fx})_i|^2|(\mathbf{Fy})_i|^2 = |(\mathbf{Fx})_0|^2, \tag{1.47}$$

since for $y_i = 1/\sqrt{n}$ for all $i\in[n]$ we have $(\mathbf{Fy})_i = \delta_{i0}$. Then with $x_i \ge 0$ for $i\in I$ we get

$$\left|\frac{1}{\sqrt{n}}\sum_{i\in[n]}x_ie^{2\pi\imath\frac{i\cdot 0}{n}}\right|^2 = \frac{1}{n}\left|\sum_{i\in I}x_i\right|^2 = \frac{1}{n}\left(\sum_i|x_i|\right)^2 = \frac{1}{n}\|\mathbf{x}\|_1^2\;\|\mathbf{y}\|^2 \le \frac{s}{n}\|\mathbf{x}\|^2\|\mathbf{y}\|^2, \tag{1.48}$$

where in the last step equality holds if only if $x_i = 1/\sqrt{s}$ for $i \in I$. This shows, that the upper bound is not sharp if $|\operatorname{supp}(\mathbf{x})| = s, |\operatorname{supp}(\mathbf{y})| = f$ with $1 < s, f < n$!

The lower bound for $\mathbf{x} = \mathbf{y}$ follows from the CAUCHY-SCHWARTZ inequality, see for example [Web94], since and $\|\mathbf{1}\|^2 = n$ by

$$\|\mathbf{x} \underline{\odot} \mathbf{x}\|^2 = \frac{1}{n\|\mathbf{U}\|_\infty^2} \sum_{i=0}^{n-1} |(\mathbf{U}\mathbf{x})_i|^4 = \frac{1}{n^2\|\mathbf{U}\|_\infty^2} \|\mathbf{1}\|^2 \left\||\mathbf{U}\mathbf{x}|^2\right\|^2 \tag{1.49}$$

$$\geq \frac{1}{n^2\|\mathbf{U}\|_\infty^2} |\langle \mathbf{1}, |\mathbf{U}\mathbf{x}|^2 \rangle|^2 = \frac{1}{n^2\|\mathbf{U}\|_\infty^2} \|\mathbf{U}\mathbf{x}\|^4 = \frac{1}{n^2\|\mathbf{U}\|_\infty^2} \|\mathbf{x}\|^4 \tag{1.50}$$

for every $\mathbf{x} \in \mathbb{K}^n$. The last equality follows by the unitary property of $\mathbf{U}$. If $\mathbf{x}$ is a Euclidean unit vector, i.e., for some $i \in [n]$ we set $\mathbf{x} = \mathbf{e}_i$ and $\mathbf{U} = \mathbb{1}_n$, then we get equality since $\|\mathbf{x} \underline{\odot} \mathbf{x}\| = 1$. Setting $\mathbf{x} = \mathbf{1}/\sqrt{n}$ yields equality for $\mathbf{U} = \mathbf{F}$ with $\|\mathbf{x} \underline{\odot} \mathbf{x}\| = 1/\sqrt{n}$. □

Remark. Note that a lower bound strictly larger than zero for arbitrary $\mathbf{x} \neq \mathbf{y} \in \mathbb{K}^n$ only exists if $\mathbf{U}\mathbf{x}$ and $\mathbf{U}\mathbf{y}$ are disjoint. If the support sets intersects in exactly one index, then there exist for each $\epsilon > 0$ normalized vectors with $\langle \mathbf{x}, \mathbf{y} \rangle = \epsilon$, resulting in $\alpha = \epsilon$.

Hence, for $k = n$ we have with $(\mathbb{K}^n, \|\cdot\|, \underline{\odot})$ a *commutative normed algebra with unit.*

Corrolary 1. *The generalised commutative point product in* (1.40) *defines a* commutative normed algebra $(\ell^2_{[n]}, \underline{\odot})$.

1.3.1 Circular Convolution and Correlation

In this work we will focus on circular (cyclic) convolution type systems, see for example [OSB99], [Chu08].

Definition 8 (Circular Convolution and Correlation)**.** *The* circluar convolution $\circledast$ *and the* circular correlation $\circledast$ *on* $\mathbb{K}^n$ *are given for* $\mathbf{x}, \mathbf{y} \in \mathbb{K}^n$ *component-wise by*

$$(\mathbf{x} \circledast \mathbf{y})_l = \sum_{k=0}^{n-1} x_k y_{l \ominus k} \quad \textit{for} \quad l \in [n], \tag{1.51}$$

$$(\mathbf{x} \circledast \mathbf{y})_l = \sum_{k=0}^{n-1} x_k \overline{y}_{l \oplus k} \quad \textit{for} \quad l \in [n], \tag{1.52}$$

with the modulo$-n$ addition $\oplus$ *given for* $l, k \in \mathbb{Z}$ *by*

$$l \oplus k := l + k \mod n = (l + k) - \left\lfloor \frac{l+k}{n} \right\rfloor n, \tag{1.53}$$

with inverse operation $l \ominus k := l \oplus (-k)$ *and* floor *operation* $\lfloor a \rfloor := \max_{n \in \mathbb{Z}}\{n \leq a\}$.

Let us define the *circulant matrix* $\mathbf{Y} := \mathrm{circ}(\mathbf{y})$, see for example [Dav79, (3.1.1)], generated by $\mathbf{y}$ and powers of the right shift-matrix $\mathbf{S}$ (permutation matrix) as

$$\mathbf{Y} = \begin{pmatrix} y_0 & y_1 & \cdots & y_{n-1} \\ y_{n-1} & y_0 & \cdots & y_{n-2} \\ \vdots & & \ddots & \vdots \\ y_1 & y_2 & \cdots & y_0 \end{pmatrix} = \sum_i y_i \mathbf{S}^i \text{ with } \mathbf{S} := \begin{pmatrix} 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \end{pmatrix} \text{ and } \mathbf{S}^0 = \mathbb{1}_n. \tag{1.54}$$

Then we can write the convolution (1.51) in matrix form as

$$\mathbf{x} \circledast \mathbf{y} = \mathbf{Y}^T \mathbf{x}, \tag{1.55}$$

where $\mathbf{Y}^T$ is the transpose of $\mathbf{Y}$ given by

$$\mathbf{Y}^T = \begin{pmatrix} y_0 & y_{n-1} & \cdots & y_1 \\ y_1 & y_0 & \cdots & y_2 \\ \vdots & & \ddots & \vdots \\ y_{n-1} & y_{n-2} & \cdots & y_0 \end{pmatrix} = \sum_i y_i \mathbf{S}^{-i}. \tag{1.56}$$

Note that $\mathbf{S}^{-1} = \mathbf{S}^T$ is the left shift-matrix. We will need the following properties of the Fourier matrix, see for example [Dav79, p.33].

Proposition 1 (Fourier matrix)**.**

$$\mathbf{F}^* = \overline{\mathbf{F}} \quad , \quad \mathbf{F}^T = \mathbf{F} \quad , \quad (\mathbf{F}^*)^T = \mathbf{F}^* \qquad \textit{(symmetry)} \tag{1.57}$$

$$\mathbf{F}\mathbf{F}^* = \mathbf{F}^*\mathbf{F} = \mathbb{1}_n := \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \qquad \textit{(unitary)} \tag{1.58}$$

$$\mathbf{F}^2 = (\mathbf{F}^*)^2 = \mathbf{\Gamma} := \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1 \\ \vdots & & \iddots & \vdots \\ 0 & 1 & \dots & 0 \end{pmatrix} \qquad \textit{(time-reversal)} \tag{1.59}$$

$$(\mathbf{F}^*)^4 = \mathbf{F}^4 = \mathbf{\Gamma}^2 = \mathbb{1}_n \qquad \textit{(involution)}. \tag{1.60}$$

One of the remarkable properties of circulant matrices, is its diagonalization by the Fourier matrix, i.e., all circulant matrix share the same set of eigenvectors, see for example [Dav79, Thm. 3.2.2], given by

$$\mathbf{Y} = \sqrt{n}\mathbf{F}^* \mathrm{diag}(\mathbf{F}^*\mathbf{y})\mathbf{F} \tag{1.61}$$

In fact, (1.61) serves as a definition for circulant matrices, see for example [Dav79, Thm. 3.2.3]. For any $\boldsymbol{\lambda} \in \mathbb{C}^n$ it holds for the diagonal matrix $\mathrm{diag}(\boldsymbol{\lambda})$

$$\mathbf{\Gamma}\mathrm{diag}(\boldsymbol{\lambda})\mathbf{\Gamma} = \mathbf{\Gamma} \begin{pmatrix} \lambda_0 & 0 & \cdots & 0 \\ 0 & 0 & \dots & \lambda_1 \\ \vdots & & \iddots & \vdots \\ 0 & \lambda_{n-1} & \cdots & 0 \end{pmatrix} = \mathrm{diag}(\mathbf{\Gamma}\boldsymbol{\lambda}) \tag{1.62}$$

and we get for the transpose of $\mathbf{Y}$:

$$
\begin{aligned}
\mathbf{Y}^T &= (\sqrt{n}\mathbf{F}^*\mathrm{diag}(\mathbf{F}^*\mathbf{y})\mathbf{F})^T = \sqrt{n}\mathbf{F}^T\mathrm{diag}(\mathbf{F}^*\mathbf{y})\mathbf{F}^{*T} && (1.63)\\
(1.57) \to &= \sqrt{n}\mathbf{F}\mathrm{diag}(\mathbf{F}^*\mathbf{y})\mathbf{F}^* && (1.64)\\
(1.60) \to &= \sqrt{n}\mathbf{F}\mathbf{\Gamma}\mathbf{\Gamma}\mathrm{diag}(\mathbf{F}^*\mathbf{y})\mathbf{\Gamma}\mathbf{\Gamma}\mathbf{F}^* && (1.65)\\
(1.62) \to &= \sqrt{n}\mathbf{F}\mathbf{\Gamma}\mathrm{diag}(\mathbf{\Gamma}\mathbf{F}^*\mathbf{y})\mathbf{\Gamma}\mathbf{F}^* && (1.66)\\
(1.57), (1.59) \to &= \sqrt{n}\mathbf{F}^*\mathrm{diag}(\mathbf{F}\mathbf{y})\mathbf{F}. && (1.67)
\end{aligned}
$$

Hence, $\mathbf{Y}^T$ is also a circulant matrix generated by $\mathbf{\Gamma}\mathbf{y}$, with the eigenvalues $\boldsymbol{\lambda} = \sqrt{n}\mathbf{F}\mathbf{y}$. We can hence write the circular convolution as a point product:

$$
\mathbf{x} \circledast \mathbf{y} := \mathbf{Y}^T\mathbf{x} = \sqrt{n}\mathbf{F}^*\mathrm{diag}(\mathbf{F}\mathbf{y})\mathbf{F}\mathbf{x} = \sqrt{n}\mathbf{F}^*(\mathbf{F}\mathbf{y} \odot \mathbf{F}\mathbf{x}) = \sqrt{n}\mathbf{F}^*(\mathbf{F}\mathbf{x} \odot \mathbf{F}\mathbf{y}) = \mathbf{y} \circledast \mathbf{x} \quad (1.68)
$$

which establishes the commutativity. Note that this is not present for the circular correlation (the right shift with complex conjugated $\mathbf{y}$), since

$$
\mathbf{x} \circledstar \mathbf{y} := \mathbf{X}\overline{\mathbf{y}} = \sqrt{n}\mathbf{F}^*\mathrm{diag}(\mathbf{F}\mathbf{\Gamma}\mathbf{x})\mathbf{F}\overline{\mathbf{y}} = \sqrt{n}\mathbf{F}^*(\mathbf{F}\mathbf{\Gamma}\mathbf{x} \odot \mathbf{F}\overline{\mathbf{y}}) = \mathbf{\Gamma}\mathbf{x} \circledast \overline{\mathbf{y}}. \quad (1.69)
$$

The difference between convolution and correlation is therefore reflected by commutativity respectively non-commutativity of the associated algebra.

Lemma 3. *The normalized circular convolution* $\underline{\circledast}$ *and correlation* $\underline{\circledstar}$

$$
\begin{aligned}
\underline{\circledast}(\mathbf{x}, \mathbf{y}) &= \mathbf{F}^*(\mathbf{F}\mathbf{x} \odot \mathbf{F}\mathbf{y}) && (1.70)\\
\underline{\circledstar}(\mathbf{x}, \mathbf{y}) &= \mathbf{F}^*(\mathbf{F}^*\mathbf{x} \odot \mathbf{F}\overline{\mathbf{y}}). && (1.71)
\end{aligned}
$$

with involution *(vector)*

$$
\begin{aligned}
{}^* : \mathbb{K}^n &\to \mathbb{K}^n\\
\mathbf{x} &\mapsto \mathbf{x}^* = \overline{\mathbf{\Gamma}\mathbf{x}}
\end{aligned} \quad (1.72)
$$

define a finite dimensional convolution algebra $\mathfrak{B}_n^{\circledast} = (\ell_n^2, \underline{\circledast}, {}^*)$ *over* $\mathbb{K} \in \{\mathbb{C}, \mathbb{R}\}$*, which is a*

(i) commutative,

(ii) unital, i.e., there exists a unique unit $\mathbf{e}_0 := (1, 0, \dots, 0)^T$ *with* $\|\mathbf{e}_0\| = 1$*,*

(iii) associative and a

(iv) normed **−algebra.*

The correlation algebra $\mathfrak{B}_n^{\circledstar} = (\ell_n^2, \underline{\circledstar}, {}^*)$ *is non-commutative and has no unit element for* $\mathbb{K} = \mathbb{R}$ *and* $\mathbb{K} = \mathbb{C}$.

Proof. Obviously $\|\mathbf{F}\|_\infty = \|\mathbf{F}^*\|_\infty = \max_{l,k} |e^{-2\pi\imath\frac{lk}{n}}/\sqrt{n}| = 1/\sqrt{n}$. By Corollary 1 we obtain a commutative Banach algebra for the product in (1.70), which proves (i). Since $\mathbf{F}^*\mathbf{F} = \mathbb{1}_n$ and $\mathbf{F}\mathbf{e}_0 = n^{-1/2}\mathbf{1}$ we have $\mathbf{e}_0 \underline{\circledast} \mathbf{y} = \mathbf{F}^*\mathbf{F}\mathbf{y} = \mathbf{y}$ and by commutativity $\mathbf{y}\underline{\circledast}\mathbf{e}_0 = \mathbf{y}$, which shows (ii). Associativity (iii) follows from the point product and unitary property of $\mathbf{F}$. Indeed, (1.72) defines an involution, see for example [Grö01], since

$$(\mathbf{x}^*)^* = \overline{\mathbf{\Gamma}\overline{\mathbf{\Gamma}\mathbf{x}}} = \mathbf{\Gamma}\mathbf{\Gamma}\mathbf{x} = \mathbb{1}_n\mathbf{x} = \mathbf{x}. \tag{1.73}$$

Moreover, we get

$$(\mathbf{x}\underline{\circledast}\mathbf{y})^* = \overline{\mathbf{\Gamma}\mathbf{F}^*(\mathbf{F}\mathbf{x} \odot \mathbf{F}\mathbf{y})} \tag{1.74}$$
$$= \mathbf{\Gamma}\mathbf{F}(\overline{\mathbf{F}\mathbf{x}} \odot \overline{\mathbf{F}\mathbf{y}}) \tag{1.75}$$
$$= \mathbf{F}^*(\mathbf{F}\overline{\mathbf{\Gamma}\mathbf{x}} \odot \mathbf{F}\overline{\mathbf{\Gamma}\mathbf{y}}) \tag{1.76}$$
$$= \mathbf{F}^*(\mathbf{F}\mathbf{x}^* \odot \mathbf{F}\mathbf{y}^*) = \mathbf{y}^*\underline{\circledast}\mathbf{x}^*. \tag{1.77}$$

Note, that $\lambda^* = \overline{\lambda}$ and hence making the normed algebra $\mathfrak{B}_n^{\underline{\circledast}}$ to a normed *−algebra (iv), see for example [Pal94]. □

Remark. The point product $\odot$ does not have a normalized unit in $\ell^2_{[n]}$, since the unit $\mathbf{1}$ is unique and has norm $\|\mathbf{1}\| = \sqrt{n}$. Moreover, it is not possible to scale the point product to obtain a unital normed algebra, since it destroys the normed algebra property (1.22). Instead, one has to scale the norm.

1.4 Semi-Discrete Convolution

In the last chapter, we will deal with infinite dimensional Hilbert spaces. As already mentioned, convolution does not yield Banach algebras in the ℓ^2 or L^2-norm for infinite dimensions. One way to control the energy of the output signal is to convolve compactly supported signals $p \in L^2_T$ with sequences $\mathbf{c} \in \ell^2$, see [Boo87].

Definition 9 (Semi-discrete Convolution)**.** *Let be $q \geq 2$ and $T \subset \mathbb{R}$ a compact subset, then we define the* semi-discrete convolution *on $\ell^q \times L^2_T$ by*

$$(p *' \mathbf{c})(t) = \sum_{k=-\infty}^{\infty} c_k p(t-k) \quad \textit{for} \quad t \in \mathbb{R}. \tag{1.78}$$

We will show in Chapter 4, that indeed under some more restrictions on p, the semi-discrete convolution generates time-continuous signals in L^2, see [AU94]. Here, we have to deal with convergence problems, since infinite point products or general products on infinite groups need not converge in the same space.

2 Sparse and Restricted Multiplications

In this chapter we will consider signals $\mathbf{x} \in \ell^2$ with support length not larger than n. Hence we can identify the signals with finite dimensional vectors in $\mathbb{K}^n$ and refer by $\mathfrak{B}_n = (\mathbb{K}^n, \|\cdot\|, \mathcal{B}_n)$ to an $n-$dimensional normed algebra over $\mathbb{K}$.
Applying further sparsity constraints to the input signals we end up with a *sparse multiplication* in n dimensions.

Definition 10 (Sparse Multiplication)**.** *Let $s \le f \le n$ be integers and $\mathcal{B}_n$ a multiplication in $\mathbb{K}^n$, then we call*

$$\mathcal{B}_n(\Sigma_s^n, \Sigma_f^n) = \bigcup_{\substack{I,J\subset[n]\\|I|\le s,|J|\le f}} \mathcal{B}_n(\mathbb{K}_I^n, \mathbb{K}_J^n) \quad \textit{with} \quad \mathbb{K}_I^n = \operatorname*{span}_{i\in I}\{\mathbf{e}_i\}\ ,\ \mathbb{K}_J^n = \operatorname*{span}_{j\in J}\{\mathbf{e}_j\} \tag{2.1}$$

*the $(s,f)-$*sparse multiplication set generated by *$\mathcal{B}_n$, i.e., the set of multiplications of all s sparse and f sparse vectors in the Euclidean basis.*

Remark. To obtain sparse multiplication in another basis, we can simply apply unitary transformations to Σ_s^n and Σ_f^n and obtain a new multiplication $\mathcal{B}_n'$, see (1.40).

2.1 Representation and Stable Embedding

For describing an arbitrary signal set Z a representation is needed, which is given by a map $\mathcal{A}$ and a set U, called the pre-image of Z under $\mathcal{A}$. A stable representation is given by an invertible mapping (one-to-one), which preserves the distance in U up to a controllable distortion factor. To define such properties a metric d on U and a metric ρ on Z is needed, see for example [AMR02]. The pair (U, d) then is called a *metric space*. Note, that this is not necessarily a linear space! In *Banach geometry*, the property of preserving distance is know as *quasi-isometry*, *near-isometry* or simply $(1+\delta)-$*isometry (or $\delta-$embedding)*, see for example [Mat02]. In signal processing one is faced with the problem of a *robust* or *stable* reconstruction of the signals, where stability refers to an embedding (sampling) which is stable against Gaussian noise. Therefore, the metric of choice is the *Euclidean metric*

$$d_E(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\| \tag{2.2}$$

given by the ℓ^2-norm. Whenever distance, measuring in our case in the ℓ^2-norm, is almost preserved by the mapping we call the map $\mathcal{A}$ a *stable embedding*. See also [BCW10] for linear embeddings and [IM04] for $\delta-$embedding in metric spaces.

Definition 11 (Stable embedding)**.** *Let* $U \subset \mathbb{K}^n, \delta \in (0,1)$ *and* $\mathcal{A}\colon \mathbb{K}^n \to \mathbb{K}^m$ *a map. Then* $\mathcal{A}$ *is a* $\delta-$stable embedding *of* U *in* $\mathbb{K}^m$*, if and only if*

$$(1-\delta)\,\|\mathbf{u}_1 - \mathbf{u}_2\| \leq \|\mathcal{A}(\mathbf{u}_1 - \mathbf{u}_2)\| \leq (1+\delta)\,\|\mathbf{u}_1 - \mathbf{u}_2\| \quad \textit{for} \quad \mathbf{u}_1, \mathbf{u}_2 \in U. \tag{2.3}$$

Remark. Note, that neither U nor $\mathcal{A}$ need to have linear structures, i.e. there could exist $\mathbf{u}_1, \mathbf{u}_2 \in U$ with $\mathbf{u}_1 - \mathbf{u}_2 \notin U$ and $\mathcal{A}(\mathbf{u}_1 - \mathbf{u}_2) \neq \mathcal{A}(\mathbf{u}_1) - \mathcal{A}(\mathbf{u}_2)$.

In principle, there are two possibilities to represent Z in a linear fashion: either we linearize $X \times Y$ by the direct sum and use a non-linear map for representing the image Z or lifting $X \times Y$ to the $s \times f$ matrices, then Z is the image of all rank-1 matrices under a linear map, see Fig. 2.2. In both cases, we need a norm and hence a vector space. Note that all norms on finite dimensional spaces are equivalent (up to multiplication constants), hence it actually does not matter which norm we choose.

2.1.1 Lipschitz Embedding

The linear space with the smallest dimension generated by $\mathbb{K}^s$ and $\mathbb{K}^f$ is the *direct sum*

$$\mathbb{K}^s \oplus \mathbb{K}^f := \left\{ \lambda \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{y}_1 \end{pmatrix} + \mu \begin{pmatrix} \mathbf{x}_2 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \lambda \mathbf{x}_1 + \mu \mathbf{x}_2 \\ \lambda \mathbf{y}_1 + \mu \mathbf{y}_2 \end{pmatrix} \;\middle|\; \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{K}^s, \mathbf{y}_1, \mathbf{y}_2 \in \mathbb{K}^f, \lambda, \mu \in \mathbb{K} \right\}. \tag{2.4}$$

The map in (2.4) is a one-to-one mapping between $\mathbb{K}^s \times \mathbb{K}^f$ and $\mathbb{K}^s \oplus \mathbb{K}^f$. We can define a Hilbert space by the usual scalar product with the induced euclidean metric as

$$d_E(\mathbf{u}_1, \mathbf{u}_2) = \sqrt{|\langle \mathbf{u}_1 - \mathbf{u}_2, \mathbf{u}_1 - \mathbf{u}_2 \rangle|} = \|\mathbf{u}_1 - \mathbf{u}_2\| \quad \text{for} \quad \mathbf{u}_1, \mathbf{u}_2 \in \mathbb{K}^s \oplus \mathbb{K}^f, \tag{2.5}$$

which also defines a product topology, see for example [Wer02, p.35]. Hence, if we define the map

$$\begin{aligned} \psi\colon \mathbb{K}^s \oplus \mathbb{K}^f &\to \mathbb{K}^n \\ \mathbf{u} \mapsto \psi(\mathbf{u}) &:= \mathcal{B}_n^{s,f}((u_0, \dots, u_{s-1})^T, (u_s, \dots, u_{s+f-1})^T), \end{aligned} \tag{2.6}$$

we can represent Z by an $s+f$ dimensional parameter space $\mathbb{K}^{s+f} := \mathbb{K}^s \oplus \mathbb{K}^f$! But to guarantee preservation of the distance by ψ in Z, one needs a *stable embedding* (2.3), i.e.

$$\frac{1}{\alpha}\,\|\mathbf{u}_1 - \mathbf{u}_2\| \leq \|\psi(\mathbf{u}_1) - \psi(\mathbf{u}_2)\| \leq \alpha\,\|\mathbf{u}_1 - \mathbf{u}_2\| \quad \text{for} \quad \mathbf{u}_1, \mathbf{u}_2 \in \mathbb{K}^{s+f}, \tag{2.7}$$

for some $\alpha > 0$. The map ψ is then called an *$\alpha-$bi-Lipschitz map*, see e.g. [Rob09],[HK99],[ALN08],[Ver11]. In fact, it is easy to show, that ψ defined by the bilinear map in (2.6) is never a stable embedding, since for non-zero $\mathbf{u}_1 = (\mathbf{x}, \mathbf{0}_f)^T, \mathbf{u}_2 = (\mathbf{0}_s, \mathbf{y})^T$ one always gets

$$\|\mathbf{u}_1 - \mathbf{u}_2\| = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 > \left\|\mathcal{B}_n^{s,f}(\mathbf{x}, \mathbf{0}_f) - \mathcal{B}_n^{s,f}(\mathbf{0}_s, \mathbf{y})\right\| = \|\mathbf{0} - \mathbf{0}\| = 0. \tag{2.8}$$

This is inherited due to the non-trivial "nullset" of $\mathcal{B}_n^{s,f}$, which contains $(\mathbf{0}_s, \mathbb{K}^f) \cup (\mathbb{K}^s, \mathbf{0}_f)$, i.e., ψ is not injective. Note, that ψ does not define a linear map. Hence one has to exclude at least the nullspace, to obtain injectivity of ψ.

2.1.2 Injectivity, Unicity and Bilinear Maps

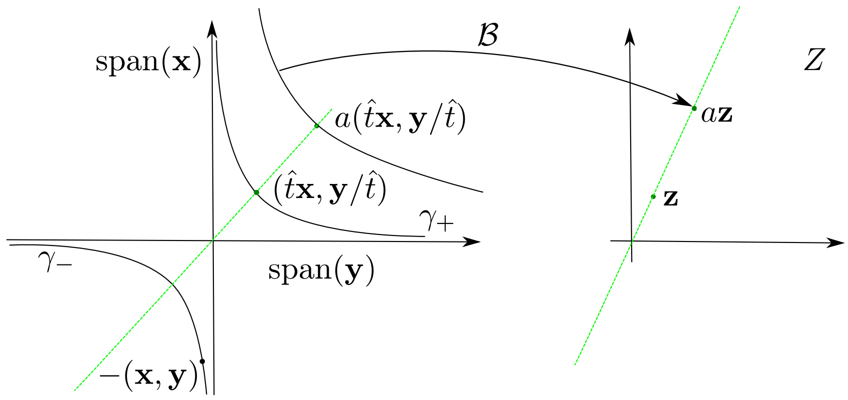


Figure 2.1: No injectivity for bilinear maps.

Although $\oplus$ and vec are one-to-one and hence invertible, the invertibility of the chain in Fig. 2.2 will be annihilated by the missing injectivity of the bilinear maps, preventing a stable embedding, since it holds for $(\mathbf{x}, \mathbf{y}) \in \mathbb{K}^s \times \mathbb{K}^f$ and any $t \notin \{0, 1\}$

$$\mathcal{B}_n^{s,f}(\mathbf{x}, \mathbf{y}) = \frac{t}{t}\mathcal{B}_n^{s,f}(\mathbf{x}, \mathbf{y}) = \mathcal{B}_n^{s,f}(t\mathbf{x}, \mathbf{y}/t), \tag{2.9}$$

but $(\mathbf{x}, \mathbf{y}) \neq (t\mathbf{x}, \mathbf{y}/t)$. In other words, for any image point $\mathbf{z}$ there exists an uncountable set of representation points given by the map

$$\gamma(t) = (t\mathbf{x}, \mathbf{y}/t) \quad \text{for} \quad t \in \mathbb{R} \setminus 0. \tag{2.10}$$

This in fact, defines two paths on a hyperbola for any $t_0 > 0$

$$\begin{aligned} \gamma_+(t) &= \gamma(t) \quad \text{for} \quad t \in [t_0, 1/t_0] \\ \gamma_-(t) &= \gamma(t) \quad \text{for} \quad t \in [-1/t_0, -t_0], \end{aligned} \tag{2.11}$$

see Fig. 2.1. Obviously, for every $(\mathbf{x}, \mathbf{y})$ pair, there exists two representation pairs, having the smallest distance to the origin, measured in the ℓ^2−norm of $\mathbb{K} \oplus \mathbb{K}$. This minimum is attained for $\hat{t} = \sqrt{\|\mathbf{y}\| / \|\mathbf{x}\|}$ by

$$\gamma(\hat{t}) = \left(\sqrt{\frac{\|\mathbf{y}\|}{\|\mathbf{x}\|}}\mathbf{x}, \sqrt{\frac{\|\mathbf{x}\|}{\|\mathbf{y}\|}}\mathbf{y} \right) \tag{2.12}$$

lying on the green axial line in Fig. 2.1 with distance to the origin

$$\left\|\gamma(\hat{t})\right\| = \sqrt{\|\mathbf{y}\| \, \|\mathbf{x}\| + \|\mathbf{y}\| \, \|\mathbf{x}\|} = \sqrt{2 \, \|\mathbf{y}\| \, \|\mathbf{x}\|}. \tag{2.13}$$

Hence, any pair on the green line, is unique up to a global sign. The length of $\mathbf{z}$ can then be scaled by some $a > 0$ which defines another hyperbola. So the whole linear cone $\mathrm{span}(\mathbf{z})$ is represented by $\mathrm{span}(\mathbf{x}) \oplus \mathrm{span}(\mathbf{y})$. If we expand this to $\mathbb{K} = \mathbb{C}$, the green line becomes a plane in a $4-$dimensional space ($\mathbb{C} \simeq \mathbb{R}^2$), where the representation points are on a sphere with radius $\sqrt{2\,\|\mathbf{y}\|\,\|\mathbf{x}\|}$ represented by the global phase $e^{\imath\omega}$.

This brings us to the following definition:

Definition 12 (Unicity)**.** *Let be U, V linear spaces. We call a map $\mathcal{A} : U \to V$* unique up to a global phase $e^{\imath\omega}$ *with $\omega \in [0, 2\pi)$ if*

$$\forall \mathbf{u}_1, \mathbf{u}_2 \in U \colon \mathcal{A}(\mathbf{u}_1) = \mathcal{A}(\mathbf{u}_2) \Leftrightarrow \mathbf{u}_1 = e^{\imath\omega}\mathbf{u}_2 \tag{2.14}$$

holds. We also denote this as the unicity *property of $\mathcal{A}$.*

Remark. The definition is borrowed from [CSV12, p.3].

To obtain uniqueness for a bilinear map, we have to restrict one input to the sphere. Hence $\mathbf{x}, \mathbf{y}$ defines $\mathbf{z}$ up to a phase

$$\begin{aligned} \mathcal{A} \colon X^{1,1} \times Y &\to \mathbb{K}^n \\ (\mathbf{x}, \mathbf{y}) &\mapsto \mathcal{A}(\mathbf{x}, \mathbf{y}) := \mathcal{B}(\mathbf{x}, \mathbf{y}). \end{aligned} \tag{2.15}$$

But obviously, $\mathcal{A}$ is not anymore a bilinear map, since it does not fulfill property (1.12).

2.1.3 Linear Embedding (Lifting Technique)

The natural way to get rid of the unicity problem is the tensor calculus, see for example [Gre67]. Here, we can lift a bilinear map $\otimes$, called the tensor map, into an $s \cdot f$ dimensional linear space $\mathbb{K}^{sf} := \mathbb{K}^s \otimes \mathbb{K}^f$, called the *tensor space* of $\mathbb{K}^s$ and $\mathbb{K}^f$. Defining the usual scalar product in $\mathbb{K}^{sf}$, the tensor space becomes a Hilbert space. The *tensor calculus* ensures the existence of a unique matrix $\mathbf{B}$ in the canonical basis[1], which represents Z in a linear fashion as the image of all simple tensors $\mathscr{K}_1 = \otimes(\mathbb{K}^s, \mathbb{K}^f) \subset \mathbb{K}^s \otimes \mathbb{K}^f$. Moreover, $\mathscr{K}_1$ is isomorphic to all rank-1 matrices

$$\mathscr{M}_1 := \left\{ \mathbf{x}\mathbf{y}^* \;\middle|\; \mathbf{x} \in \mathbb{K}^s, \mathbf{y} \in \mathbb{K}^f \right\} \subset \mathbb{K}^{s \times f} \tag{2.16}$$

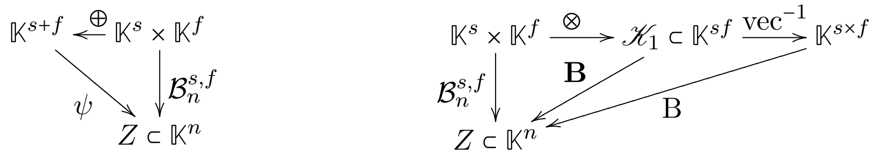
by the isomorphism[2]

$$\mathrm{vec}(\mathbf{x}\mathbf{y}^*) = (x_1\overline{y}_1, \ldots, x_1\overline{y}_f, x_2\overline{y}_1, \ldots, x_2\overline{y}_f, \ldots, x_s\overline{y}_1, \ldots, x_s\overline{y}_f)^T =: \mathbf{x} \otimes \bar{\mathbf{y}}. \tag{2.17}$$

Hence, the hyperbola in (2.11) is mapped to the tensor $\mathbf{u} = \mathbf{x} \otimes \overline{\mathbf{y}}$. The span is exactly a one-dimensional subspace in $\mathbb{K}^{sf}$ and corresponds to $\mathrm{span}(\mathbf{x}) \oplus \mathrm{span}(\overline{\mathbf{y}})$ in the previous section. Moreover, we can identify $\mathrm{span}(\mathbf{x} \otimes \overline{\mathbf{y}})$ by an element in the *projective space* $\mathbb{KP}^{sf-1}$, which is

[1] Canonical refers here to the Euclidean basis $\mathcal{E} = \{\mathbf{e}_0, \ldots, \mathbf{e}_{n-1}\}$, but any other fixed basis would also do.

[2] This is again a bilinear map and $\otimes \circ \mathrm{vec}^{-1}$ defines just another tensor product.



(a) Direct sum.

(b) Lifting to tensor space.

Figure 2.2: Direct sum and tensor space linearization of $\mathcal{B}_n^{s,f}$

equivalent to the *Grassmanian manifold* $G(1, sf)$, the set of all one dimensional subspaces in $\mathbb{K}^{sf}$, see for example [Lee13]. Then the matrix $\mathbf{B}$ transforms to the linear map $\mathrm{B} = \mathrm{B}_n \colon \mathbb{K}^{s\times f} \to \mathbb{K}^n$, see Fig. 2.2b. This linearization via tensor calculus was recently used in low-rank matrix recovery and is known as “lifting” [RFP10; CSV12].

Unfortunately, $\mathscr{K}_1$ and $\mathscr{M}_1$ are not linear spaces, but the span $\mathscr{K} = \mathbb{K}^{sf}$ and $\mathscr{M} = \mathbb{K}^{s\times f}$, which are by definition the smallest linear spaces containing $\mathscr{K}_1$ respectively $\mathscr{M}_1$. Note, by defining the equivalence relation $\mathbf{u} \sim \tilde{\mathbf{u}} \Leftrightarrow \mathbf{u} = t\mathbf{u}$ for some $t \in \mathbb{K}$ the set $\mathscr{K}_1 / \sim$ is the Grassmanian $G(1, sf)$. Defining the Frobenius norm (Schatten−2 norm) for $s \times f$ matrices by

$$\|\mathbf{U}\|_F := \left(\sum_{k=0}^{\min\{s,f\}-1} \sigma_k(\mathbf{U})^2 \right)^{1/2} \quad \text{for} \quad \mathbf{U} \in \mathbb{K}^{s\times f}, \tag{2.18}$$

where $\sigma_k(\mathbf{U})$ denotes the k−largest singular value of $\mathbf{U}$, see e.g. [HJ90, p.421]. One can easily show that the Frobenius and the Euclidean norm define a *cross norm* on $\mathbb{K}^s \times \mathbb{K}^f$, since it holds

$$\|\mathbf{x}\mathbf{y}^*\|_F = \|\mathbf{x}\| \, \|\mathbf{y}\| = \|\mathbf{x} \otimes \mathbf{y}\| \,. \tag{2.19}$$

Moreover, the map vec in (2.17) is *isometric*, see for example [Gre67], and hence $\mathbb{K}^{sf}$ is isomorph to $\mathbb{K}^{s\times f}$. In terms of *low-rank matrix recovery*, B_m represents linear (random) measurements of low-rank matrices $\mathbf{U} \in \mathbb{K}^{s\times f}$ for some $m \le n$ and the question arises, whenever from $\mathrm{B}_m(\mathbf{U}) \in \mathbb{K}^m$ a reconstruction of $\mathbf{U}$ is possible. A stable embedding of $\mathscr{M}_1$ in $\mathbb{K}^m$ would guarantee a unique reconstruction of all matrices in $\mathscr{M}_1$, i.e., we have to show the existence of $\alpha, \beta > 0$ s.t.

$$\alpha \|\mathbf{U}_1 - \mathbf{U}_2\|_F \le \|\mathrm{B}_m(\mathbf{U}_1 - \mathbf{U}_2)\| \le \beta \|\mathbf{U}_1 - \mathbf{U}_2\|_F \quad \text{for} \quad \mathbf{U}_1, \mathbf{U}_2 \in \mathscr{M}_1. \tag{2.20}$$

In [RFP10, Def.3.1] this is called the (asymmetric) *restricted isometry property* (RIP) of B for rank−2 matrices[3]. Hence, the RIP of B_m on $\mathscr{M}_2$ guarantees a stable embedding of $\mathscr{M}_1$ in $\mathbb{K}^m$.

If we express the middle term by our bilinear map, we get for $\mathbf{U}_1 = \mathbf{x}_1\mathbf{y}_1^*, \mathbf{U}_2 = \mathbf{x}_2\mathbf{y}_2^* \in \mathscr{M}_1$

$$\|\mathcal{B}(\mathbf{x}_1, \mathbf{y}_1) - \mathcal{B}(\mathbf{x}_2, \mathbf{y}_2)\| = \|\mathrm{B}(\mathbf{x}_1\mathbf{y}_1^*) - \mathrm{B}(\mathbf{x}_1\mathbf{y}_2^*)\| = \|\mathrm{B}(\mathbf{x}_1\mathbf{y}_1^* - \mathbf{x}_2\mathbf{y}_2^*)\| \,, \tag{2.21}$$

[3] In [RFP10] the map B is denoted with A.

but $\mathcal{B}(\mathbf{x}_1, \mathbf{y}_1) - \mathcal{B}(\mathbf{x}_2, \mathbf{y}_2) \neq \mathcal{B}(\mathbf{x}_1 - \mathbf{x}_2, \mathbf{y}_1 - \mathbf{y}_2)$. This is also seen in the cross norm

$$\|\mathbf{x}_1 - \mathbf{x}_2\| \, \|\mathbf{y}_1 - \mathbf{y}_2\| = \|(\mathbf{x}_1 - \mathbf{x}_2)(\mathbf{y}_1 - \mathbf{y}_2)^*\| \neq \|\mathbf{x}_1 \mathbf{y}_1^* - \mathbf{x}_2 \mathbf{y}_2^*\| \,. \tag{2.22}$$

Hence, our bilinear embedding has two problems:

1. For any linearization method, we do not get rid of the injectivity problem, i.e., neither $\psi : \mathbb{K}^{s+f} \to \mathbb{K}^n$ nor $\otimes : \mathbb{K}^s \times \mathbb{K}^f \to \mathscr{K}_1$ is injective.
2. The RNMP on $X \times Y$ cannot control differences.

Hence we cannot use the RNMP for a stable embedding without further restrictions, since difference in $\mathbb{K}^s$ and $\mathbb{K}^f$ does not correspond to difference in Z or $\mathscr{K}^1$. In the next section we will investigate these properties of bilinear maps in more detail and show a stable embedding result for quadratic maps.

2.1.4 RNMP and Stability

In this section we will use a more abstract definition of stability and injectivity in terms of unicity. Let us first define the symmetry property for arbitrary maps.

Definition 13. *Let X and Y be arbitrary sets. Then the map*

$$\mathcal{C} \colon X \times X \to Y \quad , \quad (\mathbf{x}, \mathbf{y}) \mapsto \mathcal{C}(\mathbf{x}, \mathbf{y}) \tag{2.23}$$

is called symmetric *if it holds*

$$\mathcal{C}(\mathbf{x}, \mathbf{y}) = \mathcal{C}(\mathbf{y}, \mathbf{x}) \quad \textit{for} \quad \mathbf{x}, \mathbf{y} \in X. \tag{2.24}$$

Moreover, $\mathcal{C}$ is called a symmetric homomorphisms, *if for each $\mathbf{x}$ respectively $\mathbf{y}$ the maps $\mathcal{C}(\mathbf{x}, \cdot)$ and $\mathcal{C}(\cdot, \mathbf{y})$ are homomorphism between the groups $(X, +)$ and $(Y, +)$.*

Let us call the map $\mathcal{A}(\mathbf{x}) := \mathcal{C}(\mathbf{x}, \mathbf{x})$ from X to Y defined by a symmetric homomorphism $\mathcal{C}$ a *quadratic map*. It then holds the relation:

$$\mathcal{A}(\mathbf{x}_1) - \mathcal{A}(\mathbf{x}_2) = \mathcal{C}(\mathbf{x}_1, \mathbf{x}_1) - \mathcal{C}(\mathbf{x}_2, \mathbf{x}_2) + \mathcal{C}(\mathbf{x}_1, \mathbf{x}_2) - \mathcal{C}(\mathbf{x}_1, \mathbf{x}_2) \tag{2.25}$$

$$\text{Homomorphism} \to = \mathcal{C}(\mathbf{x}_1, \mathbf{x}_1 + \mathbf{x}_2) - \mathcal{C}(\mathbf{x}_2 + \mathbf{x}_1, \mathbf{x}_2) \tag{2.26}$$

$$\text{Symmetry} \to = \mathcal{C}(\mathbf{x}_1 - \mathbf{x}_2, \mathbf{x}_1 + \mathbf{x}_2). \tag{2.27}$$

If X and Y are subsets of normed linear spaces $\mathscr{X}$ respectively $\mathscr{Y}$, then the RNMP for $\mathcal{C}$ on $X \times X$ corresponds to a stability property of $\mathcal{A}$ up to global sign. This is a weakened notion of stability as defined in Definition 11.

Definition 14 (Stability up to global sign)**.** *Let $\mathscr{X}, \mathscr{Y}$ be normed spaces, X some arbitrary subset in $\mathscr{X}$ and $\mathcal{C}$ a symmetric homomorphism. Then the map*

$$\mathcal{A}: \mathscr{X} \to \mathscr{Y} \quad , \quad \mathbf{x} \mapsto \mathcal{A}(\mathbf{x}) = \mathcal{C}(\mathbf{x}, \mathbf{x}) \tag{2.28}$$

is called stable on X up to global sign *if there exist a constant $c > 0$ such that it holds*

$$\|\mathcal{A}(\mathbf{x}_1) - \mathcal{A}(\mathbf{x}_2)\| \geq c \|\mathbf{x}_1 - \mathbf{x}_2\| \|\mathbf{x}_1 + \mathbf{x}_2\| \quad \textit{for} \quad \mathbf{x}_1, \mathbf{x}_2 \in X. \tag{2.29}$$

Remark. ELDAR and MENDELSON [EM12] called this *stability* of $\mathcal{A}$ (up to a global sign) for the real case. Moreover, stability implies injectivity of $\mathcal{A}$ on X, which is a much weaker property than stability. Note that their definition refers to a quadratic map without further restriction, which only makes sense for $\mathbb{K} = \mathbb{R}$.
However, a generalization to the complex case was done recently by EHLER, FORNASIER and SIGL for finite dimensions in [EFS13], but with the ℓ^2−norm substituted by the Hilbert Schmidt (Frobenius) norm:

$$\sum_{i=0}^{n-1} \left| |\langle \mathbf{a}_i, \mathbf{x}_1 \rangle|^2 - |\langle \mathbf{a}_i, \mathbf{x}_2 \rangle|^2 \right| \geq c \|\mathbf{x}_1 \mathbf{x}_1^* - \mathbf{x}_2 \mathbf{x}_2^*\|_F . \tag{2.30}$$

Hence it cannot be distinguished between units of modulus.

We will discuss our general approach in Section 2.3 and give an explicit application to the *Phase Retrieval* problem. Indeed, by demanding the symmetry condition and the explicit definition in (2.28), we can obtain stability up to a global sign even in the complex case.

2.2 RNMP for the Discrete Convolution

In this section, we will prove a fundamental norm inequality for the discrete convolution, the RNMP of the discrete convolution on all signals with finite support. This result will enable the concrete application of our compressed embedding results in the next chapter. Moreover, there seem to be countless applications from this inequality as well as a strong relation to known problems in different fields of mathematics.

The following theorem is a generalization of a result in [WJ13a], in the sense of the extension to complex vectors, which actually in the proof only is an extension from the SZEGÖ factorization to the FEJER-RIESZ factorization. Moreover, we already have embedded the finite dimensional signals in ℓ^2 and instead formulated the result for the discrete convolution of signals with finite support[4].

Theorem 2. *Let s,f be natural numbers with $s \leq f$. Then there exist a constant $\alpha_{\tilde{n}(s,f)} > 0$ with $\tilde{n}(s,f) = (s+f)^{10^8(s+f)^3 \log^3(s+f)}$, such that for all $\mathbf{x} \in \ell^2_s$ and $\mathbf{y} \in \ell^2_f$ it holds:*

$$\alpha_{\tilde{n}(s,f)} \|\mathbf{x}\| \|\mathbf{y}\| \leq \|\mathbf{x} * \mathbf{y}\| \leq \sqrt{s} \|\mathbf{x}\| \|\mathbf{y}\| . \tag{2.31}$$

Moreover, $\alpha^2_{\tilde{n}}$ is the smallest eigenvalue of all $\tilde{n} \times \tilde{n}$ Toeplitz matrices with symbol given by a normalized non-negative trigonometric polynomial of order $\tilde{n}-1$, which is a strictly decreasing sequence in $\tilde{n}$. For $s=1$ we get equality with $\alpha_1 = 1$.

Remark. We can relate the discrete convolution to the circular convolution by taking the smallest $n/2 \in \mathbb{N}$ such that $\operatorname{supp}(\mathbf{x}), \operatorname{supp}(\mathbf{y}) \subset [-n/2-1, n/2-1]$. Shifting $\mathbf{x}, \mathbf{y}$ to $\ell^2_{[n]}$ does not change the norm, i.e., we can identify the signals with vectors $\mathbf{x}, \mathbf{y} \in \mathbb{K}^n$. Appending now $n-1$ zeros to $\mathbf{x}$ and $\mathbf{y}$ yields equality between the discrete convolution and the circular convolution. Note that the number of zeros is sharp, since for $s=f=2$, $n' = 2n-2$ and every $n \geq 4$ it follows that $\|\mathbf{x}' \circledast \mathbf{y}'\| = 0$ for $\mathbf{x}', \mathbf{y}' \in (\mathbb{K}^n, \mathbf{0}_{n-2})$ when the non-zero components are $x'_0 = x'_{n-1} = y'_0 = \sqrt{1/2}$ and $y'_{n-1} = -y'_0$ and therefore $\alpha = 0$. Hence, we have for our sparse zero-padded circular convolution $\|\mathbf{x}' \circledast \mathbf{y}'\|_{\ell^2_{[\tilde{n}]}} = \|\mathbf{x} * \mathbf{y}\|$.

Proof: The upper-bound in (2.31) follows for any $\mathbf{x}' \in (\Sigma^n_s, \mathbf{0}_{n-1})$ and $\mathbf{y}' \in (\Sigma^n_f, \mathbf{0}_{n-1})$ by Lemma 2, setting $\underline{\odot} = \sqrt{n}\circledast$.

Before we prove the lower bound (2.31) under sparsity constrains we reformulate the general optimization problem of the lower bound (1.31) for the circular convolution in $\mathbb{K}^n$. By the absolute homogeneity of the norm and the convolution we get a *bi-quadratic* optimization problem

$$(P_{bq}) \qquad \alpha^2_{\text{opt}} := \min_{\mathbf{x},\mathbf{y} \in (\mathbb{K}^n)^{1,1}} \|\mathbf{x} \circledast \mathbf{y}\|^2 . \tag{2.32}$$

[4] Actually, the estimate of the dimension $\tilde{n}$ of the constant $\alpha_{\tilde{n}}$ in [WJ12b], was quite to optimistic. Nevertheless, the authors conjecture an $\mathcal{O}(s \cdot f)$ scaling of $\tilde{n}$, but the algorithm in [WJ12b] was wrong and has to be fare more extended. The following proof uses tools from additive combinatorics by embedding the original problem.

The objective function of this optimization problem is given by (1.68) as:

$$b(\mathbf{x},\mathbf{y}) := \|\mathbf{x} \circledast \mathbf{y}\|^2 = \left\|\mathbf{Y}^T\mathbf{x}\right\|^2 = \left\langle \mathbf{Y}^T\mathbf{x}, \mathbf{Y}^T\mathbf{x}\right\rangle = \left\langle \mathbf{x}, (\mathbf{Y}^T)^*\mathbf{Y}^T\mathbf{x}\right\rangle. \tag{2.33}$$

Obviously, the matrix $\mathbf{B}_\mathbf{y} = (\mathbf{Y}^T)^*\mathbf{Y}^T$ is positive semidefinite (Hermitian) and the problem (P_{bq}) reduces for fixed $\mathbf{y}$ to a *quadratic problem*

$$(P_q) \qquad \min_{\mathbf{x}\in\mathbb{K}^n, \|\mathbf{x}\|=1} b(\mathbf{x},\mathbf{y}) = \min_{\mathbf{x}\in\mathbb{K}^n, \|\mathbf{x}\|=1} \langle \mathbf{x}, \mathbf{B}_\mathbf{y}\mathbf{x}\rangle = \lambda(\mathbf{B}_\mathbf{y}), \tag{2.34}$$

which defines the smallest eigenvalue of $\mathbf{B}_\mathbf{y}$. It is known for the real case, $\mathbb{K} = \mathbb{R}$, that the problem is NP hard to solve, even by restricting to s respectively f dimensional subspaces X_I and Y_J, see [LNQY09, (1.1)] and [WJ13a]. Nevertheless, the $n \times n$ Hermitian matrix $\mathbf{B}_\mathbf{y}$ has a rich structure

$$\begin{aligned} \mathbf{B}_\mathbf{y} = (\mathbf{Y}^T)^*(\mathbf{Y})^T &= n\overline{\mathbf{F}\text{diag}(\mathbf{F}\mathbf{y})\mathbf{F}^*}\mathbf{F}^*\text{diag}(\mathbf{F}\mathbf{y})\mathbf{F} && (2.35)\\ &= n\mathbf{F}^*\text{diag}(\overline{\mathbf{F}\mathbf{y}})\text{diag}(\mathbf{F}\mathbf{y})\mathbf{F} && (2.36)\\ &= \sqrt{n}\mathbf{F}^*\text{diag}(\mathbf{F}^*(\sqrt{n}\mathbf{F}|\mathbf{F}\mathbf{y}|^2))\mathbf{F}. && (2.37)\end{aligned}$$

Hence, this is a *Hermitian circulant matrix* generated by the (circular) time-reverse *circular auto-correlation* of $\mathbf{y}$

$$\mathbf{\Gamma}(\mathbf{y} \circledast \mathbf{y}) = \sqrt{n}\mathbf{\Gamma}\mathbf{F}^*(\mathbf{F}\mathbf{y} \odot \mathbf{F}\mathbf{\Gamma}\overline{\mathbf{y}}) = \sqrt{n}\mathbf{F}|\mathbf{F}\mathbf{y}|^2 \tag{2.38}$$

having eigenvalues [Dav79, Thm. 3.2.2]

$$\lambda_k = n|\mathbf{F}\mathbf{y}|^2 = n\mathbf{F}(\mathbf{y} \circledast \mathbf{y}) \quad \text{for} \quad k \in [n]. \tag{2.39}$$

Since $\mathbf{\Gamma}(\mathbf{y} \circledast \mathbf{y})$ generates $\mathbf{B}_\mathbf{y}$ we have that $\mathbf{y} \circledast \mathbf{y}$ generates $\mathbf{B}_\mathbf{y}^T$ and the elements, ith row i'th column, are therefore given by the samples (1.52)

$$(\mathbf{B}_\mathbf{y})_{ii'} = (\mathbf{B}_\mathbf{y}^T)_{i'i} = \sum_j y_j \overline{y}_{j\oplus(i'\ominus i)} =: b_{i'\ominus i}(\mathbf{y}) \quad , \quad i, i' \in [n]. \tag{2.40}$$

Note that a circulant matrix is also a Toeplitz matrix [Dav79, p.70]. Hence $\mathbf{B}_\mathbf{y}$ is an Hermitian circulant Toeplitz matrix.

Unfortunately, the *circular auto-correlation* in (2.40) can generate a matrix having zero eigenvalues, at least for even dimensions n, see remark below of the Theorem 2. Hence, we have to demand further constraints. We will show later, that a sufficient condition to lower bound the smallest eigenvalue away from zero is given by the (aperiodic) *linear auto-correlation* (change modulo addition to normal addition).

To get rid of the boundary effects in the circular auto-correlation we have to embed $\mathbb{K}^n$ in $\mathbb{K}^{n'}$ with $n' := 2n - 1$ by adding $n - 1$ zeros $\mathbf{0}_{n-1}$ to each vector. This increases in the first place the dimension to an $2n-1\times 2n-1$ Hermitian Toeplitz matrix $\mathbf{B'}_{\mathbf{y}'}$ for $\mathbf{y}'=(\mathbf{y},\mathbf{0}_{n-1})^T$, given by

$$\mathbf{B'}_{\mathbf{y}'} := \left(\begin{array}{cccc|ccc} b_0 & b_1 & \cdots & b_{n-1} & b_n & \cdots & b_{2n-2} \\ \overline{b}_1 & b_0 & \cdots & b_{n-2} & b_{n-1} & \cdots & b_{2n-3} \\ \vdots & & \ddots & \vdots & \vdots & \ddots & \vdots \\ \overline{b}_{n-1} & \overline{b}_{n-2} & \cdots & b_0 & b_1 & \cdots & b_{n-1} \\ \hline \overline{b}_n & \overline{b}_{n-1} & \cdots & \overline{b}_1 & b_0 & \cdots & b_{n-1} \\ \vdots & & \ddots & \vdots & \vdots & \ddots & \vdots \\ \overline{b}_{2n-2} & \overline{b}_{2n-3} & \cdots & \overline{b}_n & \overline{b}_{n-1} & \cdots & b_0 \end{array}\right) = \left(\begin{array}{c|c} \mathbf{B}_{\mathbf{y}'} & \mathbf{E}_{\mathbf{y}'} \\ \hline \mathbf{E}^*_{\mathbf{y}'} & \mathbf{B}^{n-1}_{\mathbf{y}'} \end{array}\right). \tag{2.41}$$

But in the second place, the dimension boils down to n, since $\mathbf{x}'=(\mathbf{x},\mathbf{0}_{n-1})^T$ cuts out the $n\times n$ principal submatrix $\mathbf{B}_{\mathbf{y}'}$, which is still a Hermitian Toeplitz matrix (not anymore circulant), i.e., we get

$$(P'_q) \qquad \min_{\mathbf{x}'\in(\mathbb{K}^n,\mathbf{0}_{n-1}),\|\mathbf{x}'\|=1} b(\mathbf{x}',\mathbf{y}') = \min_{\mathbf{x}\in\mathbb{K}^n,\|\mathbf{x}\|=1} \langle \mathbf{x},\mathbf{B}_{\mathbf{y}'}\mathbf{x}\rangle = \lambda(\mathbf{B}_{\mathbf{y}'}). \tag{2.42}$$

Then we obtain for the first row respectively complex conjugated column in $\mathbf{B}_{\mathbf{y}'}$

$$\overline{b_{-k}(\mathbf{y}')} = b_k(\mathbf{y}') = \sum_{j=0}^{n'-1} y'_j \overline{y'}_{j\oplus k} = \sum_{j=0}^{n-1} y_j \overline{y'}_{j+k} = \sum_{j=0}^{n-1-k} y_j \overline{y}_{j+k} =: b^a_k(\mathbf{y}) \quad , \quad k\in[0,n-1] \tag{2.43}$$

and $\mathbf{B}_{\mathbf{y}'} =: \mathbf{B}^a_{\mathbf{y}}$ is given by samples of the *linear auto-correlation* (2.43) of $\mathbf{y}\in\mathbb{K}^n$.

The second constraint of the theorem is given by the sparsity of $\mathbf{x}$ and $\mathbf{y}$. Let us define the set of all subsets in $[n]$ with length s by

$$[n]_s := \{\, A\subset[n] \mid |A| = s \,\}. \tag{2.44}$$

Then, for any $\mathbf{x}\in\Sigma^n_s, \mathbf{y}\in\Sigma^n_f$ there exist $I\in[n]_s, J\in[n]_f$, such that $\operatorname{supp}(\mathbf{x})\subset I, \operatorname{supp}(\mathbf{y})\subset J$. Applying this to our quadratic optimization problem (P'_q), we see that $\mathbf{x}$ cuts out from $\mathbf{B}^a_{\mathbf{y}}$ an $s\times s$ Hermitian Toeplitz matrix $\mathbf{B}^{a,I}_{\mathbf{y}}$ with coefficients $k\in I\ominus i_0 = I - i_0 = \{k_0,\dots,k_{s-1}\} =: K\in[n]_s$ given by

$$b^a_{k_\theta}(\mathbf{y}) = \sum_{j=0}^{n-1-|k_\theta|} y_j\overline{y}_{j+k_\theta} =: b^{a,I}_\theta(\mathbf{y}) \quad , \quad \theta\in[s]. \tag{2.45}$$

Since $\mathbf{B}^a_{\mathbf{y}}$ is Hermitian so is $\mathbf{B}^{a,I}_{\mathbf{y}}$, hence $b^{a,I}_\theta = \overline{b^{a,I}_{-\theta}}$. Surely, for $\mathbf{y}\in\Sigma^n_f$ the sum in (2.45) has at most f summands, but since we have to consider all support sets J we have to sum over all terms. Assume now there is exists for sufficiently large n an universal $\tilde{n}=\tilde{n}(s,f)\le n$, such that for each $I\in[n]_s$ and $\mathbf{y}\in\Sigma^n_f$ with support in $J\in[n]_s$ there exist a vector $\tilde{\mathbf{y}}\in\Sigma^{\tilde{n}}_f$ with $\operatorname{supp}(\tilde{\mathbf{y}})\subset\tilde{J}\in[\tilde{n}]_f$ and a set $\tilde{K}:=\{\tilde{k}_\theta\}_{\theta=0}^{s-1}\in[\tilde{n}]_s$ satisfying

$$b^{a,\tilde{n}}_{\tilde{k}_\theta}(\tilde{\mathbf{y}}) := \sum_{j=0}^{\tilde{n}-1-|\tilde{k}_\theta|} \tilde{y}_j\overline{\tilde{y}}_{j+\tilde{k}_\theta} = \sum_{j=0}^{n-1-|k_\theta|} y_j\overline{y}_{j+k_\theta} = b^a_{k_\theta}(\mathbf{y}) \quad , \quad \theta\in[s]. \tag{2.46}$$

Note that $\tilde{k}_\theta \neq k_\theta$. In fact, (2.46) implies for a fixed $\mathbf{y} \in \Sigma_f$ with $\|\mathbf{y}\| = 1$ and $I \in [n]_s$ the existence of an $s \times s$ principal submatrix $\mathbf{B}_{\mathbf{y}}^{a,I}$ in the $\tilde{n} \times \tilde{n}$ Hermitian Toeplitz matrix

$$\mathbf{B}_{\tilde{\mathbf{y}}}^{a,\tilde{n}} = \begin{pmatrix} b_0^{a,\tilde{n}}(\tilde{\mathbf{y}}) & \dots & b_{\tilde{n}-1}^{a,\tilde{n}}(\tilde{\mathbf{y}}) \\ \vdots & \ddots & \vdots \\ b_{\tilde{n}-1}^{a,\tilde{n}}(\tilde{\mathbf{y}}) & \dots & b_0^{a,\tilde{n}}(\tilde{\mathbf{y}}) \end{pmatrix} \quad , \quad b_k^{a,\tilde{n}}(\tilde{\mathbf{y}}) := \sum_{j=0}^{\tilde{n}-1-k} \tilde{y}_j \overline{\tilde{y}}_{j+k} \quad \text{for} \quad k \in [\tilde{n}] \tag{2.47}$$

with positive *symbol* given for $\omega \in [0, 2\pi)$ and $\|\tilde{\mathbf{y}}\| = 1$ by

$$\begin{aligned} b^{a,\tilde{n}}(\tilde{\mathbf{y}}, \omega) = \sum_{k=1-\tilde{n}}^{\tilde{n}-1} b_k^{a,\tilde{n}}(\tilde{\mathbf{y}}) e^{\imath k\omega} = 1 + \sum_{k=1}^{\tilde{n}-1} \left(a_k \cos(k\omega) + c_k \sin(k\omega)\right) \\ \text{with} \quad a_k := 2\Re(b_k^{a,\tilde{n}}(\tilde{\mathbf{y}})) \quad \text{and} \quad c_k := -2\Im(b_k^{a,\tilde{n}}(\tilde{\mathbf{y}})), \end{aligned} \tag{2.48}$$

see for example [BG05]. We call (2.48) a *normalized trigonometric polynomial of order* $\tilde{n} - 1$ since $a_0 = \|\tilde{\mathbf{y}}\|^2 = 1$.

Then by CAUCHY'S Interlacing Theorem, see for example [BG05, Prop.9.19], the smallest eigenvalue of any $\mathbf{B}_{\mathbf{y}}^{a,I}$ is bounded from below by the smallest eigenvalue of $\mathbf{B}_{\tilde{\mathbf{y}}}^{a,\tilde{n}}$, i.e.,

$$\alpha_{s,f,n} \geq \min_{I \in [n]_s} \min_{\substack{\mathbf{y} \in \Sigma_f^n \\ \|\mathbf{y}\|=1}} \lambda(\mathbf{B}_{\mathbf{y}}^{a,I}) \geq \min_{\substack{\tilde{\mathbf{x}}, \tilde{\mathbf{y}} \in \mathbb{K}^{\tilde{n}} \\ \|\tilde{\mathbf{x}}\| = \|\tilde{\mathbf{y}}\| = 1}} \left\langle \tilde{\mathbf{x}}, \mathbf{B}_{\tilde{\mathbf{y}}}^{a,\tilde{n}} \tilde{\mathbf{x}} \right\rangle = \min_{\substack{\tilde{\mathbf{y}} \in \mathbb{K}^{\tilde{n}} \\ \|\tilde{\mathbf{y}}\|=1}} \lambda(\mathbf{B}_{\tilde{\mathbf{y}}}^{a,\tilde{n}}) =: \alpha_{\tilde{n}}, \tag{2.49}$$

where the argument on the right hand side is independent of I and J. By the well-known FEJER-RIESZ Factorization, see for example [Dim04, Thm.3], we know that the symbol of $\mathbf{B}_{\tilde{\mathbf{y}}}^{a,\tilde{n}}$ in (2.48) is a *non-negative trigonometric polynomial*[5] of order $\tilde{n} - 1$ for every normalized $\tilde{\mathbf{y}} \in \mathbb{K}^{\tilde{n}}$ since $b_k^{a,\tilde{n}}(\tilde{\mathbf{y}})$ is given by (2.46), i.e., $0 \leq \min_\omega b^{a,\tilde{n}}(\tilde{\mathbf{y}}, \omega)$. By [BG05, (10.2)] we then have $\lambda(\mathbf{B}_{\tilde{\mathbf{y}}}^{a,\tilde{n}}) > 0$. Hence $\mathbf{B}_{\tilde{\mathbf{y}}}^{a,\tilde{n}}$ is invertible and the *determinant* $\det(\mathbf{B}_{\tilde{\mathbf{y}}}^{a,\tilde{n}}) \neq 0$. Using

$$\frac{1}{\lambda(\mathbf{B}_{\tilde{\mathbf{y}}}^{a,\tilde{n}})} = \left\| \mathbf{B}_{\tilde{\mathbf{y}}}^{a,\tilde{n}} \right\|_2 \tag{2.50}$$

in [BG05, p.59], we can estimate the smallest eigenvalue (singular value) with [BG05, Thm. 4.2] by the determinant as

$$\lambda(\mathbf{B}_{\tilde{\mathbf{y}}}^{a,\tilde{n}}) \geq |\det(\mathbf{B}_{\tilde{\mathbf{y}}}^{a,\tilde{n}})| \frac{1}{\sqrt{\tilde{n}}(\sum_k |b_k^{a,\tilde{n}}(\tilde{\mathbf{y}})|^2)^{(\tilde{n}-1)/2}}, \tag{2.51}$$

where the ℓ^2–norm of the sequence $b_k^{a,\tilde{n}}(\tilde{\mathbf{y}})$ can be upper bounded for $\tilde{n} > 1$ by the CAUCHY-SCHWARTZ inequality to

$$\sum_k |b_k^{a,\tilde{n}}(\tilde{\mathbf{y}})|^2 \leq 1 + 2 \sum_{k=1}^{\tilde{n}-1} |\sum_{j=0}^{\tilde{n}-1} \tilde{y}_j \overline{\tilde{y}}_{j+k}|^2 \leq 1 + 2 \sum_{k=1}^{\tilde{n}-1} \|\tilde{\mathbf{y}}\|^4 = 1 + 2(\tilde{n} - 1) < 2\tilde{n}, \tag{2.52}$$

[5] Note, there exist $\tilde{\mathbf{y}} \in \mathbb{K}^{\tilde{n}}$ with $\|\tilde{\mathbf{y}}\| = 1$ and $b^{a,\tilde{n}}(\tilde{\mathbf{y}}, \omega) = 0$ for some $\omega \in [0, 2\pi)$. Thats the reason why things are more complicated here. Moreover, we want to find a universal lower bound over all $\tilde{\mathbf{y}}$, which is equivalent to a universal lower bound over all non-negative trigonometric polynomials of order $\tilde{n} - 1$. By the best knowledge of the authors, there exist no analytic lower bound for $\alpha_{\tilde{n}}$.

which is independent of $\tilde{\mathbf{y}} \in (\mathbb{K}^{\tilde{n}})^{1,1}$. Since the determinant is a continuous function in $\tilde{\mathbf{y}}$ over a compact set, the minimum is attained and is denoted by $0 < d_{\tilde{n}} := \min_{\tilde{\mathbf{y}}} |\det(\mathbf{B}_{\tilde{\mathbf{y}}}^{a,\tilde{n}})|$. Note, that $d_{\tilde{n}}$ is a decreasing sequence, since we extend the minimum to a larger set by increasing $\tilde{n}$. Hence we get

$$\min_{\tilde{\mathbf{y}}} \left(|\det(\mathbf{B}_{\tilde{\mathbf{y}}}^{a,\tilde{n}})| \frac{1}{\sqrt{\tilde{n}}(2\tilde{n})^{(\tilde{n}-1)/2}} \right) = \frac{\sqrt{2}}{(2\tilde{n})^{\tilde{n}/2}} d_{\tilde{n}}. \tag{2.53}$$

This is a valid lower bound by (2.51) for the smallest eigenvalue of all possible $\mathbf{B}_{\tilde{\mathbf{y}}}^{a,\tilde{n}}$. Hence we have shown

$$\alpha_{\text{opt}} \geq \min_{\substack{\tilde{\mathbf{y}} \in \tilde{\Sigma}_f \\ \|\tilde{\mathbf{y}}\|=1}} \min_{\substack{\tilde{\mathbf{x}} \in \tilde{\Sigma}_s \\ \|\tilde{\mathbf{x}}\|=1}} b(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) > \sqrt{2}(2\tilde{n})^{-\frac{\tilde{n}}{2}} d_{\tilde{n}} > 0. \tag{2.54}$$

We will now show, how $\tilde{n}$ can be chosen for $2 \leq s \leq sf \leq n$. If $s = 1$, then $K = I - i_0 = \{0\} = \{k_0\}$ and $\mathbf{B}_{\mathbf{y}}^{a,I} = \mathbb{1}_n$ for all $I \in [n]_1$ and $\mathbf{y} \in \Sigma_f$. Hence the smallest (and single) eigenvalue is 1 for every $\mathbf{B}_{\tilde{\mathbf{y}}}^{a,1}$, i.e., $\alpha_1 = 1$. Hence, we will demand $s \geq 2$.
The embedding of $\mathbb{K}^s$ respectively $\mathbb{K}^f$ in $\mathbb{K}^n$ with support $I \in [n]_s$ resp. $J \in [n]_f$ is given by

$$x_i := \sum_{\theta=0}^{s-1} \delta_{i,i_\theta} \hat{x}_\theta \quad \text{and} \quad y_i := \sum_{\gamma=0}^{f-1} \delta_{i,j_\gamma} \hat{y}_\gamma \quad , \quad i \in [n], \tag{2.55}$$

$$\tilde{y}_{\tilde{j}} := \sum_{\gamma=0}^{f-1} \delta_{\tilde{j},\tilde{j}_\gamma} \hat{y}_\gamma \quad , \quad \tilde{j} \in [\tilde{n}]. \tag{2.56}$$

According to (2.46) we have to show that for each $\mathbf{x} \in \Sigma_s^n, \mathbf{y} \in \Sigma_f^n$ there exists sets $\tilde{K} = \{\tilde{k}_\theta\}_{\theta \in [s]}, \tilde{J} = \{\tilde{j}_\gamma\}_{\gamma \in [f]} \subset [\tilde{n}]$ and a vector $\tilde{\mathbf{y}} \in \mathbb{K}^{\tilde{n}}$ given in (2.56), such that it holds:

$$\sum_{\tilde{j}=0}^{\tilde{n}-1-|\tilde{k}_\theta|} \tilde{y}_{\tilde{j}} \overline{\tilde{y}}_{\tilde{j}+\tilde{k}_\theta} - \sum_{j=0}^{n-1-|k_\theta|} y_j \overline{y}_{j+k_\theta} = 0 \quad , \quad \theta \in [s]. \tag{2.57}$$

Since the first index in K is 0, we will define $[n]_s^0 := (0, \{ A \subset [n] \setminus 0 \mid |A| = s-1 \})$.

Equation (2.57) is equivalent, by using the embedding in (2.55) and (2.56), to

$$\begin{aligned} \sum_{\gamma,\gamma'} \left[\sum_{\tilde{j}=0}^{n-1-\tilde{k}_\theta} \delta_{\tilde{j},\tilde{j}_\gamma} \delta_{\tilde{j}+\tilde{k}_\theta,\tilde{j}_{\gamma'}} - \sum_{j=0}^{n-1-k_\theta} \delta_{j,j_\gamma} \delta_{j+k_\theta,j_{\gamma'}} \right] \hat{y}_\gamma \overline{\hat{y}}_{\gamma'} & \\ = \sum_{\gamma,\gamma'} \left(\delta_{\tilde{j}_\gamma+\tilde{k}_\theta,\tilde{j}_{\gamma'}} - \delta_{j_\gamma+k_\theta,j_{\gamma'}} \right) \hat{y}_\gamma \overline{\hat{y}}_{\gamma'} = 0 & \end{aligned} \quad \text{for all} \quad \hat{\mathbf{y}} \in \mathbb{R}^f. \tag{2.58}$$

Hence we have to show for every $(K, J) \in [n]_s^0 \times [n]_f$ the existence of $(\tilde{K}, \tilde{J}) \in [\tilde{n}]_s^0 \times [\tilde{n}]_f^0$, such that it holds for all $\theta \in [s], \gamma, \gamma' \in [f]$

$$\mathcal{D}_{\theta\gamma\gamma'}(\tilde{K}, \tilde{J}) := \delta_{\tilde{j}_\gamma+\tilde{k}_\theta,\tilde{j}_{\gamma'}} = \delta_{j_\gamma+k_\theta,j_{\gamma'}} =: \mathcal{D}_{\theta\gamma\gamma'}(K, J), \tag{2.59}$$

where $\mathcal{D}$ is a tensor[6] in $[s] \times [f] \times [f]$.

Approach via Additive Combinatorics

The tensor $\mathcal{D}$ in (2.59) is given by addition of natural numbers. The numbers in $J, K \subset [n]$ could be arbitrarily large, since $n \to \infty$. Hence we can assume $J, K \subset \mathbb{N}$ or even $J, K \subset \mathbb{Z}$. The problem of finding sets $\tilde{J}, \tilde{K}$ in $[\tilde{n}]$ satisfying (2.59) can be formulated in *additive number theory* by setting

$$A := J \cup K\ ,\ |A| \leq s + f. \tag{2.60}$$

Then the addition

$$j + k = j' + 0 \quad \text{for} \quad j, j' \in J,\ k \in K \tag{2.61}$$

can be realized by numbers in A setting $a_1 = j, a_2 = k, a_1' = j', a_2' = 0$. Note that $0 \in K$ and hence in A. If such numbers in A satisfy the equality in (2.61), then we want to map these numbers one-to-one to numbers in $[\tilde{n}]$ obeying the same addition. Such a map ϕ is precisely known as a *Freiman Isomorphism ϕ of order* 2 between A and $\tilde{A} = \phi(A) \subset \mathbb{Z}$, i.e., $\phi : A \to \tilde{A}$ is bijective and satisfies

$$a_1 + a_2 = a_1' + a_2' \Leftrightarrow \phi(a_1) + \phi(a_2) = \phi(a_1') + \phi(a_2') \quad , \quad a_1, a_2, a_1', a_2' \in A. \tag{2.62}$$

Let us stress a very important fact here, a *Freiman homomorphism*, i.e., a map $\phi : A \to \mathbb{Z}$

$$a_1 + a_2 = a_1' + a_2' \Rightarrow \phi(a_1) + \phi(a_2) = \phi(a_1') + \phi(a_2') \quad , \quad a_1, a_2, a_1', a_2' \in A. \tag{2.63}$$

is not a group homomorphism, it is a weaker definition! So $\phi(0)$ is not necessarily 0. However, since the translation map on $\mathbb{Z}$ is a Freiman homomorphism and the composition of two Freiman homomorphisms is again a Freiman homomorphism [TV06, Sec. 5.3], we can always assume, that our Freiman homomorphism (isomorphism) is normalized, i.e., $\phi(0) = 0$. This is necessary to show (2.59), since we then have

$$j + k = j' \qquad j, k, j' \Leftrightarrow \phi(j) + \phi(k) = \phi(j') \in A. \tag{2.64}$$

Hence, if ϕ is a Freiman isomorphism (of order 2) we get $\tilde{J} = \phi(J), \tilde{K} = \phi(K)$ satisfying[7] (2.59).

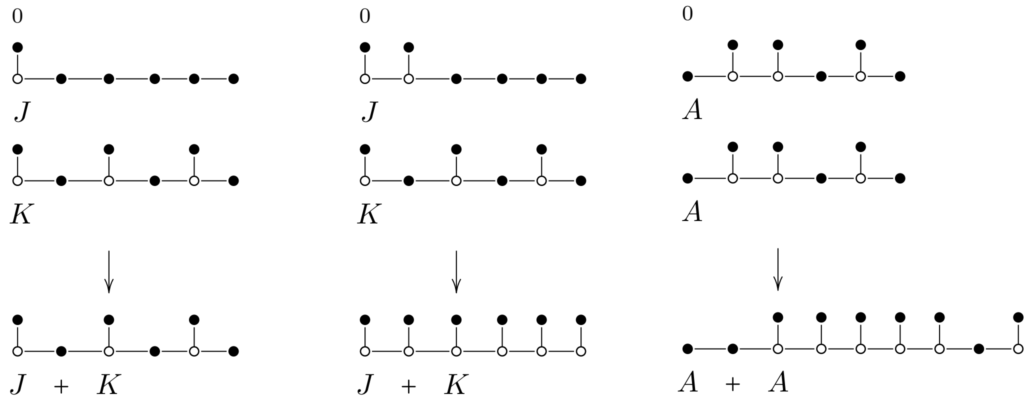
In additive number theory a lot is known about sets with small *doubling constants*[8]

$$\sigma = \sigma(A) := \frac{|A + A|}{|A|}. \tag{2.65}$$

[6]In fact the tensor $\mathcal{D}(K, J)$ on the right in (2.59) defines an *equivalence relation* in $[n]_s^0 \times [n]_f$, i.e. $(K, J) \sim_{\mathcal{D}} (K', J') \Leftrightarrow \mathcal{D}(K, J) = \mathcal{D}(K', J')$. The task is to show, that for each equivalence relation $[(K, J)]_{\mathcal{D}}$ there exists an element $(\tilde{K}, \tilde{J}) \in [(K, J)]_{\mathcal{D}}$ such that $(\tilde{K}, \tilde{J}) \in [\tilde{n}]_s^0 \times [\tilde{n}]_f^0 \subset [n]_s^0 \times [n]_f$.

[7]Note, that $\tilde{J}$ and $\tilde{K}$ are not necessarily ordered, which is for (2.59) not relevant. Going back to the Toeplitz-Matrix, we can in (2.46) order the coefficients, i.e., ordering $\tilde{K}$ and find the submatrix $\mathbf{B}_{\mathbf{y}}^I$ in (2.47).

[8]The Minkowski sum is given as $A + A = 2A = \{a_1 + a_2 \mid a_1, a_2 \in A\}$.



(a) $|J + K| = |J||K| = 3$. (b) $|J + K| = |J||K| = 6$. (c) $|A + A| = \frac{|A|+1}{2}|A| = 6$.

Figure 2.3: Set addition and convolution, $\mathrm{supp}(\chi_J * \chi_K) = J + K$.

Unfortunately, J, K and hence A are arbitrary, i.e., we can only assume [TV06, Lem.2.1]

$$\sigma \le \frac{s + f + 1}{2}. \tag{2.66}$$

On the other hand consider a normalized *d−dimensional (generalized arithmetic) progression*

$$P := \left\{ \langle \mathbf{b}, \mathbf{v} \rangle = \sum_{i=1}^{d} b_i v_i \;\middle|\; b_i \in [0, n_i], i = 1, \dots, d \right\}, \tag{2.67}$$

where $\mathbf{v} \in \mathbb{Z}^d$ is the basis vector and $\mathbf{n} \in \mathbb{Z}^d$ defines the edges of the box $B = [0, n_1] \times \cdots \times [0, n_d]$ in $\mathbb{Z}^d$, then the cardinality is $|P| \le \mathrm{Vol}(B) = l(P) := \Pi_{i=1}^{d}(n_i + 1)$. If $|P| = \mathrm{Vol}(B)$, i.e. cardinality equals the volume of B or equivalent $\langle \cdot, \mathbf{v} \rangle$ is injective on B, then P is called *proper*. Hence, a proper progression has a very small doubling constant given by $\sigma \le 2^d$, since

$$B + B = [0, n_1] \times \cdots \times [0, n_d] + [0, n_1] \times \cdots \times [0, n_d] = [0, 2n_1] \times \cdots \times [0, 2n_d] = 2B, \tag{2.68}$$

yielding $|2P| \le \mathrm{Vol}(2B) = \Pi_i(2n_i + 1) \le 2^d \mathrm{Vol}(B) = 2^d |P|$. If P is 2−proper, i.e., $P + P$ is proper, then $\langle \cdot, \mathbf{v} \rangle$ is an Freiman isomorphism between B and P. Moreover, $\langle \cdot, \mathbf{v} \rangle$ is an homomorphism on $\mathbb{Z}^d$.

If A has the largest possible doubling constant, then it is a *Sidon set*, see [TV06, Def. 4.27]. For these sets, it is known, that they are Freiman isomorphic to each other [TV06, p. 5.3.9], hence also to the geometric series $\phi(A) = V = \{1, 2, 4, 8, \dots, 2^{d-1}\}$, which span a proper *geometric progression*

$$G = \left\{ \sum_{i=1}^{d} b_i 2^{i-1} \;\middle|\; b_i \in [0, 1] \right\} \tag{2.69}$$

with $d = |A|$, see also Fig. 2.3. Hence $\phi(A) = V \subset G \subset [0, 2^{|A|} - 1]$, i.e. for Sidon sets we have $\tilde{n} = 2^{s+f}$. In fact, there is the following conjecture by KONYAGIN and LEV in [KL00]

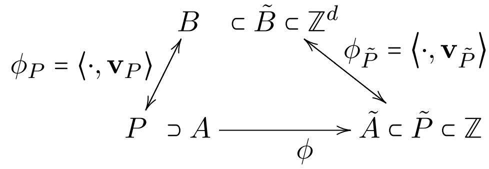


Figure 2.4: Freiman diagram.

Conjecture 1. *Let $A \subset \mathbb{Z}$. Then there exists a Freiman isomorphism to a subset in* $[0, 2^{|A|-2}]$.

However, in general we will have to proceed as follows (deriving a larger number $\tilde{n}$ in the conjecture)

(i) For all $A \subset \mathbb{Z}$ with $|A| \leq s + f$ it holds for the doubling constant $\sigma \leq \alpha = (s + f + 1)/2$.

(ii) By using Theorem 4, a sharp and explicit version of the CHANG-FREIMAN theorem, the set A is contained in a 2−proper progression P with dimension $d \leq \lceil 2\alpha \rceil$ and cardinality $|P| = \prod_{i=1}^{d}(n_i + 1)$, depending only on s and f. Then $A + A \subset 2P := P + P$, where $2P$ is still a proper progression. Hence ϕ_P is a bijective Freiman homomorphism from $2B$ to $2P$, see remark for Proposition 7, and by Lemma 4 a Freiman isomorphism from B to P (respectively ϕ_P^{-1} from P to B).

(iii) Since the box $B \subset \mathbb{Z}^d$ for P has edges $n_i \geq 1$ and by properness $\mathrm{Vol}(B) = |P|$, the largest possible edge n is given by $(n + 1)2^{d-1} = |P|$ assuming all other $d - 1$ edges to be 1, i.e. $n+1 = 2^{1-d}|P|$. Then we will show with Proposition 7 that $B \subset \tilde{B} = [0, n]^d$ is Freiman isomorphic to the proper geometric progression $\tilde{P} = \left\{ \sum b_i(2n + 1)^{i-1} \;\middle|\; \mathbf{b} \in [0, n]^d \right\}$ where the largest element is less than $(2(n + 1))^{d-1}$.

(iv) Hence there exists a Freiman isomorphism $\phi = \phi_{\tilde{P}} \circ \phi_P^{-1}$, since the restrictions are Freiman isomorphic [TV06, p. 221], between A and $\phi(A) \subset [0, (2^{-(d-2)}|P|)^{d-1}]$ where $|P|$ and d are bounded by Theorem 4 with (i) by the worst-case $\alpha = (s + f + 1)/2$ of A.

Chang's Result with Explicit Constant

The main ingredient for our proof is a recently established result of CHANG [Cha02], which gives an upper bound on the cardinality of a proper progression P containing the (unstructured) finite set $A \subset \mathbb{Z}$. We will consider an upper bound for the doubling constant σ of A, which is greater than $5/2$. This is reasonable, since we get for the worst doubling constants $(s+f+1)/2 \geq 5/2$, since $s, f \geq 2$.

Theorem 3 (Chang). *Let $A \subset \mathbb{Z}$ be a finite set with doubling constant σ. Let us set $\alpha = \max\{\sigma, 2.5\}$. Then there exists a d−dimensional proper progression P containing A with*

$$d \leq \lceil \alpha - 1 \rceil \tag{2.70}$$

$$\log \frac{|P|}{|A|} < c\alpha^2 (\log \alpha)^3, \tag{2.71}$$

where $c = 26\bar{c}$ and $\bar{c} = 10^6$.

A refined version of GREEN and TAO [GT06] gives

Theorem 4. *Let $A \subset \mathbb{Z}$ be a finite set and define α as before. Then there exists a 2−proper progression P containing A with*

$$d \leq \lceil 2\alpha \rceil \tag{2.72}$$

$$|P| \leq 2^{\alpha} 10^{c\alpha^2 (\log \alpha)^3}. \tag{2.73}$$

Chang's Theorem follows from a sharp version of the Freiman Theorem.

Theorem 5 (Freiman). *Let $A \subset \mathbb{Z}$ be a finite set with doubling constant σ and set α as before. Then A is contained in a d−dimensional progression, where*

$$d < \bar{c}\alpha^2 (\log \alpha)^2 \tag{2.74}$$

$$\log \frac{l(P)}{|A|} < \bar{c}\alpha^2 (\log \alpha)^2. \tag{2.75}$$

Here, the constant $\bar{c}$ as defined in Theorem 3 is in fact given by a deep result of RUDIN [Rud60] for *dissociated sets*, which could be explicitly given by GREEN in [Gre04]. Here, a dissociated set is a synonym for a Sidon set, see for example [TV06].

Proposition 6. *Let be $n \in \mathbb{N}$, a_k complex numbers, $\Lambda \subset \mathbb{Z}_n$ a dissociated set and $p > 2$, then*

$$\left(\frac{1}{n} \sum_{x \in [n]} \Big| \sum_{k \in \Lambda} a_k e^{2\pi \imath k x / n} \Big|^p \right)^{1/p} \leq 12\sqrt{p} \left(\sum_{k \in \Lambda} |a_k|^2 \right)^{1/2}. \tag{2.76}$$

Proof of Freiman's Theorem. We will repeat the steps in Chang's proof to derive the explicit constant c in (2.73) and (2.71). Here we will refer to equations and assertions in [Cha02] by $[\cdot]$. The first constant comes into play in [Proposition 2.1], on the dimension of a proper progression in $2A - 2A$. The Proof of [Proposition 2.1] is given in [Section 3]. Here, the [Lemma 3.1] established an upper bound of the cardinality of the dissociated set Λ, given by Proposition 6

$$|\Lambda| < 12\rho^{-2} \log(1/\delta) \tag{2.77}$$

for $\frac{1}{100} > \delta > 0$. To see this, we substitute in $[(3.9)]$ by (2.76), since

$$\|g\|_p = \left(\frac{1}{n}\sum_x \Big|\sum_k a_k e^{2\pi \imath xk/n}\Big|^p\right)^{1/p} \leq 12\sqrt{p} \tag{2.78}$$

and the sequence a_k defined in $[(3.5)]$ is bounded by

$$\sum_{k\in\Lambda} |a_k|^2 \leq \frac{\sum_{k\in\Lambda} |\hat{f}(k)|^2}{\sum_{l\in\Lambda} |\hat{f}(l)|^2} = 1. \tag{2.79}$$

The Proof of [Proposition 2.1] continues on page 415. To apply [Lemma 3.2] we choose $\rho^{-2} = 101\alpha$ satisfying

$$100\alpha < \rho^{-2} < 1000\alpha \tag{2.80}$$

and set $\epsilon = 1/10$. Since $|\mathcal{R}| = \delta N$ in [Lemma 3.2] we can by a Theorem 8.9 in [Nat96] (see [(3.1)] and [(3.2)] in [Cha02]) lower bound the constant δ by

$$\delta > \frac{1}{320\alpha^{16}}. \tag{2.81}$$

Now, the hypothesis in [Lemma 3.2] is satisfied and we get for the bound in (2.77)

$$d = |\Lambda| < 12 \cdot 101\alpha \log(320\alpha^{16}) < 1212 \cdot 16\alpha \log(3\alpha/2) \tag{2.82}$$

$$< 1212 \cdot 16\alpha \log(\alpha^{3/2}) < 3 \cdot 10^4 \alpha \log \alpha, \tag{2.83}$$

which is the dimension d of the progression P contained in $2A - 2A$. Using the ϵ and formula [(1.19)] in [Cha02] we get

$$|P| > \frac{|A|}{8(10d^2)^d}. \tag{2.84}$$

With this we can bound t in [(2.17)], which is the iteration to construct the proper progression which contains A by

$$10^t \leq 8\alpha^4(10d^2)^d \Leftrightarrow t \leq \log(8\alpha^4(10d^2)^d) < 5\log(\alpha) + d\log(10d^2) \tag{2.85}$$

$$< 5\log\alpha + 2d\log(10d/3) \tag{2.86}$$

$$< 5\log\alpha + 2(3\cdot 10^4)\alpha\log\alpha \cdot \log(10^5\alpha\log\alpha) \tag{2.87}$$

$$< 5\alpha\log\alpha + (6\cdot 10^4)\alpha(\log\alpha)(\log\alpha^{13}\alpha^2) \tag{2.88}$$

$$< 5\alpha\log\alpha + (90\cdot 10^4)\alpha(\log\alpha)(\log\alpha) \tag{2.89}$$

$$< 91\cdot 10^4\alpha(\log\alpha)^2. \tag{2.90}$$

Then substituting (2.83) and (2.90) in [(2.22)] we get

$$\bar{d} < d + 10\alpha t < 3\cdot 10^4\alpha\log\alpha + 91\cdot 10^4\alpha^2(\log\alpha)^2 < 10^6\alpha^2(\log\alpha)^2. \tag{2.91}$$

Setting $\bar{c} := 10^6$ we get from [(2.23)] for the length

$$l(\bar{P}) \leq 3^{\bar{d}}\alpha^4|A| < 3^{\bar{d}+4\alpha}|A| < 10^{\bar{c}\alpha^2(\log\alpha)^2}|A| \tag{2.92}$$

and hence the result

$$\bar{d} < \bar{c}\alpha^2(\log\alpha)^2 \quad , \quad \log\frac{l(\bar{P})}{|A|} < \bar{c}\alpha^2(\log\alpha)^2. \tag{2.93}$$

□

Now, we are ready to prove Chang's Theorem, using methods developed in [Rus91; Bil99].

Proof of Chang's Theorem. The idea is to start with the (not proper) progression $\bar{P}$ derived in the Freiman Theorem. Then we go via Freiman Isomorphism to progressions with smaller dimension but larger cardinality. This uses the proof of [Theorem 1.2] in [Bil99]. First, Bilu uses symmetrized sets, i.e., the progression

$$\bar{P}(q_1,\ldots,q_{\bar{d}};n_1,\ldots,n_{\bar{d}}) \tag{2.94}$$

has to be symmetrized by the box $B_1 = \Pi_i[-n_i+1, n_i-1]$ and we get for the volume

$$\phi(B_1\cap\mathbb{Z}^{\bar{d}}) \supset A \quad , \quad \mathrm{Vol}(B_1) \leq 2^{\bar{d}}l(\bar{P}). \tag{2.95}$$

Then the Freiman Theorem 5 gives

$$\log\frac{\mathrm{Vol}(B_1)}{|A|} \leq \log 2^{\bar{d}} + \log\frac{l(\bar{P})}{|A|} < \bar{d}+\bar{d} = 2\bar{d} \tag{2.96}$$

Going to a smaller box B_2 with dimension $m_2 \leq m_1 \leq \bar{d}$ we get by following Chang until [(4.12)] with $T = 2$ in [(4.11)]

$$\log\frac{\mathrm{Vol}(B_2)}{|A|} \leq \log(4\bar{d})^{\bar{d}} + \log\frac{\mathrm{Vol}(B_1)}{|A|} \leq \bar{d}\log(4\bar{d}) + 2\bar{d} \leq \bar{d}\log(4\bar{c}\alpha^4) + 2\bar{d} \tag{2.97}$$

$$\leq \bar{d}\log\alpha^{14}\alpha^4 + 2\bar{d} = 18\bar{d}\log\alpha + 2\bar{d}\log\alpha^3 = 24\bar{d}\log\alpha. \tag{2.98}$$

Note, since $\alpha \geq 2.5$ we have $\alpha^3 > 10$ and hence $\log(\alpha^3) > 1$. In the next step, we taking $m' \leq \alpha - 1$ and get by Chang

$$\mathrm{Vol}(B') \leq m_2!\left(\frac{m_2}{2}\right)^{m_2}\mathrm{Vol}(B_2) \tag{2.99}$$

yielding $\log\frac{\mathrm{Vol}(B')}{|A|} \leq 24\bar{d}\log\alpha$. Now we want pass to a parallelepiped. Setting $T = 2\alpha(\lceil\alpha\rceil!)^2$, which gets

$$m'' \leq m' \leq \lceil\alpha - 1\rceil \tag{2.100}$$

$$\mathrm{Vol}(B'') \leq (2m'T)^{m'-m''}\mathrm{Vol}(B') \tag{2.101}$$

with [(4.22)] and the rough estimate $\lceil\alpha\rceil! \le \alpha^{2\alpha+2}$ for any $\alpha > 0$

$$\log\frac{\mathrm{Vol}(B'')}{|A|} \le \log(2m'2\alpha(\lceil\alpha\rceil!)^2)^{m'-m''} + 24\bar{d}\log\alpha \tag{2.102}$$
$$\le \alpha\log(4\alpha^2\lceil\alpha\rceil!^2) + 24\bar{d}\log\alpha \tag{2.103}$$
$$\le 2\alpha\log(2\alpha\lceil\alpha\rceil!) + 24\bar{d}\log\alpha \tag{2.104}$$
$$\le 2\alpha\log(\alpha^2\alpha^{2\alpha+2}) + 24\bar{d}\log\alpha \le 2\alpha\log(\alpha^{4\alpha}) + 24\bar{d}\log\alpha \tag{2.105}$$
$$\le 8\cdot 9\alpha^2(\log\alpha)^3 + 24\bar{d}\log\alpha \le 25\bar{d}\log\alpha. \tag{2.106}$$

In the final step we pass to the parallelepiped and get by [(4.23)]

$$|P| \le \lceil\alpha\rceil!\left(\frac{3}{2}(\lceil\alpha\rceil!)^2\right)^{\alpha}\mathrm{Vol}(B'') \tag{2.107}$$

and hence

$$\log\frac{|P|}{|A|} \le \log(\lceil\alpha\rceil!\cdot\left(\alpha^{4\alpha+5}\right)^{\alpha}) + \log\frac{\mathrm{Vol}(B'')}{|A|} \tag{2.108}$$
$$\le \log(\alpha^{2\alpha+2}\alpha^{6\alpha^2}) + 25\bar{d}\log\alpha \le \log(\alpha^{6\alpha^2+2\alpha^2}) + 25\bar{d}\log\alpha \tag{2.109}$$
$$\le 26\bar{c}\alpha^2(\log\alpha)^3 = c\alpha^2(\log\alpha)^3 \tag{2.110}$$

with

$$c = 266\bar{c} = 266\cdot 10^6. \tag{2.111}$$

□

Remark. Before we proceed, let us stress the fact, that the bounds of Chang are optimal in the scaling of α. But since we are always focused on the worst case, the bound, which could be derived by using RUSZA and BILU (see [Cha02] in [(4.1)]), may be still better. Nevertheless, the bound by Chang will be sufficient for our purpose.

Now, using the result of Tao and Green in Theorem 4 we get for the progression P

$$|P| \le 2^{\alpha}10^{c\alpha^2(\log\alpha)^3}2\alpha \quad , \quad d \le \lceil 2\alpha\rceil, \tag{2.112}$$

containing A, such that the sumset $Q := P + P$ is still a proper progression. Next we need:

Lemma 4. *Let $A, B \subset \mathbb{Z}$ containing 0 and $\phi : A + A \to B$ be an bijective Freiman homomorphism, then $\phi : A \to \phi(A) \subset B$ is an Freiman isomorphism.*

Proof. Since $0 \in A$ we have

$$A \subset A + A \quad , \quad \phi(A) \subset B. \tag{2.113}$$

Now we show, that ϕ acts like a group homomorphism between A and B. For any $a_1, a_2 \in A$ we set $a_1' := a_1 + a_2, a_2' = 0 \in A + A$ and get by the Freiman homomorphism property (2.63)

$$a_1 + a_2 = a_1' + a_2' \Rightarrow \phi(a_1) + \phi(a_2) = \phi(a_1') + \phi(a_2') = \phi(a_1 + a_2) + \phi(0) = \phi(a_1 + a_2)$$
$$\Rightarrow \phi(a_1) + \phi(a_2) = \phi(a_1 + a_2). \tag{2.114}$$

Now, since ϕ is bijective we have $\phi^{-1} : B \to A + A$ and with (2.114) it holds for any $a_1, a_2 \in A$

$$\phi^{-1}(\phi(a_1) + \phi(a_2)) = \phi^{-1}(\phi(a_1 + a_2)) = a_1 + a_2 = \phi^{-1}(\phi(a_1)) + \phi^{-1}(\phi(a_2)) \tag{2.115}$$

and hence $\phi^{-1} : \phi(A) \to A$ acts also like a homomorphism and we get for $b_1, b_2, b_1', b_2' \in \phi(A)$

$$b_1 + b_2 = b_1' + b_2' \Rightarrow \phi^{-1}(b_1 + b_2) = \phi^{-1}(b_1' + b_2') \tag{2.116}$$
$$(2.115) \to \Rightarrow \phi^{-1}(b_1) + \phi^{-1}(b_2) = \phi^{-1}(b_1') + \phi^{-1}(b_2'), \tag{2.117}$$

which is the definition of a Freiman homomorphism. □

Next, we need to prove an exercise in [TV06, p.221]

Proposition 7. *Let $n, d \in \mathbb{N}$ and define $m = 2n + 1, \mathbf{v} = (m^0, m^1, \ldots, m^{d-1}), B = [0, n]^d$, then*

$$\langle \cdot, \mathbf{v} \rangle : B \to P = \langle B, \mathbf{v} \rangle \subset \mathbb{Z} \tag{2.118}$$

is a Freiman isomorphism of order 2 between B and P.

Proof. Since we know, that $\phi = \langle \cdot, \mathbf{v} \rangle$ is an group homomorphism [TV06, p.114] between the ambient group $\mathbb{Z}^d$ and $\mathbb{Z}$, we get for any $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_1', \mathbf{b}_2' \in 2B$

$$\mathbf{b}_1 + \mathbf{b}_2 = \mathbf{b}_1' + \mathbf{b}_2' \Rightarrow \langle \mathbf{b}_1 + \mathbf{b}_2, \mathbf{v} \rangle = \langle \mathbf{b}_1' + \mathbf{b}_2', \mathbf{v} \rangle \tag{2.119}$$
$$\text{group homomorphism} \to \Leftrightarrow \langle \mathbf{b}_1, \mathbf{v} \rangle + \langle \mathbf{b}_2, \mathbf{v} \rangle = \langle \mathbf{b}_1', \mathbf{v} \rangle + \langle \mathbf{b}_2', \mathbf{v} \rangle, \tag{2.120}$$

which shows that $\langle \cdot, \mathbf{v} \rangle$ is a surjective Freiman homomorphism (2.62) from $2B$ to $\langle 2B, \mathbf{v} \rangle$. Hence, if we can show injectivity on $2B = [0, 2n]^d$ we have bijectivity. Since we have for each $\mathbf{b} \in 2B$

$$\langle \mathbf{b}, \mathbf{v} \rangle = \sum_{i=1}^{d} b_i v_i = \sum_{i=1}^{d} b_i m^{i-1}, \tag{2.121}$$

we only need to show that each $\mathbf{b} \in 2B$ creates a different number, i.e., we have to show that for all $i \in [1, d]$ it holds

$$\forall b_i \in [0, 2n] \colon b_i m^{i-1} < m^i, \tag{2.122}$$

which is true, since this is equivalent to

$$\forall b_i \in [0, 2n] \colon b_i < m^{i-(i-1)} = 2n + 1. \tag{2.123}$$

By Lemma 4 we have therefore a Freiman isomorphism between B and $P = \phi(B)$. □

Remark. In fact, this holds for any $2-$proper progression, since this equivalent with injectivity for $\langle \cdot, \mathbf{v} \rangle$ from $2B$ to $2P$.

Considering the last tow steps (ii) and (iii) we get with (2.112) for

$$\tilde{n} \leq (2 \cdot 2^{-(d-1)}|P|)^{d-1} \leq (2^{-(d-2)}|P|)^{d-1} \leq (10^{c\alpha^2(\log\alpha)^3})^{2\alpha-1} = (10^{c\alpha^2(\log\alpha)^3})^{s+f} \quad (2.124)$$
$$(2.125)$$

inserting $\alpha = (s+f+1)/2 \leq s+f$ we get with $c \leq 10^8$ by (2.111) finally our estimate

$$\tilde{n}(s,f) \leq 10^{c(s+f)^3 \log^3(s+f)} \quad (2.126)$$

■

2.2.1 Algorithmic implementation

The problem in (2.34) can be approximated by discretization of $\mathbf{y}$ in $D = \{0, \sqrt{1/d}, \dots, \sqrt{d/d}\}$ with $d \in \mathbb{N}$. Hence $D^{\tilde{n}}$ is a $\tilde{n}$-dimensional uniform grid of the cube. For each fixed $\mathbf{y}_d \in D^{\tilde{n}}$ with $\|\mathbf{y}_d\| = 1$ we get $\mathbf{B}_{\mathbf{y}_d}$ and obtain the approximate solution

$$\alpha_{\tilde{n}}^{d} = \min_{\mathbf{y}_d \in D^{\tilde{n}}, \|\mathbf{y}_d\|=1} \lambda(\mathbf{B}_{\mathbf{y}_d}), \tag{2.127}$$

which is an $(1 - 1/d)$-approximation solution to $\alpha_{\tilde{n}}$, see [LNQY09], i.e.,

$$\alpha_{\tilde{n}} \geq \alpha_{\tilde{n}}^{\text{low},d} := \alpha_{\tilde{n}}^{d} - \frac{\tilde{n}}{d}. \tag{2.128}$$

The price, is the size of the cube grid: the number of possible grid points $\mathbf{y}_d$ are of the order $|D|^{\tilde{n}} = (d+1)^{\tilde{n}} < \tilde{n}^{\tilde{n}}$ and hence sub-exponential. We could establish in Fig. 2.5 with MATLAB global lower bounds, drawn as dotted green lines, for $\alpha_{\tilde{n}}$. For $\tilde{n} > 6$ the computational time was too large to establish a global lower bound.

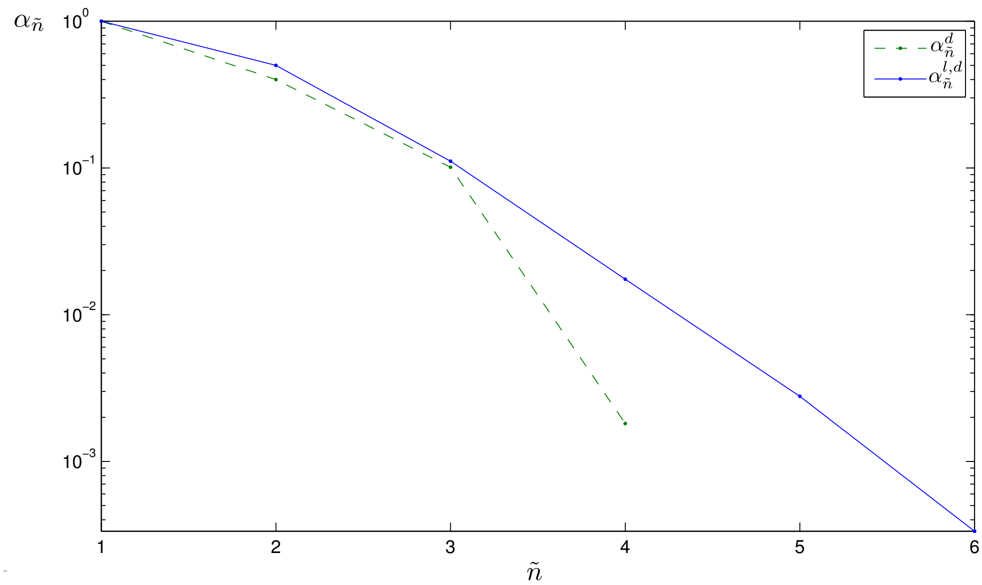


Figure 2.5: Approximation results of the lower bound $\alpha_{\tilde{n}}$.

2.3 Circular Convolution and Phase Retrieval

The part in this section was published in [WJ13b].

Recovering a signal from intensity (magnitude) measurements is known as the *phase retrieval problem*. This problem has a long history beginning in the 70's by GERCHBERG and SAXTON [GS72] and later by FIENUP [Fie78], who gave explicit reconstruction algorithms for the phase from magnitude Fourier measurements. Since the magnitude of a linear measurement can not distinguish between numbers of unit modulus, stability and injectivity for such measurements can only hold up to a global phase respectively sign, i.e., up to a factor $e^{i\omega}$ respectively ± 1. One of the challenging tasks in phase retrieval is to determine the necessary and sufficient number of the magnitude of linear measurements for stability or injectivity. For example, CANDES et.al. [CSV12] have shown stable reconstruction of any $n-$dimensional complex-valued signal from the magnitude of $\mathcal{O}(n \log n)$ linear Gaussian-random measurements. A more principal result from BALAN et al. in [BCE06] shows that a generic frame exists with injectivity at $4n - 2$ measurements. Moreover, they could give a fast reconstruction algorithm in [BBCE07]. In a recent result [BCMN13], BANDEIRA et al. conjecture that $4n - 4$ measurements are necessary for injectivity. However, a practical construction and implementation of these measurements seem to be rather hard, but serve as a theoretical bound.

More recently, non-linear or interference–based approaches are considered to provide unique phase reconstruction. For example, WANG [Wan13] presented a method where interference with a known signal $\mathbf{y} \in \mathbb{C}^n$ helps to recover a signal $\mathbf{x} \in \mathbb{C}^n$ up to a global sign from only $3n$ Fourier measurements $|\mathbf{F}(\mathbf{x} + \omega\mathbf{y})|^2$ where $\omega \in \mathbb{C}$ is a root of unity. For *real* k–sparse signals, ELDAR and MENDELSON [EM12] established a stable recovery from $\mathcal{O}(k \log(en/k))$ sub-Gaussian random measurements. A very recent result [EFS13] from EHLER, FORNASIER and SIGL extended this to the complex case, allowing also a stable recovery with exponentially high probability by providing an explicit reconstruction algorithm. LU and VETTERLI also use sparsity for spectral factorization of real valued impulse responses [LV11]. Moreover, they give a reconstruction algorithm from the autocorrelation. A recent result by WANG and XU [WX13] states injectivity for $k-$sparse complex-valued signals from $4k - 2$ generic measurements as long as $k < n$. Unfortunately, so far there doesn't exists a constructive or deterministic frame providing a recovery or even stable recovery.

Here, we will show a concrete measurement procedure allowing stable recovery of any vector $\mathbf{x} \in \mathbb{C}^n$ with $x_0 \in \mathbb{R}$ up to global sign from magnitudes of $4n - 3$ linear measurements. The measurements can be implemented as linear mappings on, for example, $(\mathrm{Re}(\mathbf{x}), \mathrm{Im}(\mathbf{x}))$ or $(\mathbf{x}, \bar{\mathbf{x}})$. We want to stress the fact, that our measurements are *not complex–linear*, since we perform a non-linear symmetrization on the signal to obtain a symmetric auto-convolution, allowing magnitude measurements from $4n - 3$ linear Fourier measurements. However, this will have implications on certain (compressive) signal processing tasks since such type of measurements occur prior to I/Q–down conversion into a complex baseband model. To prove stability for magnitude Fourier measurements, we will use our result in [WJ12b] for the $(s, f)-$sparse zero-padded circular convolution. In view of sparsity, zero padding can also be seen as a particular structured sparse signal subclass in $\mathbb{C}^{4n-3}$.

Before we show a stability property for magnitude Fourier measurements, we will formulate the RNMP for the sparse zero-padded circular convolution Theorem 2 in finite dimensions.

Corrolary 2. *Let $s, f, n \in \mathbb{N}$ with $s \leq f \leq n$. Then there exists a constant $\alpha_{n'(s,f,n)} > 0$ with $n'(s,f,n) := \min\{\tilde{n}(s,f), n\}$, such that for all $\mathbf{x} \in \Sigma_s^n, \mathbf{y} \in \Sigma_f^n$ zero padded by $n-1$ zeros $\mathbf{0}_{n-1}$ it holds*

$$\alpha_{n'(s,f,n)} \|\mathbf{x}\| \|\mathbf{y}\| \leq \left\| (\mathbf{x}, \mathbf{0}_{n-1})^T \circledast (\mathbf{y}, \mathbf{0}_{n-1})^T \right\| \leq \sqrt{s} \|\mathbf{x}\| \|\mathbf{y}\| . \tag{2.129}$$

Proof. The proof is essentially the same as for Theorem 2 by setting $\tilde{\mathbf{x}} = (\mathbf{x}, \mathbf{0}_{n-1}) \in \mathbb{K}^{2n-1}$. Since the dimension n is fixed in the corollary, we can set the constant $n'(s,f)$ in Theorem 2 to

$$n'(s,f,n) := \min\{\tilde{n}(s,f), n\}. \tag{2.130}$$

□

Remark. As already mentioned in the proof of Theorem 2, the zero padding is necessary to obtain a lower bound strictly greater than zero.

However, we conjecture, that a bound α is also present for the circular convolution on prime dimensions. See the principal result of the DONOHO-STARK inequality in Appendix A.1.

Moreover, setting $n' = 2n - 1$ and using the definition (1.55) for the circular convolution it holds for any $\mathbf{x}, \mathbf{y} \in \mathbb{K}^{n'}$

$$\mathbf{x} \circledast \mathbf{y} = \mathbf{Y}^T \mathbf{x} = \sum_{i=0}^{n'} x_i \mathbf{S}_{n'}^{-i} \mathbf{y} = \sum_{i=0 \oplus 1}^{n' \oplus 1} x_{i \ominus 1} \mathbf{S}_{n'}^{-(i \ominus 1)} \mathbf{y} = \sum_{i=0}^{2n-2} (\mathbf{S}_{n'} \mathbf{x})_i \mathbf{S}_{n'}^{-i} \mathbf{S}_{n'} \mathbf{y} \tag{2.131}$$

$$= \mathbf{S}_{n'} \mathbf{x} \circledast \mathbf{S}_{n'} \mathbf{y}. \tag{2.132}$$

We can repeat this arbitrarily often and hence whenever the zeros are contained in a cyclic block of length $n-1$ the inequality (2.129) holds.

2.3.1 Recovery from Symmetrized Magnitude Fourier Measurements

We will ignore sparsity and assume $s = f = n$ in Corollary 2, i.e., $n'(s,f,n) = n$, which yields the RNMP for the zero padded circular convolution in $\mathbb{K}^n$. In fact, the zero padding yields equivalence between circular and discrete convolution. But before we do so, we will symmetrize the signal $\mathbf{x} \in \mathbb{K}^n$ with the *symmetrization map* $\mathcal{S}^\circ$ as

$$\mathcal{S}^\circ : \mathbb{K}^n \to \mathbb{K}^{2n-1} \quad , \quad \mathbf{x} \mapsto \mathcal{S}^\circ(\mathbf{x}) := \begin{pmatrix} \mathbf{x} \\ \mathbf{x}_-^\circ \end{pmatrix}, \tag{2.133}$$

where the time reversal operation is given by $\mathbf{x}_- := \mathbf{S\Gamma x} = (x_{n-1}, x_{n-2}, \ldots, x_1, x_0)^T$ and the time-reversal without the first coefficient $\mathbf{\Gamma}^\circ \mathbf{x} = \mathbf{x}_-^\circ := (x_{n-1}, x_{n-2}, \ldots, x_1)^T$. See definition (1.59) and (1.54). Let us stress the fact, that only for $\mathbb{K} = \mathbb{R}$ the symmetrization map is linear. To be more precise, $\mathcal{S}^\circ$ is a homomorphism between $(\mathbb{K}^n, +)$ and $(\mathbb{K}^{2n-1}, +)$ but not a homomorphism between $(\mathbb{C}^n, \cdot)$ and $(\mathbb{C}^{2n-1}, \cdot)$. However, if we **demand** $x_0 = \overline{x_0}$ (denoted by $\mathbb{K}_0^n$), then we get the following symmetry (involution invariance),

$$\mathcal{S}^\circ(\mathbf{x})^* = \mathbf{\Gamma}_{2n-1}\overline{\begin{pmatrix}\mathbf{x}\\ \overline{\mathbf{x}_-^\circ}\end{pmatrix}} = \mathbf{\Gamma}_{2n-1}\begin{pmatrix}\overline{\mathbf{x}}\\ \mathbf{x}_-^\circ\end{pmatrix} = \begin{pmatrix}\mathbf{x}\\ \overline{\mathbf{x}_-^\circ}\end{pmatrix} = \mathcal{S}^\circ(\mathbf{x}) \quad \text{for} \quad \mathbf{x} \in \mathbb{K}_0^n. \tag{2.134}$$

Here, the (algebra) involution is defined as $\mathbf{x}^* = \overline{\mathbf{\Gamma x}}$. Hence, we obtain equality between the circular convolution and correlation for symmetrized vectors in $\mathbb{K}^{2n-1}$, i.e.,

$$\mathcal{C}(\mathbf{x}, \mathbf{y}) := \mathcal{S}^\circ(\mathbf{x}) \circledast \mathcal{S}^\circ(\mathbf{y}) = \mathcal{S}^\circ(\mathbf{x}) \circledast \mathcal{S}^\circ(\mathbf{y})^* = \mathcal{S}^\circ(\mathbf{x}) \circledast \mathcal{S}^\circ(\mathbf{y}) \quad \text{for} \quad \mathbf{x}, \mathbf{y} \in \mathbb{K}_0^n. \tag{2.135}$$

Note, that the condition (2.134) is also **necessary** for obtaining equality between circular correlation and convolution, even if $\mathbf{x} = \mathbf{y}$ in (2.135). Moreover, $\mathcal{C}$ is a *symmetric homomorphism* on $(\mathbb{K}_0^n, +)$, i.e., $\mathcal{C}(\mathbf{x}, \mathbf{y}) = \mathcal{C}(\mathbf{y}, \mathbf{x})$. Let us define the map

$$\mathcal{A}^\circ : \mathbb{K}_0^n \to \mathbb{K}^{2n-1} \quad , \quad \mathbf{x} \mapsto \mathcal{A}^\circ(\mathbf{x}) := \mathcal{C}(\mathbf{x}, \mathbf{x}) \quad \text{for} \quad \mathbf{x} \in \mathbb{K}_0^n, \tag{2.136}$$

which we call the *symmetrized circular auto-convolution* of $\mathbf{x}$. Let us assume $x_{n-1} \neq 0$, then the impulse response $\mathbf{h} := (h_0, \ldots, h_{2n-2})^T = \mathbf{S}_{2n-1}^{n-1}\mathcal{S}^\circ(\mathbf{x}) = (\overline{x_{n-1}}, \ldots, \overline{x_1}, x_0, x_1, \ldots, x_{n-1})^T$, defines a *linear-phase filter* $H(z) = \sum_{k=0}^{2n-2} h_k z^{-k}$ for $z \in \mathbb{C}$ since $\overline{h_0} = h_{2n-2} \neq 0$ and

$$\overline{h_k} = h_{2n-2-k} \quad \text{for} \quad k \in [2n-1]. \tag{2.137}$$

The impulse response or filter is then called *Hermitian* or *conjugate symmetric* of order $2n-2$, see for example [Vai93, Cha.2]. Hence, with (2.132) we have[9] for $\mathbf{x} \in \mathbb{K}_0^n$

$$\mathcal{A}^\circ(\mathbf{x}) = \mathcal{S}^\circ(\mathbf{x}) \circledast \mathcal{S}^\circ(\mathbf{x}) = \mathbf{S}_{2n-1}^{n-1}\mathcal{S}^\circ(\mathbf{x}) \circledast \mathbf{S}_{2n-1}^{n-1}\mathcal{S}^\circ(\mathbf{x}) = \mathbf{h} \circledast \mathbf{h} \tag{2.138}$$

which is the circular auto-convolution of a linear-phase filter. For any such map it holds due to (2.27) the relation:

$$\mathcal{A}^\circ(\mathbf{x}_1) - \mathcal{A}^\circ(\mathbf{x}_2) = \mathcal{C}(\mathbf{x}_1 - \mathbf{x}_2, \mathbf{x}_1 + \mathbf{x}_2) \tag{2.139}$$

and with (2.135)

$$\mathcal{A}^\circ(\mathbf{x}_1) - \mathcal{A}^\circ(\mathbf{x}_2) = \mathcal{S}^\circ(\mathbf{x}_1 - \mathbf{x}_2) \circledast \mathcal{S}^\circ(\mathbf{x}_1 + \mathbf{x}_2) \quad \text{for} \quad \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{K}_0^n. \tag{2.140}$$

To apply Corollary 2 we define the *zero-padded symmetrization*:

$$\mathcal{S}_z^\circ : \mathbb{K}^n \to \mathbb{K}^{4n-3} \quad , \quad \mathbf{x} \mapsto \mathcal{S}_z^\circ(\mathbf{x}) := \mathcal{S}^\circ\begin{pmatrix}\mathbf{x}\\ \mathbf{0}_{n-1}\end{pmatrix}, \tag{2.141}$$

[9] Note, that $\mathbf{S}_{2n-1}^{n-1}$ centers the impulse response such that it becomes a causal FIR filter.

then $\mathcal{S}_z^\circ(\mathbf{x}) \in \mathbb{K}_0^{4n-3}$ for every $\mathbf{x} \in \mathbb{K}^n$. Moreover we have

$$\mathbf{S}_{\tilde{n}}^{n-1}\mathcal{S}_z^\circ(\mathbf{x}) = \mathbf{S}_{\tilde{n}}^{n-1}\mathcal{S}^\circ\begin{pmatrix}\mathbf{x}\\ \mathbf{0}_{n-1}\end{pmatrix} = \mathbf{S}_{\tilde{n}}^{n-1}\begin{pmatrix}\mathbf{x}\\ \mathbf{0}_{2n-2}\\ \overline{\mathbf{x}_-^\circ}\end{pmatrix} = \begin{pmatrix}\overline{\mathbf{x}_-^\circ}\\ \mathbf{x}\\ \mathbf{0}_{2n-2}\end{pmatrix} = \begin{pmatrix}\mathbf{h}\\ \mathbf{0}_{2n-2}\end{pmatrix}. \tag{2.142}$$

By (2.132) we have therefore

$$\mathbf{S}_{\tilde{n}}^{n-1}\mathcal{S}_z^\circ(\mathbf{x}) \circledast \mathbf{S}_{\tilde{n}}^{n-1}\mathcal{S}_z^\circ(\mathbf{x}) = \mathcal{S}_z^\circ(\mathbf{x}) \circledast \mathcal{S}_z^\circ(\mathbf{x}) =: \mathcal{A}_z^\circ(\mathbf{x}). \tag{2.143}$$

Theorem 8. *Let $n \in \mathbb{N}$, then $\tilde{n} = 4n-3$ absolute-square Fourier measurements of zero padded symmetrized vectors in $\mathbb{C}^{\tilde{n}}$, given by* (2.141)*, are stable up to a global sign for $\mathbf{x} \in \mathbb{C}_0^n$, i.e. for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{C}_0^n$ it holds*

$$\left\| |\mathbf{F}_{\tilde{n}}\mathcal{S}_z^\circ(\mathbf{x}_1)|^2 - |\mathbf{F}_{\tilde{n}}\mathcal{S}_z^\circ(\mathbf{x}_2)|^2 \right\|^2 \geq c_{\tilde{n}}^2\left(\|\mathbf{x}_1-\mathbf{x}_2\|^2 + \|\mathbf{x}_1^\circ-\mathbf{x}_2^\circ\|^2\right)\left(\|\mathbf{x}_1+\mathbf{x}_2\|^2 + \|\mathbf{x}_1^\circ+\mathbf{x}_2^\circ\|^2\right) \tag{2.144}$$

with $c_{\tilde{n}} = \alpha_{\tilde{n}}/\sqrt{\tilde{n}} > 0$.

Remark. To see the stability up to a global sign, we can lower bound the RHS in (2.144) and get

$$\left\| |\mathbf{F}_{\tilde{n}}\mathcal{S}_z^\circ(\mathbf{x}_1)|^2 - |\mathbf{F}_{\tilde{n}}\mathcal{S}_z^\circ(\mathbf{x}_2)|^2 \right\| \geq c_n \|\mathbf{x}_1 - \mathbf{x}_2\| \, \|\mathbf{x}_1 + \mathbf{x}_2\|. \tag{2.145}$$

Proof. We will first show, that the inverse Fourier transformation of the auto-correlation is indeed a magnitude Fourier measurement by using (2.134) and the definition (1.68):

$$\mathbf{F}^*(\mathbf{x} \circledast \mathbf{x}) = \mathbf{F}^*(\mathbf{x} \circledast \mathbf{x}) = \sqrt{n}\mathbf{\Gamma}(\mathbf{F}\mathbf{\Gamma}\mathbf{x} \odot \mathbf{F}\overline{\mathbf{x}}) = \sqrt{n}\mathbf{\Gamma}\mathbf{\Gamma}(\mathbf{F}\mathbf{x} \odot \mathbf{\Gamma}\mathbf{F}\overline{\mathbf{x}}) = \sqrt{n}|\mathbf{F}\mathbf{x}|^2. \tag{2.146}$$

Hence we get for the Fourier transform of the symmetrized auto-convolution (2.136)

$$\mathbf{F}_{2n-1}^*\mathcal{A}^\circ(\mathbf{x}) = \mathbf{F}_{2n-1}^*(\mathcal{S}^\circ(\mathbf{x}) \circledast \mathcal{S}^\circ(\mathbf{x})) = \sqrt{2n-1}|\mathbf{F}_{2n-1}\mathcal{S}^\circ(\mathbf{x})|^2 \quad \text{for} \quad \mathbf{x} \in \mathbb{C}_0^n. \tag{2.147}$$

Similar, we obtain for the zero padded version (2.143)

$$\mathbf{F}_{4n-3}^*\mathcal{A}_z^\circ(\mathbf{x}) = \sqrt{4n-3}|\mathbf{F}_{4n-3}\mathcal{S}_z^\circ(\mathbf{x})|^2 \quad \text{for} \quad \mathbf{x} \in \mathbb{C}_0^n. \tag{2.148}$$

Putting everything together, we get with (2.135),(2.141), (2.143) and Corollary 2 by setting $\tilde{n} = 4n-3$

$$\left\| |\mathbf{F}_{\tilde{n}}\mathcal{S}_z^\circ(\mathbf{x}_1)|^2 - |\mathbf{F}_{\tilde{n}}\mathcal{S}_z^\circ(\mathbf{x}_2)|^2 \right\|^2 = \frac{1}{\tilde{n}}\left\|\mathbf{F}_{\tilde{n}}^*(\mathcal{A}_z^\circ(\mathbf{x}_1) - \mathcal{A}_z^\circ(\mathbf{x}_2))\right\|^2 \tag{2.149}$$

$$\mathbf{F}_{\tilde{n}}^* \text{ is unitary} \rightarrow = \frac{1}{\tilde{n}}\left\|\mathcal{A}_z^\circ(\mathbf{x}_1) - \mathcal{A}_z^\circ(\mathbf{x}_2)\right\|^2 \tag{2.150}$$

$$(2.140) \rightarrow = \frac{1}{\tilde{n}}\left\|\mathcal{S}_z^\circ(\mathbf{x}_1-\mathbf{x}_2) \circledast \mathcal{S}_z^\circ(\mathbf{x}_1+\mathbf{x}_2)\right\|^2$$

$$(2.143) \rightarrow = \frac{1}{\tilde{n}}\left\|\mathbf{S}_{\tilde{n}}^{n-1}\mathcal{S}_z^\circ(\mathbf{x}_1-\mathbf{x}_2) \circledast \mathbf{S}_{\tilde{n}}^{n-1}\mathcal{S}_z^\circ(\mathbf{x}_1+\mathbf{x}_2)\right\|^2 \tag{2.151}$$

$$\mathbf{S}_{\tilde{n}}^i \text{ is unitary, Corollary 2} \rightarrow \geq \frac{\alpha_{\tilde{n}}^2}{\tilde{n}}\left\|\mathbf{S}_{\tilde{n}}^{n-1}\mathcal{S}_z^\circ(\mathbf{x}_1-\mathbf{x}_2) \circledast \mathbf{S}_{\tilde{n}}^{n-1}\mathcal{S}_z^\circ(\mathbf{x}_1+\mathbf{x}_2)\right\|^2 \tag{2.152}$$

$$= \frac{\alpha_{\tilde{n}}^2}{\tilde{n}}\left(\|\mathbf{x}_1-\mathbf{x}_2\|^2 + \|\mathbf{x}_1^\circ-\mathbf{x}_2^\circ\|^2\right)\left(\|\mathbf{x}_1+\mathbf{x}_2\|^2 + \|\mathbf{x}_1^\circ+\mathbf{x}_2^\circ\|^2\right) \tag{2.153}$$

□

If we consider $\mathbb{K} = \mathbb{R}$, then (2.144) is equivalent with a *stable embedding* of $T = \mathbb{R}^n$ in $\mathbb{R}^{4n-3}$ by $A = |\mathbf{F}_{\tilde{n}}\mathcal{S}_z^{\circ}(\cdot)|^2$, see also [EM12] where ELDAR and MENDELSON used the ℓ^1−norm on the left side. Now, since $\mathcal{S}^{\circ}$ is not a linear map in $\mathbb{C}^n$ we can indeed distinguish the complex phase by the Fourier measurements. Hence, a suitable symmetrization design provides injectivity for magnitude Fourier measurements.

To get rid of the odd definition $\mathbb{C}_0^n$ one can symmetrize $\mathbf{x} \in \mathbb{C}^n$ by

$$\mathcal{S}_z(\mathbf{x}) := \begin{pmatrix} \mathbf{0}_n \\ \mathbf{x} \\ \overline{\mathbf{x}_-} \\ \mathbf{0}_{n-1} \end{pmatrix} \in \mathbb{C}^{4n-1} \tag{2.154}$$

satisfying $\mathbf{\Gamma}_{4n-1}\overline{\mathcal{S}_z(\mathbf{x})} = \mathcal{S}_z(\mathbf{x}) \in \mathbb{C}^{4n-1}$. The price are two additional dimensions for embedding all $\mathbf{x} \in \mathbb{C}^n$ stably. Hence we have:

Corrolary 3. *Let $n \in \mathbb{N}$, then $\tilde{n} = 4n - 1$ absolute-square Fourier measurements of zero padded and symmetrized vectors given by* (2.154) *are stable up to a global sign for $\mathbf{x} \in \mathbb{C}^n$, i.e. for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{C}^n$ it holds*

$$\left\| |\mathbf{F}_{\tilde{n}}\mathcal{S}_z(\mathbf{x}_1)|^2 - |\mathbf{F}_{\tilde{n}}\mathcal{S}_z(\mathbf{x}_2)|^2 \right\| \geq 2\tilde{c}_n \|\mathbf{x}_1 - \mathbf{x}_2\| \|\mathbf{x}_1 + \mathbf{x}_2\| \tag{2.155}$$

with $\tilde{c}_n = \alpha_{\tilde{n}}/\sqrt{\tilde{n}} > 0$.

Proof. The same steps as in the proof of Theorem 8. □

An extension to sparse signals as in [WJ12a] is difficult to apply, since randomly chosen Fourier samples do not provide a measure concentration.

2.3.2 Phase Retrieval in the Real Case

If we restrict Theorem 8 to real-valued vectors, the symmetrization map $\mathcal{S}^{\circ}$ in (2.133) becomes linear. Hence, we have a stable reconstruction up to a global sign from the magnitude of $4n-3$ linear measurements, given by

$$\mathbf{A} = \mathbf{F}_{4n-3} \begin{pmatrix} \mathbb{1}_{n\times n} \\ \mathbf{0}_{n-1\times n} \\ \mathbf{0}_{n-1\times n} \\ \mathbf{\Gamma}^{\circ}_{n-1\times n} \end{pmatrix}, \tag{2.156}$$

where $\mathbf{\Gamma}^{\circ}$ denotes the matrix $\mathbf{\Gamma}$ without the first row.

Interest in phase retrieval increased in the recent years due to applications in X-ray diffraction. Moreover, the understanding of sparse signal restriction by probabilistic methods, allowing compressed sensing techniques, has pushed the phase retrieval problem further to sparse signals [CSV12],[BM13], [RCLV13],[OYDS12],[SBE13]. We will develop in the next chapter a general approach for a stable reconstruction of bilinear images with compressed sensing methods. Although, our machinery relies heavily on the RNMP, which is difficult to show on linear spaces, we will give in the next chapter a principle result of possible sampling rates for a stable reconstruction.

3 Compressive Sampling on Sparse Multiplications

Compressed Sensing raised in the last decade to a powerful tool in signal processing by merging probabilistic methods and Fourier analysis. For an excellent survey to this field the reader is referred to [FR11],[EK12] and [DDEK12].

In a recent work [HB11], HEDGE and BARANIUK considered (s, f)-sparse circular convolutions in $\mathbb{R}^n$ and formulated a *restricted isometry property* (RIP) for the set of all differences in the union of the image of all (s, f)-sparse circular convolutions. It turns out that their proof approach leads to difficult mathematical problems, which according to the authors' knowledge are still unsolved. Since the proof in [BDDW08, Lemma 5.1] relies on a linear structure, which may not be present for the image of bilinear maps, more strict conditions on $\mathcal{B}$ and the input sets X and Y are needed to control the norm of the output set. However, we will show in Lemma 5, that the RNMP of $\mathcal{B}$ on $X \times Y$ is a sufficient condition to obtain the RIP on the image set from $\mathcal{O}((s+f)\log n)$ random measurements. Furthermore, the authors conjecture that the RNMP and convexity is also a necessary condition for establishing the RIP from measurement with additive scaling in the sparsity, instead of multiplicative. This result implies an universal stable compressive sampling on the image set from any unknown but fixed sparse channel state in Corollary 4. Nevertheless, if the channel state itself vary in time, such a universal compressive measurement does not imply a stable reconstruction of $\mathbf{x}$ and $\mathbf{y}$, since the image set is then not a linear set.

Moreover, our developed framework is expendable to all bilinear operations having the RNMP on convex cones in an arbitrary basis. This enables compressed sensing on "sparse" output sets, which can not be written as a finite union of subspaces, and leads to *generalized structured sparsity models* [BW09; DE11].

3.1 RIP for Multiplications on Subspaces

Our main result provides a generalized compressed sensing framework by a stable random embedding of certain (s, f)-sparse signal models which can not be formulated as a k-sparse signal model Σ_k. Moreover, our model can be related to matrix recovery of $s \times f$ rank-1 matrices if $\mathcal{B}$ is norm multiplicative ($\alpha = \beta$) on $X \times Y$. Since differences of rank-1 matrices are not rank-1 but rank-2, embedding require additional structure. In the first step for a successful embedding result, we will proof the restricted isometry property (RIP) for random linear maps if $(\mathcal{B}, X, Y)$ obeys the RNMP.

Lemma 5 (RIP on Bilinear Image)**.** *Let be $s, f, n, m \in \mathbb{N}$ with $s \le f \le m \le n$, $\delta \in (0,1)$ and $X, Y \subset \mathbb{R}^n$ are s respectively f dimensional convex cones. If the bilinear map $\mathcal{B}: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ has the global RNMP on $X \times Y$ with bounds α and β, then a realization of a sub-Gaussian matrix $\mathbf{\Phi}: \mathbb{R}^n \to \mathbb{R}^m$ with $[\mathbf{\Phi}]_{ij} \sim \mathcal{N}(0, 1/m)$ fulfills for every $\mathbf{z} \in \mathcal{B}(X, Y)$*

$$(1-\delta)\|\mathbf{z}\| \le \|\mathbf{\Phi z}\| \le (1+\delta)\|\mathbf{z}\| \tag{3.1}$$

with probability

$$\ge 1 - 2N_{\delta/d}(X^1)N_{\delta/d}(X^1)e^{-c_0(\delta)m} \tag{3.2}$$

and constants

$$d = d(\alpha,\beta) := \begin{cases} 7\frac{\beta}{\alpha}(2+\sqrt{\alpha}) & , \quad \alpha \ne \beta \\ 12 & , \quad \alpha = \beta \end{cases} \tag{3.3}$$

$$c_0(\delta) := (3\delta^2 - \delta^3)/48 \tag{3.4}$$

$$\epsilon := \delta/d. \tag{3.5}$$

Remark. Here $N_\epsilon(X^1) := N(X^1, X^\epsilon) := \min\{n \mid \exists \{\mathbf{p}_i\}_{i=1}^n : X^1 \subset \bigcup_i (X^\epsilon + \mathbf{p}_i)\}$ denotes the covering number of X^1 by the covering sets $\{X^\epsilon + \mathbf{p}_i\}$. The determination of the covering numbers is a Banach geometrical problem and well studied for various precompact subsets of Banach-spaces, see for example [Pis89], [Ver11]. In fact, it is even not necessary to define a norm or even a metric on X and Y to use the covering arguments. Note, precompactness of a set is already given by the existence of a finite number of covering sets.

Proof. The main part follows a technique in [BDDW08], where BARANIUK et al. considered a linear subspace Z of $\mathbb{R}^n$ with a $\delta/4$-net $\mathcal{R}$ for $Z^{1,1}$. By the *measure concentration* of Gaussian matrices $\mathbf{\Phi}$ one gets for every $\mathbf{r} \in \mathcal{R}$ and any $\delta \in (0,1)$ the inequality[1]:

$$|\,\|\mathbf{\Phi r}\| - \|\mathbf{r}\|\,| \le \frac{\delta}{2}\|\mathbf{r}\| \tag{3.6}$$

with probability

$$> 1 - 2e^{-c_0(\delta)m}. \tag{3.7}$$

Both, the constant δ and the dimension k of Z determine the covering number $N_\delta(Z^1) = (3/\delta)^k$, which is the cardinality of the net $\mathcal{R}$ and scale the exponential term in (3.7). Since $Z = \mathcal{B}(X, Y)$ is in general not a linear space one would have to embed Z in a linear subspace, but this would imply by the tensor calculus in the worst case a space of dimension sf. Hence we need to find a "better" covering of $Z^{1,1}$. Such a covering can be achieved in an **additive way** if we can cover

[1] Obviously, this holds for every $\mathbf{r} \in \mathbb{R}^N$ and the inequality (3.6) is equivalent to $\left|\|\mathbf{\Phi r}\|^2 - \|\mathbf{r}\|^2\right| \le \delta/2\,\|\mathbf{r}\|^2$, see [BDDW08].

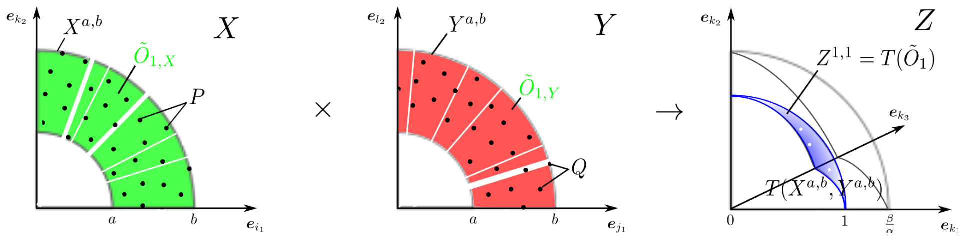


Figure 3.1: Net construction in the shells for covering the sphere in Z.

$Z^{1,1}$ by the Cartesian product of covers in X and Y, which is guaranteed by the RNMP on $X \times Y$, see Fig. 3.1. The norm multiplicativity of the global RNMP allows to arrange the nets in a multiplicative way and hence yields the additive behaviour of the dimensions. The proof is divided in an *algebraic part*, which uses Banach geometric tools and in an *probabilistic part*, where the measure concentration of the random linear map $\mathbf{\Phi}$ is applied.

The Algebraic part in [BDDW08] can be used again on X and Y as well to get an upper bound on the cardinality of $\mathcal{R}$, but now in terms of the covering numbers $N_\epsilon(X^1)$ and $N_\epsilon(Y^1)$. For this we need to control the norm of $\mathbf{z}$ by elements in $X \times Y$ which is possible if $\mathcal{B}$ has the global RNMP, since the representation set $O \subset X \times Y$ does not contain "bad" pairs for $Z := \mathcal{B}(X,Y)$. It is in fact not necessary to give an explicit parametrization of O, merely the existence of suitable representation pairs will suffice.

This means, the sphere $Z^{1,1}$ of the image can be represented by $\mathcal{B}(O_1) = \bigcup_{(\mathbf{x},\mathbf{y})\in O_1} \mathcal{B}(\mathbf{x},\mathbf{y})$ with $O_1 := \{(\mathbf{x},\mathbf{y}) \in O \mid \|\mathcal{B}(\mathbf{x},\mathbf{y})\| = 1\} \subset X \times Y$ and by Definition 6 it holds

$$\alpha \|\mathbf{x}\| \|\mathbf{y}\| \le 1 \le \beta \|\mathbf{x}\| \|\mathbf{y}\| \quad \text{for} \quad (\mathbf{x},\mathbf{y}) \in O_1. \tag{3.8}$$

Let us rescale the vectors in the pair $(\mathbf{x},\mathbf{y}) \in O_1$ by setting

$$\tilde{\mathbf{x}} := \frac{\mathbf{x}}{\|\mathbf{x}\|} \quad , \quad \tilde{\mathbf{y}} := \|\mathbf{x}\| \mathbf{y}. \tag{3.9}$$

Since X and Y are linear cones, we get by bilinearity:

$$\mathcal{B}(\tilde{\mathbf{x}},\tilde{\mathbf{y}}) = \mathbf{z} = \mathcal{B}(\mathbf{x},\mathbf{y}) \tag{3.10}$$

and from positive homogeneity of $\|\cdot\|$ we have with (3.8):

$$\alpha \|\tilde{\mathbf{y}}\| \le 1 \le \beta \|\tilde{\mathbf{y}}\| \quad \Rightarrow \quad \frac{1}{\beta} \le \|\tilde{\mathbf{y}}\| \le \frac{1}{\alpha}. \tag{3.11}$$

Since $0 < \alpha \le 1 \le \beta$ we can also choose a representation set $\tilde{O}_1$, contained in the symmetrized set of convex shells $X^{a,b} \times Y^{a,b}$ with common inner and outer radii $a := \beta^{-\frac{1}{2}} \le 1 \le \alpha^{-\frac{1}{2}} =: b$, i.e.

$$\tilde{O}_1 \subset \tilde{O}_{1,X} \times \tilde{O}_{1,Y} \subset X^{a,b} \times Y^{a,b}, \tag{3.12}$$

where $\tilde{O}_{1,X}$ and $\tilde{O}_{1,Y}$ are the projections of $\tilde{O}_1$ to X respectively Y and represent axial lines in the shells, see Fig. 3.1. Note, that $\tilde{O}_1$ can not be written as a Cartesian product, if some pairs are not allowed.

The algebraic part of the proof continues as follows: Any realization of $\mathbf{\Phi}$ is a linear map on a finite–dimensional normed space $\mathbb{R}^n$ and hence bounded, i.e., we can define

$$1 + \rho := \sup_{\mathbf{z} \in Z} \frac{\|\mathbf{\Phi z}\|}{\|\mathbf{z}\|} = \sup_{\mathbf{z} \in Z^{1,1}} \|\mathbf{\Phi z}\|, \tag{3.13}$$

where $1 + \rho \geq 0$ denotes the smallest upper bound (operator norm of $\mathbf{\Phi}$ restricted to Z). If we can show that $\rho \leq \delta$ we have shown the upper bound in (3.1). Note, that the supremum is not necessarily attained, since $Z^{1,1}$ is not necessarily compact (closed). In fact, this will not be a problem, since we can find a sequence $\mathbf{z}_n$ which can be arbitrarily close to $1 + \rho$.

What we really need, is the freedom in the representation of $Z^{1,1}$. Let $\mathcal{P} \subset X^{a,b}$ and $\mathcal{Q} \subset Y^{a,b}$ be ϵ-nets for $X^{a,b}$ resp. $Y^{a,b}$ with $\epsilon \in (0,1)$ and define $\mathcal{R} := \{\mathbf{r} = \mathcal{B}(\mathbf{p},\mathbf{q}) \mid (\mathbf{p},\mathbf{q}) \in \mathcal{P} \times \mathcal{Q}\} \subset Z$. It follows that[2] $|\mathcal{R}| \leq N_{\epsilon/b}(X^1) N_{\epsilon/b}(Y^1)$ sets $\mathcal{B}(X^\epsilon(\mathbf{p}), Y^\epsilon(\mathbf{q})) \subset Z$ with $X^\epsilon(\mathbf{p}) := X^\epsilon + \mathbf{p}$ and $Y^\epsilon(\mathbf{q}) := Y^\epsilon + \mathbf{q}$ cover the image $Z^{1,1}$ by (3.12), since every $(\mathbf{x},\mathbf{y}) \in \tilde{O}_1$ is contained in a Cartesian product $X^\epsilon(\mathbf{p}) \times Y^\epsilon(\mathbf{q})$ for some net point $(\mathbf{p},\mathbf{q}) \in \mathcal{P} \times \mathcal{Q}$. Note that this covering sets for $Z^{1,1}$ are not necessarily convex!

Since we want to control the norm of $\mathbf{z} - \mathbf{r}$ we will partition the difference by a sum of three elements in Z

$$\begin{aligned}
\mathcal{B}(\mathbf{x},\mathbf{y}) - \mathcal{B}(\mathbf{p},\mathbf{q}) &= \mathcal{B}(\mathbf{x},\mathbf{y}) - \mathcal{B}(\mathbf{p},\mathbf{y}) + \mathcal{B}(\mathbf{p},\mathbf{y}) - \mathcal{B}(\mathbf{p},\mathbf{q}) && (3.14)\\
&= \mathcal{B}(\mathbf{x}-\mathbf{p},\mathbf{y}) + \mathcal{B}(\mathbf{p},\mathbf{y}-\mathbf{q}) && (3.15)\\
&= \mathcal{B}(\mathbf{x}-\mathbf{p},\mathbf{y}-\mathbf{q}) + \mathcal{B}(\mathbf{x}-\mathbf{p},\mathbf{q}) + \mathcal{B}(\mathbf{p},\mathbf{y}-\mathbf{q}), && (3.16)
\end{aligned}$$

which gives for the norm by using the ϵ–net property and continuity (upper bound β) of $\mathcal{B}$

$$\begin{aligned}
\|\mathbf{z}-\mathbf{r}\| &\leq \|\mathcal{B}(\mathbf{x}-\mathbf{p},\mathbf{y}-\mathbf{q})\| + \|\mathcal{B}(\mathbf{x}-\mathbf{p},\mathbf{q})\| + \|\mathcal{B}(\mathbf{p},\mathbf{y}-\mathbf{q})\| && (3.17)\\
&\leq \beta(\|\mathbf{x}-\mathbf{p}\| \, \|\mathbf{y}-\mathbf{q}\| + \|\mathbf{x}-\mathbf{p}\| \, \|\mathbf{q}\| + \|\mathbf{p}\| \, \|\mathbf{y}-\mathbf{q}\|) && (3.18)\\
&\leq \beta(\epsilon^2 + \epsilon \|\mathbf{q}\| + \epsilon \|\mathbf{p}\|) && (3.19)
\end{aligned}$$

since for any $\mathbf{z}$ we can find $(\mathbf{p},\mathbf{q}) \in P \times Q$ with distance less than ϵ to $\mathbf{x}$ respectively $\mathbf{y}$. Note, the RNMP is only valid on Z and we need therefore the convexity of X and Y to guarantee that $\mathbf{x} - \mathbf{p} \in X$ and $\mathbf{y} - \mathbf{q} \in Y$! To bound the distance furhter, we need a bound for the net points $\mathbf{q}$ and $\mathbf{p}$ which are given by the RNMP lower bound as $b = \alpha^{-1/2} < \infty$. Moreover, this controls at the same time the norm of $\mathbf{r}$ itself by using the lower triangle inequality

$$|1 - \|\mathbf{r}\|\,| = |\,\|\mathbf{z}\| - \|\mathbf{r}\|\,| \leq \|\mathbf{z}-\mathbf{r}\| \leq \beta(\epsilon^2 + 2b\epsilon) \leq \beta(2b+1)\epsilon := c\epsilon. \tag{3.20}$$

[2] Note, that for the covering number it holds $N(X^b, X^\epsilon) = N(X^1, X^{\epsilon/b}) = N_{\epsilon/b}(X^1)$. Moreover $X^{a,b} \subset X^b$.

Hence, the set $\mathcal{R}$ is a $c\epsilon$–cover[3] for $Z^{1,1}$ with the sets $\mathcal{B}(X^{\epsilon}(\mathbf{p}), Y^{\epsilon}(\mathbf{q}))$. Applying $\mathbf{\Phi}$ on both sides and using the lower triangle inequality we get

$$|\,\|\mathbf{\Phi z}\| - \|\mathbf{\Phi r}\|\,| \leq \|\mathbf{\Phi}\mathcal{B}(\mathbf{x}-\mathbf{p},\mathbf{y}-\mathbf{q}) + \mathbf{\Phi}\mathcal{B}(\mathbf{x}-\mathbf{p},\mathbf{q}) + \Phi\mathcal{B}(\mathbf{p},\mathbf{y}-\mathbf{q})\|\,. \tag{3.21}$$

Using the upper triangle inequality and the bound[4] $1+\rho$ in (3.13) and (3.19) we get similar:

$$\begin{aligned} |\,\|\mathbf{\Phi z}\| - \|\mathbf{\Phi r}\|\,| &\leq (1+\rho)\big(\,\|\mathcal{B}(\mathbf{p},\mathbf{y}-\mathbf{q})\| + \|\mathcal{B}(\mathbf{x}-\mathbf{p},\mathbf{q})\| + \|\mathcal{B}(\mathbf{x}-\mathbf{p},\mathbf{y}-\mathbf{q})\|\,\big) & (3.22)\\ &\leq (1+\rho)c\epsilon. & (3.23)\end{aligned}$$

The Probabilistic part comes into play by the measure concentration (3.6) of $\mathbf{\Phi}$, applied to every net point $\mathbf{r} \in \mathcal{R}$. Hence we get as upper bound for $\|\mathbf{\Phi z}\|$ with (3.6)

$$\|\mathbf{\Phi z}\| \leq (1+\rho)c\epsilon + \|\mathbf{\Phi r}\| \leq (1+\rho)c\epsilon + (1+\delta/2)\,\|\mathbf{r}\|\,, \tag{3.24}$$

which holds for all $\mathbf{z} \in Z^{1,1}$ and hence for all $\mathbf{r} \in \mathcal{R}$ with probability $\geq 1 - 2|\mathcal{R}|e^{-c_0(\delta)m}$. Since $\|\mathbf{r}\| \neq 1$ only if $\alpha \neq \beta$ we define

$$\tilde{c} = \begin{cases} c & , \ \alpha \neq \beta \\ 0 & , \ \alpha = \beta \end{cases}. \tag{3.25}$$

and get

$$\|\mathbf{\Phi z}\| \leq (1+\rho)c\epsilon + (1+\delta/2)(\tilde{c}\epsilon + 1) = 1 + \rho c\epsilon + c\epsilon + (1+\delta/2)\tilde{c}\epsilon + \delta/2. \tag{3.26}$$

By definition of the supremum in (3.13), there exist a sequence $\{\mathbf{z}_k\} \subset Z^{1,1}$, s.t., for each $\eta > 0$ there exist $l \in \mathbb{N}$ with

$$1 + \rho - \eta \leq \|\mathbf{\Phi z}_l\| \tag{3.27}$$

and hence we get with (3.26)

$$\rho - \rho c\epsilon \leq c\epsilon + (1+\delta/2)\tilde{c}\epsilon + \delta/2 + \eta \quad \Leftrightarrow \quad \rho \leq \frac{2c\epsilon + (2+\delta)\tilde{c}\epsilon + 2\eta + \delta}{2(1-c\epsilon)}. \tag{3.28}$$

Now we can set η arbitrarily small, for example $\eta = \eta'\tilde{c}\epsilon/2$ with some $\eta' > 0$, then we have

$$\rho \leq \frac{2c\epsilon + (2+\delta+\eta')\tilde{c}\epsilon + \delta}{2(1-c\epsilon)}. \tag{3.29}$$

Since $\delta < 1$ we can actually set $\eta' = 1 - \delta > 0$. Hence it follows

$$\rho \leq \frac{2c\epsilon + 3\tilde{c}\epsilon + \delta}{2(1-c\epsilon)}. \tag{3.30}$$

[3] Strictly speaking, the cover is not a net, since in general $\mathbf{r} \notin Z^{1,1}$ and the cover sets are not convex nor symmetric.

[4] It is also here necessary, that $\mathbf{x} - \mathbf{p} \in X$ since only then the image is in Z, otherwise we could not apply the universal bound defined in (3.13). Also extending Z to a larger set would result in a circle argument.

Let us proceed by case distinction. If $\alpha = \beta$ then $\tilde{c} = 0$ and $c = 3$. Defining $\epsilon = \frac{\delta}{12} \leq 1$ with $\delta \in (0, 1)$ we get

$$\rho \leq \frac{6\epsilon + \delta}{2(1 - 3\epsilon)} \leq \delta. \tag{3.31}$$

If we have $\alpha \neq \beta$ then $\tilde{c} = c = c(\alpha, \beta)$. Defining $\epsilon = \frac{\delta}{7c} \leq 1$ with $\delta \in (0, 1)$ we get

$$\rho \leq \frac{\frac{5c\epsilon}{2} + \frac{\delta}{2}}{1 - c\epsilon} \leq \frac{\frac{5\delta + 7\delta}{14}}{1 - \frac{\delta}{7}} \overset{\delta<1}{\leq} \frac{\frac{12}{14}\delta}{\frac{6}{7}} = \delta \tag{3.32}$$

with probability

$$> 1 - 2N_{\delta/d}(X)N_{\delta/d}(Y)e^{-c_0(\delta)m} \quad ,$$
$$\text{where} \quad d := d(\alpha, \beta) = \begin{cases} 7\beta(2/\sqrt{\alpha} + 1) & , \quad \alpha \neq \beta \\ 12 & , \quad \alpha = \beta \end{cases}. \tag{3.33}$$

The lower bound $1 - \delta$ follows from this with

$$\|\mathbf{\Phi z}\| \geq \|\mathbf{\Phi r}\| - (1 + \rho)c\epsilon. \tag{3.34}$$

By considering all $\mathbf{z} \in Z^{1,1}$ we get by inserting (3.32) with same probability as in (3.33)

$$\|\mathbf{\Phi z}\| \geq \left(1 - \frac{\delta}{2}\right)\left(1 - \tilde{c}\frac{\delta}{d}\right) - (1 + \delta)\frac{c\delta}{d}. \tag{3.35}$$

If $\alpha = \beta$ then $\tilde{c} = 0, c = 3$ and $d = 12$. This gives

$$\|\mathbf{\Phi z}\| \geq 1 - \delta/2 - \delta/2 = 1 - \delta. \tag{3.36}$$

If $\alpha \neq \beta$ then $\tilde{c} = c, d = 7c$ and we get

$$\|\mathbf{\Phi z}\| \geq 1 - \delta/2 - c\frac{\delta}{d} + \frac{c\delta^2}{2d} - \frac{c\delta}{d} - \frac{c\delta^2}{d} = 1 - \frac{\delta}{2} - \frac{2c\delta}{d} - \frac{c\delta^2}{2d} \tag{3.37}$$
$$= 1 - \frac{\delta}{2} - \frac{4\delta + \delta^2}{14} \geq 1 - \frac{\delta}{2} - \frac{\delta}{2} = 1 - \delta \tag{3.38}$$

□

3.2 RIP for Sparse Multiplications

Now we can derive a similar RIP result as in [BDDW08] for all $(s,f)-$sparse multiplications in $\mathbb{R}^n$. Note, this alone does not imply a stable embedding of the image set, but to any sparse image generated by the associated linear maps $\mathrm{L}_{\mathbf{x}}$ and $\mathrm{R}_{\mathbf{y}}$, see Corollary 4.

Theorem 9 (RIP on Sparse Multiplications)**.** *Let be $s,f,n \in \mathbb{N}$ with $s,f \ll n$ and $\mathcal{B}_n$ a multiplication in $\mathbb{R}^n$ statisfying the RNMP on $\Sigma_s^n \times \Sigma_f^n$ with bounds $\alpha,\beta > 0$ and independent of n. Then for any $\delta \in (0,1)$, there exist a constant $c > 0$, s.t. for $m = \mathcal{O}((s+f)\log(n/s))$ every realization of a sub-Gaussian matrix $\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^m$ with $(\mathbf{\Phi})_{ij} \sim \mathcal{N}(0,1/m)$ fulfills for every $\mathbf{x} \in \Sigma_s^n$ the following $\delta-$stability*

$$(1-\delta)\|\mathbf{z}\| \le \|\mathbf{\Phi z}\| \le (1+\delta)\|\mathbf{z}\| \quad \textit{for} \quad \mathbf{z} \in \mathcal{B}_n(\Sigma_s^n \circledast \Sigma_f^n) \tag{3.39}$$

with probability

$$\ge 1 - 2e^{-cm}. \tag{3.40}$$

Proof. Let X and Y be two linear spaces with dimensions s resp. f in Lemma 5, then $N_{\delta/d}(X^1) \le (3d/\delta)^s$ and $N_{\delta/d}(Y^1) \le (3d/\delta)^f$, s.t. we get a failure probability for the RIP not larger than

$$2(3d/\delta)^{s+f}e^{-c_0(\delta)m}. \tag{3.41}$$

If the RNMP holds on $\Sigma_s^n \times \Sigma_f^n$ with bounds α,β, then we have $\binom{n}{s}\binom{n}{f}$ different s,f dimensional subspaces X,Y. Since $\binom{n}{s} \le (en/s)^s$, we get for $k = s+f$

$$2(en/s)^s(en/s)^f(3\delta/d)^{s+f}e^{-c_0(\delta)m} = 2(en/s)^k(3d/\delta)^k e^{-c_0(\delta)m} \tag{3.42}$$

$$= 2e^{-c_0(\delta)m + k(\ln(n/s)+1+\ln(3d/\delta))}. \tag{3.43}$$

Let now $c_1 > 0$ such that $k \le c_1 m/\ln(n/s)$, then

$$\le 2e^{-c_0(\delta)m + c_1 m\left(1+\frac{1+\ln(3d/\delta)}{\ln(n/s)}\right)}. \tag{3.44}$$

To get an exponential decay of the failure probability in m we need to show the existence of a positive constant $c_2 > 0$ such that

$$2e^{-c_0(\delta)m + c_1 m(1+(2+\ln(3d/\delta)/\ln(n/s))} = 2e^{-\overbrace{\left(c_0(\delta)-c_1(1+\frac{1+\ln(3d/\delta)}{\ln(n/k)})\right)}^{=:c}m} = 2e^{-cm}. \tag{3.45}$$

Since we have for all $s,f \in N$ for the lower bound $\alpha \le 1$ we get with

$$d(s,f) = 7\frac{\beta}{\alpha}(2+\sqrt{\alpha}) < \frac{21\beta}{\alpha} \tag{3.46}$$

the following estimation

$$c_0(\delta) - c_1(1 + \frac{1 + \ln(3d/\delta)}{\ln(n/s)}) \geq c_0(\delta) - c_1 \underbrace{\left(1 + \frac{1 + \ln(63\beta/\alpha) - \ln\delta}{\ln(n/s)}\right)}_{=:c_2(s,f,n)} = c. \tag{3.47}$$

Here, the ratio $\beta(s,f)/\alpha(s,f)$ goes to infinity for $f \to \infty$. But for fixed s, f this ratio is finite and hence for n large enough $c_2(s,f,n)$ is positive for any $\delta \in (0,1)$. Then it is easy to decrease c_1 until it gets compensated by $c_0(\delta)$.

Note, for subgaussian matrices we have $c_0(\delta) = (3\delta^2 - \delta^3)/48$ by (3.4). Therefore, if we set $c_1 = \delta^2/(48c_2)$ we get a positive c. Hence, from $m = \mathcal{O}((s+f)\log(n/s))$ random measurements we obtain the RIP with exponential high probability if $s \ll n$. □

This enables immediatley the following important universal compressed sensing result for LTI systems, described by zero-padded circular convolutions.

Corrolary 4 (Compressibility of sparse LTI systems)**.** *Let be* $s, f, n \in \mathbb{N}$ *with* $s, f \ll n$ *and* $\mathbf{0}_{n-1}$ *the zero vector in* $\mathbb{R}^{n-1}$*. Then for any* $\delta \in (0,1)$*, there exists a constant* $c > 0$*, s.t. for* $m = \mathcal{O}((s+2f)\log n)$ *every realization of a sub-Gaussian matrix* $\mathbf{\Phi} : \mathbb{R}^n \to \mathbb{R}^m$ *with* $(\mathbf{\Phi})_{ij} \sim \mathcal{N}(0, 1/m)$ *fulfills for every* $\mathbf{x} \in \Sigma_s^n$ *the following* δ*–stability*

$$(1-\delta)\|\mathbf{z}_1 - \mathbf{z}_2\| \leq \|\mathbf{\Phi}(\mathbf{z}_1 - \mathbf{z}_2)\| \leq (1+\delta)\|\mathbf{z}_1 - \mathbf{z}_2\| \quad \textit{for} \quad \mathbf{z}_1, \mathbf{z}_2 \in \begin{pmatrix} \mathbf{x} \\ \mathbf{0}_{n-1} \end{pmatrix} \circledast \begin{pmatrix} \Sigma_f^n \\ \mathbf{0}_{n-1} \end{pmatrix} \tag{3.48}$$

with probability $\geq 1 - 2e^{-cm}$*.*

Proof. Note, w.l.o.g. we can assume $s \leq f$ by commutativity of the convolution. Let $\mathbf{x} \in (\Sigma_s^n, \mathbf{0})$ an arbitrary channel parameter (generating the circular matrix $\mathbf{X}$ and hence the channel). Then we know that

$$\mathbf{z} = \mathbf{z}_1 - \mathbf{z}_2 = \mathbf{x} \circledast \mathbf{y}_1 - \mathbf{x} \circledast \mathbf{y}_2 = \mathbf{x} \circledast \mathbf{y} \in (\Sigma_s^n, \mathbf{0}) \circledast (\Sigma_{2f}^n, \mathbf{0}), \tag{3.49}$$

since $\mathbf{y} = \mathbf{y}_1 - \mathbf{y}_2 \in (\Sigma_{2f}^n, \mathbf{0})$. Hence, by using Corollary 2, assuming n is sufficiently large, and Theorem 9 we have for any realization $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ of a sub-Gaussian random matrix

$$(1-\delta)\|\mathbf{z}_1 - \mathbf{z}_2\| \leq \|\mathbf{\Phi}\mathbf{z}\| = \|\mathbf{\Phi}(\mathbf{z}_1 - \mathbf{z}_2)\| = \|\mathbf{\Phi}(\mathbf{x} \circledast \mathbf{y})\| \leq (1+\delta)\|\mathbf{z}_1 - \mathbf{z}_2\| \tag{3.50}$$

with probability larger than $1 - 2e^{-cm}$ for some constant $c > 0$ and any $m = \mathcal{O}((s+2f)\log n)$. □

3.3 From RIP on Bilinear Images to Stable Compressed Sensing

Unfortunately, the RIP on Z does not imply a RIP on $Z - Z$, since Z is not a linear set, contrary to the standard compressed sensing result, where it holds $\Sigma_k^n - \Sigma_k^n = \Sigma_{2k}^n$. Hence, it is still open, how the RNMP can guarantee a RIP on the difference set $Z - Z$ or in general, taking the image of all sparse vectors $\Sigma_s^n \times \Sigma_f^n$, how can, on the set

$$\Delta = \mathcal{B}(\Sigma_s^n, \Sigma_f^n) - \mathcal{B}(\Sigma_s^n, \Sigma_f^n) \tag{3.51}$$

a RIP for random projections onto m dimension be established? Only if for some random $\mathbf{\Phi}$ the RIP holds on Δ, we have a stable random embedding by $\mathbf{\Phi}$ of $\mathcal{B}(\Sigma_s^n, \Sigma_f^n)$.

Nevertheless, as seen in the previous Chapter 2, for the symmetrized circular convolution a stable phase retrieval could be established. If the ambient dimension is larger than $n \geq \tilde{n}(s, f)$ as derived in the Theorem 2, then an n independence of the stable embedding is obtained. Moreover, if we could establish a random measurement on the magnitude Fourier measurements, such that the measure concentration or the RIP is fulfilled, a stable reconstruction from m random measurements of the image Z could be in principal established.

4 Orthogonalization of Convolutions with Compact Support

In this chapter we will investigate semi-discrete convolution systems and develop a transformation, acting as a finite sequence on a time-limited continuous pulse. The transformation shapes the output signal in frequency and add a shift orthogonality property to the pulse. The result of this chapter was published in [WJ12a]. By abuse of notation we will here denote by capital letters $A, B, C, \ldots$ numbers and constants.

In communication theory one is interested to obtain a reliable high data rate transmission. Usually, the data rate depends by the famous *Shannon* formula on the bandwidth. But extending the bandwidth could yield disturbance of existing systems, e.g. GPS and UMTS. Hence the Federal Communications Commission (FCC) released [FCC02] a very low power spectral density (PSD) mask for ultra-wideband (UWB) systems. To ensure that sufficiently high SNR is maintained in the frequency band $F = [0, 14]$GHz, as required by the FCC, the pulses have to be designed for a high efficient frequency utilisation. This utilisation can be expressed by the pulses normalized effective signal power (NESP) [LYG03]. Several pulse shaping methods for pulse amplitude and pulse position modulation (PAM and PPM) were developed in the last decade based on a FIR prefiltering of a fixed basic pulse. A SNR optimization under the FCC mask constraints then reduces to a FIR filter optimization [Tia+06; WBV99; LYG03; WTDG06]. Since the SNR is limited, the amount of signals can be increased to achieve higher data rates or to enable multi-user capabilities. For coherent and synchronized transmission over memoryless AWGN channels, an increased number N of mutually orthogonal UWB signals inside the same time slot, known as N-ary orthogonal signal design, improves the BER performance over E_b/N_0 and hence the achievable rate of the system [SHL94, Ch.4].

Combining spectral shaping and orthogonalization is an inherently difficult problem being neither linear nor convex. Therefore most methods approach this problem **sequentially**, e.g. combining spectral optimization with a GRAM-SCHMIDT construction [And05; WTDG06; Pro01]. This usually results in an unacceptable loss in the NESP value of the pulses [WTDG04; DK07]. Moreover these orthogonal pulses are different in shape and therefore not useful for PPM. Furthermore, a big challenge in UWB impulse radio (UWB-IR) implementation are high rate sampling operations. Therefore, an analog transmission scheme is desirable [PCWD03; RM07; DK07] to avoid high sampling rates in AD/DA conversion [PCWD03].

Usually, PPM is referred to an orthogonal (non-overlapping) pulse modulation scheme. To achieve higher data rates in PPM, pulse overlapping was already investigated in optical communication [BDK84] and called OPPM. An application to UWB was studied for the binary case with a Gaussian monocycle [ZMSF01]. To the authors' knowledge, no orthogonal overlapping PPM (OOPPM) signaling has been considered based on strictly time-limited pulses. In this chapter we propose a new analog pulse shape design for UWB-IR to enable an almost OOPPM signaling, which approximates the OOPPM scheme up to a desired accuracy.

In our approach, we first design a time-limited spectral optimized pulse p and perform afterwards a Löwdin orthogonalization of the set of $2M+1$ integer translates $\{p(\cdot - k)\}_{k=-M}^{M}$. This orthogonalization method provides an implementable and stable approximation $p^{\circ,M}$ to a Nyquist pulse p°. Using the *Fourier transformation* for p given by $\hat{p}(\nu) = \int_{\mathbb{R}} p(t) e^{-i2\pi t\nu} dt$ for every $\nu \in \mathbb{R}$, the Nyquist[1] pulse p° can be expressed in the frequency-domain

$$\hat{p}^{\circ}(\nu) = \frac{\hat{p}(\nu)}{\sqrt{\sum_k |\hat{p}(\nu - k)|^2}} \tag{4.1}$$

for ν almost everywhere, which is known as the *orthogonalization trick* [Dau90]. Usually in digital signal processing, see Fig. 4.1, the time-continuous signal (pulse) is sampled by an analog to digital (AD) operation to obtain a time-discrete signal in $\mathbb{C}^{2M+1}$. Then a discretization of the orthogonalization in (4.1) yields a finite digital transformation D^M to construct a discrete signal in the Fourier domain which has again to be transformed by a DA operation to obtain finally an approximation of the Nyquist pulse [JS02].
Instead of using such an AD/DA conversion to operate in a discrete domain, we use the **democratic** LÖWDIN orthogonalization B^M, found by Per-Olov Löwdin in [Löw50; Löw70], where all $2M + 1$ linear independent pulse translates are involved simultaneously by a linear combination to generate a set of time-limited mutual orthogonal pulses. Hence the Löwdin orthogonalization is order independent. The Löwdin pulses constitute then an orthonormal basis for the span $\mathscr{V}^M(p)$ of the initial basis $\{p(\cdot - k)\}_{k=-M}^{M}$. Moreover, as we will show in our main Theorem 12, the Löwdin orthogonalization B^M is a stable approximation method to the Nyquist construction in (4.1) and operates completely in the analog domain. Note, that a discretization of (4.1) generates neither shift-orthogonal pulses nor a set of mutual orthogonal pulses. The important property of the Löwdin method is its minimal summed energy distortion to the initial basis. It turns out that all orthogonal pulses maintain the spectral efficiency "quite well".
As M tends to infinity the Löwdin orthogonalization $B = B^{\infty}$ applied to the initial pulse p delivers the Nyquist pulse p°, which allows a real-time OOPPM system with only one single matched filter at the receiver. Since the Löwdin transform B^M is hard to compute and to control we introduce an approximate Löwdin orthogonalization (ALO) $\tilde{B}^M$ and investigate its stability and convergence properties. It turns out that for fixed M the transformations B^M and $\tilde{B}^M$ are both implementable by a FIR filter bank like the spectral optimization. We call $B^M p$ and $\tilde{B}^M p$ *approximate Nyquist* pulses, since we observed that even for finite M our analog approximation yields time–limited pulses with almost shift–orthogonal character since the sample values of the autocorrelation are below a measurable magnitude for the correlator. Hence such a construction of approximate Nyquist pulses seems to be promising for an OOPPM system.

The structure of this chapter is as follows: In Section 4.1 we introduce the signal model and motivate our spectrally efficient N-ary orthogonal overlapping PPM design for UWB systems.

[1] This also called a square-root Nyquist pulse in the literature.

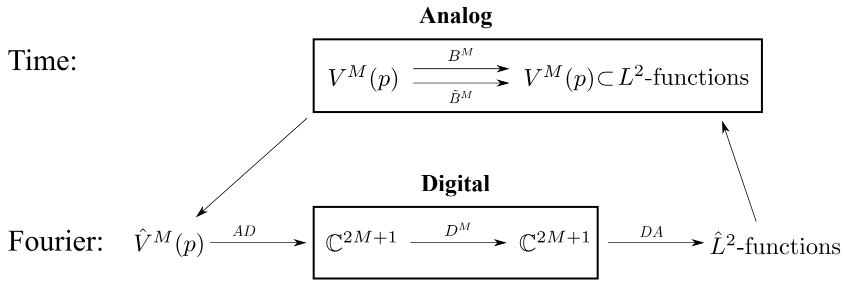


Figure 4.1: Analog and discrete approximation methods in time and frequency domains

Section 4.2 presents the state of the art in FCC optimal pulse shaping for UWB-IR based on PPM or PAM transmission by FIR prefiltering of Gaussian monocycles which is a necessary prerequisite for our design. To develop our approximation and convergence results in Section 4.3 we introduce the theory of shift–invariant spaces in Section 4.3.1 and in Section 4.3.2 the Löwdin orthogonalization for a set of N translates to provide a N-ary orthogonal overlapping transmission. Our main result is given in Section 4.3.3, where we consider the stability of the Löwdin orthogonalization B^M (for M increasing to infinity) and develop for this a simplified approximation method $\tilde{B}^M$, called ALO. In Section 4.4 we study certain properties of our filter design and investigate the combination of both approaches. Furthermore, in Section 4.4.1 we develop a connection between our result and the canonical tight frame construction. Finally in Section 4.5, we demonstrate that the ALO and Löwdin transforms yields for sufficiently large filter orders compactly supported approximate Nyquist pulses, which can be used for OOPPM having high spectral efficiency in the FCC region. Moreover, the Löwdin pulses provides also a spectrally efficient $2M+1$-ary orthogonal pulse shape modulation (PSM) [GK08].

4.1 Signal Model

To control signal power in time or frequency locally we need bounded pulses in $L^\infty := \{p : \mathbb{R} \to \mathbb{C} \mid \|p\|_\infty < \infty\}$ with

$$\|p\|_\infty := \operatorname*{ess\,sup}_{t\in\mathbb{R}} \{|p(t)|\}, \tag{4.2}$$

where the *essential supremum* is defined as the smallest upper bound for $|p(t)|$ except on a set of measure zero. If the pulse is continuous than this implies boundedness everywhere. UWB-IR technology uses ultra short pulses, i.e. strictly time-limited pulses with support contained in a finite interval $X \subset \mathbb{R}$. We call such L^2-functions *compactly supported* in X and denote its closed span by the subspace $L^2(X)$. The coding of an information sequence $\{d_n\} = \{d_n\}_{n\in\mathbb{Z}}$ is realized by pulse modulation techniques [Pro01; Mid96] of a fixed normalized basic pulse $p \in L^2([0, T_p])$ with duration T_p.

A relevant issue in the UWB-IR framework and in our work is the spectral shape of the pulse. In this section we will therefore summarize the derivation of spectral densities for common UWB modulation schemes such as PAM [Tia+06], PPM [Sch93; LYG03; WS98; Win99; RMS98] and combinations of both [NM03] to justify our spectral shaping in the next section. Antipodal PAM and N-ary PPM are linear modulation schemes which map each data symbol d_n to a pulse (symbol) $s_{d_n}(t)$ with the same power spectrum $\mathcal{E}\,|\hat{p}(\nu)|^2$. If we fix the energy $\mathcal{E}$ of the transmitted symbols and the pulse repetition time (symbol duration) T_s, we will show now for certain discrete random processes (e.g. i.i.d. processes [WJT10a]) that the power spectrum density (PSD) of the transmitted signals is given by

$$S_u(\nu) = \frac{\mathcal{E}\,|\hat{p}(\nu)|^2}{T_s}. \tag{4.3}$$

Hence an optimization of the pulse power spectrum to the FCC mask $S_{\text{FCC}}(\nu)$ over the band F in Section 4.2 increases the transmit power. To be more precise, PPM produces discrete spectral lines, induced by the periodic pulse repetition, the use of uniformly distributed pseudo-random time hopping (TH) codes $c_n \in [0, N_c]$ was suggested to reduce this effect and to enable multi-user capabilities [Sch93; Win99; WS98; Win02; WZK08]:

$$u(t) = \sum_{n=-\infty}^{\infty} \sqrt{\mathcal{E}} p(t - nT_f - c_n T_c - d_{\lfloor n/N_f \rfloor} T). \tag{4.4}$$

In [Win02] this is called framed TH by a random sequence, since the coding is repeated in each frame N_f times with a clock rate of $1/T_f$. Hence $N_f T_f = T_s$ is the symbol duration for transmitting one out of N symbol waveforms representing the encoded information symbol $d_n \in \{0, \dots, N-1\}$. To prevent ISI and collision with other users, the maximal PPM shift T and TH shift T_c have to fulfill $NT \le T_c$ and $N_c T_c \le T_f$. To ensure mutual orthogonality of all symbols one requires $T > T_p$. The PSD for independent discrete i.i.d. processes $\{c_n\}, \{d_n\}$ follows from the Wiener-Khintchine Theorem [Win02; VRE05] to [LYG03, (5)], [NM03; NM06].

$$S_u(\nu) = \frac{\mathcal{E}\,|\hat{p}(\nu)|^2}{T_f}\left[1 - \left|G_\beta(\nu)\right|^2 + \frac{\left|G_\beta(\nu)\right|^2}{T_f}\sum_k \delta\left(\nu - \frac{k}{T_f}\right)\right] \tag{4.5}$$

$$\text{with } \left|G_\beta(\nu)\right| = \frac{1}{N_c N}\left|\frac{\sin(\pi\nu T_c N_c)}{\sin(\pi\nu T_c)}\right|\left|\frac{\sin(\pi\nu T N)}{\sin(\pi\nu T)}\right| \text{ and } \delta \text{ is the Dirac distribution.} \tag{4.6}$$

However, a more effective and simple reduction method without the use of frame repetition ($N_f = 1$) or random TH ($N_c = 1$) has been proposed in [NM03; NM06]. Here antipodal PAM with $a_n \in \{-1, 1\}$ is combined with N-ary PPM modulation, for $NT \le T_s$

$$u(t) = \sqrt{\mathcal{E}} \sum_n a_n p(t - nT_s - d_n T). \tag{4.7}$$

The PSD for such i.i.d. processes is well known [Mid96, Sec.4.3]:

$$S_u(\nu) = \frac{\mathcal{E}\,|\hat{p}(\nu)|^2}{T_s}\left(\mathbb{E}[a^2] - \left|\mathbb{E}[a]\mathbb{E}[e^{-\imath 2\pi\nu dT}]\right|^2 + \frac{\left|\mathbb{E}[a]\mathbb{E}[e^{-\imath 2\pi\nu dT}]\right|^2}{T_s}\sum_n \delta\left(\nu - \frac{n}{T_s}\right)\right), \tag{4.8}$$

since we have $G_\beta(\nu) = \mathbb{E}[a]\mathbb{E}[e^{i2\pi\nu dT}]$ in (4.6). For an i.i.d. process $\{a_n\}$ with expectation $\mathbb{E}[a] = 0$ and variance $\mathbb{E}[a^2] = 1$ the PSD reduces to (4.3). Hence the effective radiation power is *essentially determined by the pulse shape* times the energy $\mathcal{E}$ per symbol duration T_s and should be bounded pointwise on F below the FCC mask S_{FCC}

$$S_u(\nu) = \mathcal{E}\frac{|\hat{p}(\nu)|^2}{T_s} \leq S_{\text{FCC}}(\nu) \quad \text{for all} \quad \nu \in F. \tag{4.9}$$

The optimal receiver for N-ary orthogonal PPM in a memoryless AWGN channel with noise power density $\mathcal{N}_0$ is the coherent correlation receiver. The bit rate R_b and average symbol error probability P_s is given as [SHL94, p. 4.1.4]

$$R_b = \frac{\log N}{T_s} \quad \text{and} \quad P_s(\mathcal{E}) \leq (N-1)\,\text{erfc}\left(\sqrt{\frac{\mathcal{E}}{\mathcal{N}_0}}\right), \tag{4.10}$$

where erfc is the complementary error function. Hence, a performance gain for fixed T_s is achieved by increasing N and/or $\mathcal{E}$.

Increasing N Usually non–overlapping pulses are necessary in PPM to guarantee orthogonality of the set $\{p(\cdot - nT)\} := \{p(\cdot - nT)\}_{n\in\mathbb{Z}}$ of pulse translates, i.e. $T > T_p$. For fixed T_p this limits the number of pulses N in $[0, T_s]$ and hence the data rate. In this chapter we will design an orthogonal overlapping PPM (OOPPM) system by keeping all overlapping translates mutually orthogonal. But such Nyquist pulses are in general not time–limited, i.e. not compactly supported. In fact, we will show that for a particular class of compactly supported pulses a non-overlapping of the shifts is necessary to obtain strict shift-orthogonality. However, we derive overlapping compactly supported pulses approximating the Nyquist pulse in (4.1) and characterize the convergence. These approximated Nyquist pulses allow a realizable N-ary OPPM implementation based on FIR filtering of time–limited analog pulses.

Increasing $\mathcal{E}$ The maximization of $\mathcal{E}$ with respect to the FCC mask was already studied in [LYG03; Tia+06] where a FIR prefiltering is used to shape the pulse such that its radiated power spectrum efficiently exploits and strictly respect the FCC mask. Note that the FCC regulation in (4.9) is a local constraint and does not force a strict band-limited design, however fast frequency decay outside the interval F is desirable for a hardware realization.

Our combined approach now relies on the construction of two prefiltering operations to shape a fixed initial pulse. The first filter shapes the pulse to optimally exploit the FCC mask and the second filter generates an approximated Nyquist pulse. The filter operations can be described as *semi–discrete convolutions* of pulses $p \in L^2$ with sequences $\mathbf{c} \in \ell^2$

$$p *'_T \mathbf{c} := \sum_{k\in\mathbb{Z}} c_k p(\cdot - kT), \tag{4.11}$$

with clock rate $1/T$. If we restrict ourselves to FIR filters of order n, we consider only sequences $\mathbf{c}^n \in \ell^2_n \cong \mathbb{C}^n$.

4.2 FCC Optimization of a Single Pulse

The first prefilter operation generates an optimized FCC pulse p. To generate a time-limited real-valued pulse we consider a real-valued initial input pulse $q \in L^2([-T_q/2, T_q/2])$ and a real-valued (causal) FIR filter $\mathbf{g}^L \in \mathbb{R}^L$. A common UWB pulse is the truncated Gaussian monocycle q [LYG03; Tia+06; BEZJ07], see also Section 4.5. The prefilter operation is then:

$$p(t) = (q *'_{T_0} \mathbf{g}^L)(t) = \sum_{k=0}^{L-1} g_k^L q(t - kT_0) \tag{4.12}$$

which results in a maximal duration $T_p = (L-1)T_0 + T_q$ of p.

To maximize the PSD according to (4.9) we have to shape the initial pulse by the filter $\mathbf{g}^L$ to exploit efficiently the FCC mask S_{FCC} in the passband $F_p \subset F$, i.e. to maximize the ratio of the pulse power in F_p and the maximal power allowed by the FCC

$$\eta(p) := \int_{F_p} |\hat{p}(\nu)|^2 d\nu \Bigg/ \int_{F_p} S_{\text{FCC}}(\nu) d\nu. \tag{4.13}$$

This is known as the direct maximization of the NESP value $\eta(p)$, see [WTDG06]. Here we already included the constants $\mathcal{E}$ and T_s in the basic pulse p. If we fix the initial pulse q, the clock rate $1/T_0$ and the filter order L, we get the following optimization problem

$$\begin{aligned} &\max_{\mathbf{g}^L \in \mathbb{R}^L} \tilde{\eta}\left(q *'_{T_0} \mathbf{g}^L\right) \\ &\text{s.t. } \forall \nu \in F : \left|\hat{\mathbf{g}}^L(\nu)\right|^2 \cdot |\hat{q}(\nu)|^2 \leq S_{\text{FCC}}(\nu), \end{aligned} \tag{4.14}$$

where $\hat{\mathbf{g}}^L$ denotes the $1/T_0$ periodic Fourier series of $\mathbf{g}^L$, which is defined for an arbitrary sequence $\mathbf{c} \in \ell^2$ as

$$\hat{\mathbf{c}}(\nu) = \sum_{n=-\infty}^{\infty} c_n e^{-2\pi i \nu n T_0}. \tag{4.15}$$

Since $\mathbf{g}^L \in \ell^2(L)$ the sum in (4.15) becomes finite for $\hat{\mathbf{g}}^L$. Liu and Wan [LW08] studied the non-convex optimization problem (4.14) with non-linear constraints numerically with fmincon, a MATLAB program. The disadvantage of this approach lies in the trap of a local optimum, which can only be overcome by an intelligent choice of the start parameters.

Alternatively (4.14) can be reformulated in a convex form by using the Fourier series of the autocorrelation $r_{\mathbf{g}^L,n} := \sum_k g_k^L g_{k-n}^L$ of the filter $\mathbf{g}^L$ [DLS02]. Since $r_{\mathbf{g}^L,n} = r_{\mathbf{g}^L,-n}$ (real-valued symmetric sequence) we can write for ν in the frequency band $[-\frac{1}{2T_0}, \frac{1}{2T_0}]$

$$\hat{\mathbf{r}}_{\mathbf{g}^L}(\nu) := \sum_{n=0}^{L-1} r_{\mathbf{g}^L,n} \phi_n(\nu) = \left|\hat{\mathbf{g}}^L(\nu)\right|^2. \tag{4.16}$$

By using the basis $\phi := \{1, 2\cos(2\pi\nu T_0), 2\cos(2\pi\nu 2T_0), \dots\}$ one gest $|\hat{p}(\nu)|^2 = \hat{\mathbf{r}}_{\mathbf{g}^L}(\nu) \cdot |\hat{q}(\nu)|^2$. Due to the symmetry of ϕ_n and S_{FCC} we can restrict the constraints in (4.14) to $F = [0, \frac{1}{2T_0}]$ and obtain the following semi–infinite linear problem:

$$\max_{\mathbf{r}\in\mathbb{R}^L} \sum_{n=0}^{L-1} r_{\mathbf{g}^L,n} c_n \quad \text{such that} \quad 0 \le \hat{\mathbf{r}}_{\mathbf{g}^L}(\nu) \le \mathcal{M}(\nu) \ \text{ for all } \ \nu \in \left[0, \frac{1}{2T_0}\right] \tag{4.17}$$

$$\text{with} \quad \mathcal{M}(\nu) := \frac{S_{\text{FCC}}(\nu)}{|\hat{q}(\nu)|^2} \quad \text{and} \quad c_n := \int_{F_p} |\hat{q}(\nu)|^2 \phi_n(\nu) d\nu. \tag{4.18}$$

Since the FCC mask is piecewise constant, we separate $\mathcal{M}(\nu)$ into five sections $\mathcal{M}_i(\nu)$ [BEJJ06] and get the inequalities

$$\forall i = 1, \dots, 5 : \ \hat{\mathbf{r}}_{\mathbf{g}^L}(\nu) \le \mathcal{M}_i(\nu) \quad \text{for} \quad \nu \in [\alpha_i, \beta_i] \tag{4.19}$$

with $\beta_1 = 1.61, \beta_2 = 1.99, \beta_3 = 3.1, \beta_4 = 10.6, \beta_5 = 14$ and $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0, \alpha_5 = \beta_4$ in GHz, see Fig. 4.2.
The necessary lower bound for $\mathbf{r}$ reads

$$\hat{\mathbf{r}}_{\mathbf{g}^L}(\nu) \ge 0 \quad \text{for} \quad \nu \in \left[0, \frac{1}{2T_0}\right] = [0, 14]\text{GHz}. \tag{4.20}$$

To formulate the constraints in (4.18) for $\mathbf{r}$ as a positive bounded cone in $\mathbb{R}^L$ we approximate $\mathcal{M}_i(\nu)$ by trigonometric polynomials[2] $\Gamma_i^L(\nu) := \sum_n \gamma_{i,n}^L \phi_n(\nu)$ of order L in the L^2-norm [BEJJ06]. The semi-infinite linear constraints in (4.19) describe a compact convex set [DLS02, (40),(41)]. To see this, let us introduce the following lower bound cones for $\theta \in [0, \frac{1}{2T_0}]$

$$K_{\text{low}}^L(\theta) = \left\{ \mathbf{r} \in \mathbb{R}^L \,\middle|\, \sum_{k=0}^{L-1} r_{\mathbf{g}^L,n} \phi_n(\nu) \ge 0, \nu \in \left[\theta, \frac{1}{2T_0}\right] \right\}. \tag{4.21}$$

For $\theta = 0$ the positive cone $K_0^L = K_{\text{low}}^L(0)$ defines the lower bound in (4.20) if we set $T_0 = \frac{1}{28\text{GHz}}$.
To formulate the non-constant upper bounds, one can use the approximation functions $\Gamma_i^L(\nu)$ [BEJJ06] given in the same basis ϕ as $\left|\hat{\mathbf{g}}^L(\nu)\right|^2$. For each $i \in \{1, \dots, 5\}$ the bounds in (4.19) are then equivalent to

$$\sum_{n=1}^{L} (\gamma_{i,n}^L - r_{\mathbf{g}^L,n}) \phi_n(\nu) \ge 0 \quad \text{for} \quad \nu \in [\alpha_i, \beta_i]. \tag{4.22}$$

For the upper bounds, we just have to set $\rho_{i,n}^L := \gamma_{i,n}^L - r_{\mathbf{g}^L,n}$ for each $i = 1, \dots, 5$ and $n \ge 1$, which leads to the upper bound cones

$$K_{\text{up}}^L(\theta_i) = \left\{ \mathbf{r} \in \mathbb{R}^L \,\middle|\, \sum_{n=0}^{L-1} \rho_{i,n}^L \phi_n(\nu) \ge 0, \nu \in \left[\theta_i, \frac{1}{2T_0}\right] \right\}, \tag{4.23}$$

$$\bar{K}_{\text{up}}^L(\theta_i) = \left\{ \mathbf{r} \in \mathbb{R}^L \,\middle|\, \sum_{n=0}^{L-1} \rho_{i,n}^L \phi_n(\nu) \ge 0, \nu \in [0, \theta_i] \right\}. \tag{4.24}$$

[2] Since the FCC mask divided by the Gaussian power spectrum is monotone increasing from 0 to 10.6GHz we can let $\Gamma_1^L, \dots, \Gamma_4^L$ overlap.

The five upper bound cones K_i^L are then

$$\forall i = 1, \dots, 4: \; K_i^L := \bar{K}_{\text{up}}^L(\beta_i) \quad \text{and} \quad K_5^L := K_{\text{up}}^L(\alpha_5). \tag{4.25}$$

Since the autocorrelation has to fulfill all these constraints, it has to be an element of the intersection. After this approximation[3] we get from (4.14) the problem

$$\max_{\mathbf{r} \in \bigcap_i K_i^L} \sum_{n=0}^{L-1} r_{\mathbf{g}^L, n} c_n. \tag{4.26}$$

This is now a convex optimization problem of a linear functional over a convex set. By the *positive real lemma* [DLS02], these cone constraints can be equivalently described by semi-positive-definite matrix equalities, s.t. the problem (4.26) is numerically solvable with the MATLAB toolbox SeDuMi [WBV99; Stu99]. The filter is obtained by a spectral factorization of $r_{\mathbf{g}^L}$. Obviously $\mathbf{g}^L$ is not uniquely determined.

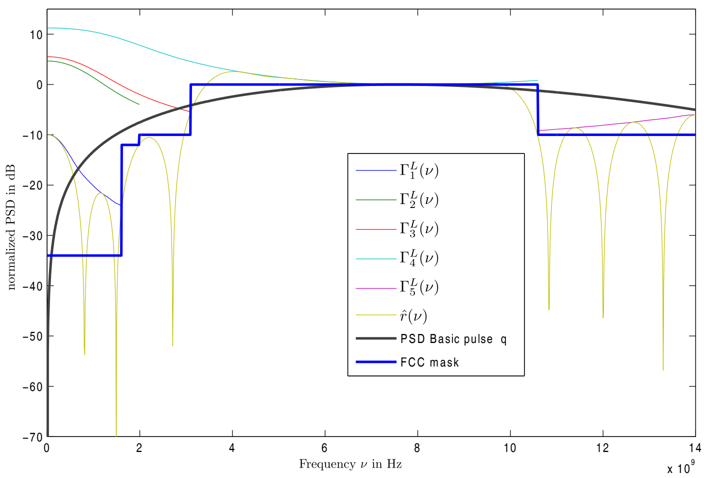


Figure 4.2: Fourier-approximations Γ_i^L of $\mathcal{M}$ for $L = 25$.

Note that this optimization problem can also be seen as the maximization of a local L^2-norm, given as the NESP value, under the constraints of local L^∞-norms.

4.3 Orthogonalization of Pulse Translates

In [Tia+06; DK07] a sequential pulse optimization was introduced, which produces mutually orthogonal pulses $p_m^\circ = q *_{T_0}' \mathbf{g}_m^L$, i.e. $(p_m^\circ, p_n^\circ) = \delta_{nm}$. Here each pulse p_m° is generated by a

[3] The Γ_i^L are approximations to the FCC mask with a certain error. Also, T_0 is now fixed via the frequency range F. If one wants to reduce T_0, one has to reformulate the cones, hence γ_i^L and extend the frequency band constraints. Increasing T_0 above $1/28$GHz is not possible, if one wants to respect the whole mask.

different FIR filter $\mathbf{g}_m^L \in \mathbb{R}^L$, which depends on the previously generated pulses $p_1^\circ, \ldots, p_{m-1}^\circ$ and produces pulses in $L^2([-T_q/2, (L-1)T_0 + T_q/2])$. This approach is similar to the Gram-Schmidt construction in that it is order-dependent, since the first pulse p_1° can be optimally designed to the FCC mask without an orthogonalization constraint. We will now present a new order–independent method to generate from a fixed initial pulse p a set of orthogonal pulses $\{p_m^\circ\}$. Therefore we introduce a new time–shift $T > 0$, namely the PPM shift in (4.4), to generate a set of $N = 2M+1$ translates $\{p(\cdot - mT)\}_{m=-M}^{M}$, i.e. M shifts in each time direction. The orthogonal pulses are then obtained by linear combinations of the translates of the initial pulse p. For a stable embedding of the finite construction we restrict the initial pulses to the set $L_{T_p}^2 := L^2([-T_p/2, T_p/2])$ of centered pulses with finite duration T_p. To study the convergence we need to introduce the concept of regular shift–invariant spaces.

4.3.1 Shift–Invariant Spaces and Riesz-Bases

To simplify notation we scale the time axis so that $T = 1$. Let us now consider the set $\mathscr{S}_0(p) := \operatorname{span}\{p(\cdot - n)\}$ of all finite linear combinations of $\{p(\cdot - n)\}$, which is certainly a subset of L^2. The L^2-closure of $\mathscr{S}_0(p)$ is a *shift–invariant* closed subspace $\mathscr{S}(p) := \overline{\mathscr{S}_0(p)} \subset L^2$, i.e. for each $f \in \mathscr{S}(p)$ also $\{f(\cdot - n)\} \subset \mathscr{S}(p)$. Since $\mathscr{S}(p)$ is generated by a single function p we call it a *principal shift–invariant* (PSI) space and p the *generator* for $\mathscr{S}(p)$. In fact, $\mathscr{S}(p)$ is the smallest PSI closed subspace of L^2 generated by p. Of course not every closed PSI space is of this form [BDR94]. In this chapter we are interested in spaces which are closed under semi-discrete convolutions (4.11) with ℓ^2 sequences, i.e. the space

$$\mathscr{V}(p) := \left\{ p *' \mathbf{c} \;\middle|\; \mathbf{c} \in \ell^2 \right\} \tag{4.27}$$

endowed with the L^2- norm. Note that $\mathscr{V}(p)$ is in general not a subspace or even a closed subspace of L^2 [BDR94; AS02]. But to guarantee stability of our filter design $\mathscr{V}(p)$ has to be closed, i.e. has to be a Hilbert subspace. More precisely, the translates of p have to form a *Riesz basis*.

Definition 15. *Let $\mathscr{H}$ be a Hilbert space. $\{e_n\} \subset \mathscr{H}$ is a Riesz basis for $\overline{\operatorname{span}\{e_n\}}$ if and only if there are constants $0 < A \le B < \infty$, s.t.*

$$A \left\| \mathbf{c} \right\|_{\ell^2}^2 \le \left\| \sum_n c_n e_n \right\|_{\mathscr{H}}^2 \le B \left\| \mathbf{c} \right\|_{\ell^2}^2 \quad \textit{for all} \quad \mathbf{c} \in \ell^2. \tag{4.28}$$

In this case $\overline{\operatorname{span}\{e_n\}}$ becomes a Hilbert-subspace of $\mathscr{H}$. For SI spaces in $L^2 = \mathscr{H}$ we get the following result.

Proposition 10 (Prop.1 in [AU94])**.** *Let $p \in L^2(\mathbb{R})$. Then $\mathscr{V}(p)$ is a closed shift–invariant subspace of L^2 if and only if*

$$A \left\| \mathbf{c} \right\|_{\ell^2}^2 \le \left\| p *' \mathbf{c} \right\|_{L^2}^2 \le B \left\| \mathbf{c} \right\|_{\ell^2}^2 \quad \textit{for all} \quad \mathbf{c} \in \ell^2 \tag{4.29}$$

holds for fixed constants $0 < A \le B < \infty$. Moreover, $\{p(\cdot - n)\}$ is a Riesz-basis for $\mathscr{V}(p)$.

If the generator p fulfills (4.29), then $\mathscr{V}(p) = \mathscr{S}(p)$ by [Jia97] and we call p a *stable generator* and $\mathscr{V}(p)$ a *regular PSI space* [BDR94]. An *orthonormal generator* (*Nyquist pulse*) p° for $\mathscr{V}(p)$ is a generator with $(p^\circ(\cdot - n), p^\circ(\cdot - m)) = \delta_{mn}$ for all $n, m \in \mathbb{Z}$ [And05, Def. 2.2-2]. Benedetto and Li [BL98] showed that the stability and orthogonality of a generator $p \in L^2$ can be described by the absolute $[0,1]$-integrable periodic function $\Phi_p \in L^1([0,1])$ of p defined for ν almost everywhere (a.e.) as

$$\Phi_p(\nu) := \sum_k |\hat{p}(\nu + k)|^2 . \tag{4.30}$$

They could show the following characterization [Chr03; BL98].

Theorem 11 (Th. 7.2.3 in [Chr03])**.** *A function $p \in L^2$ is a stable generator for $\mathscr{V}(p)$ if and only if there exists $0 < A \leq B < \infty$ such that*

$$A \leq \Phi_p(\nu) \leq B \qquad \text{for } \nu \text{ a.e.} \tag{4.31}$$

and is an orthonormal generator for $\mathscr{V}(p)$ if and only if

$$\Phi_p(\nu) = 1 \qquad \text{for } \nu \text{ a.e..} \tag{4.32}$$

Proof. For a proof see Th. 7.2.3. (ii) and (iii) in [Chr03]. In our special case we have $B = 1$. Note, that the Riesz sequence and orthonormal sequence are bases for their closed span, meaning that in our case $\mathscr{S}(p) = \mathscr{V}(p)$. □

Due to this characterization in frequency there is a simple “orthogonalization trick” for a stable generator given in (4.1), which was found by Meyer, Mallat, Daubechies and others [Mey86],[Chr03, Prop. 7.3.9]. Unfortunately, this does not provide an a priori construction in the time domain and does not lead to a support control of the orthonormal generator in time, as necessary for UWB-IR.

Contrary to an approximation in the frequency-domain we approach an approximation in time-domain via the Löwdin transformation. We will show that in the limit the Löwdin transformation for shift–sequences is in fact given in frequency by the orthogonalization trick (4.1). By using finite section methods we establish an approximation method in terms of the discrete Fourier transform (DFT) to allow an easy computation. Furthermore, we show that the Löwdin construction for stable generators is unique and optimal in the L^2-distance among all orthonormal generators and corresponds to the canonical construction of so called tight frames (given later).

4.3.2 Löwdin Orthogonalization for Finite Dimensions

Since the Löwdin transformation is originally defined for a finite set of linearly independent elements in a Hilbert space $\mathscr{H}$, we will use the *finite section method* to derive a stable approximation to the infinite case. For this we consider for any $M \in \mathbb{N}$ the symmetric orthogonal projection $\mathbf{P}_M$ from ℓ^2 to $\ell^2_M = \{\mathbf{c} \in \ell^2 | \operatorname{supp} \mathbf{c} \subset [M, M]\}$ defined for $\mathbf{c} \in \ell^2$ by $\mathbf{P}_M \mathbf{c} := \mathbf{c}^M = (0, \ldots, 0, c_{-M}, \ldots, c_M, 0, \ldots, 0)$. Then the finite section $\mathbf{G}_M$ of the infinite dimensional Gram matrix $\mathbf{G}$ of $p \in L^2$, given by

$$(\mathbf{G})_{nm} := \langle p(\cdot - m), p(\cdot - n)\rangle \quad \text{for} \quad n, m \in \mathbb{Z}, \tag{4.33}$$

can be defined as $\mathbf{G}_M := \mathbf{P}_M \mathbf{G} \mathbf{P}_M$, see [Grö01, Prop. 5.1.5]. If p satisfies (4.28) and if we restrict the semi-discrete convolution $p *' \mathbf{c}$ to ℓ^2_M, we obtain a $2M + 1$ dimensional Hilbert subspace $\mathscr{V}^M(p)$ of $\mathscr{V}(p) = \mathscr{H}$. Then the unique linear operation B^M, which generates from $\{p(\cdot - m)\}_{m=-M}^M$ an *orthonormal basis* (ONB) $\{p_m^{\circ,M}\}_{m=-M}^M$ for $\mathscr{V}^M(p)$ and simultaneously minimizes

$$\sum_{m=-M}^{M} \left\| B^M p(\cdot - m) - p(\cdot - m) \right\|_{L^2}^2 \tag{4.34}$$

is given by the (symmetric) Löwdin transformation [Löw50; Löw70; Lan36; Jun07] and can be represented in matrix form as

$$p_m^{\circ,M} := B^M p(\cdot - m) = \sum_{n=-M}^{M} (\mathbf{G}_M^{-\frac{1}{2}})_{mn} p(\cdot - n) \quad \text{for all} \quad m \in [-M, M], \tag{4.35}$$

where we call each $p_m^{\circ,M}$ a *Löwdin orthogonal (LO) pulse* or *Löwdin pulse*. Here $\mathbf{G}_M^{-\frac{1}{2}}$ denotes the (canonical) inverse square-root (restricted to ℓ^2_M) of $\mathbf{G}_M$. Note that $\mathbf{G}_M^{-\frac{1}{2}}$ is not equal to $\mathbf{P}_M \mathbf{G}^{-\frac{1}{2}} \mathbf{P}_M$. Since the sum in (4.35) is finite, the definition of the Löwdin pulses is also pointwise well-defined. In the next section, we will see that this is a priori not true for the infinite case. If we identify the corresponding m'th row of the inverse square-root of $\mathbf{G}_M$ with vectors $\mathbf{h}_m^M = ((\mathbf{G}_M^{-\frac{1}{2}})_{m-M}, \ldots, (\mathbf{G}_M^{-\frac{1}{2}})_{m+M})$ we can describe (4.35) by a FIR filter bank as

$$p_m^{\circ,M} = p *' \mathbf{h}_m^M \quad , \quad m \in [-M, M]. \tag{4.36}$$

Unfortunately, none of these Löwdin pulses is a shift-orthogonal pulse, which would be necessary for an OOPPM transmission. In the next section we will thus show that the Löwdin orthogonalization converges for M to infinity to an IIR filter $\mathbf{b}$ given as the centered row of $\mathbf{G}^{-\frac{1}{2}}$. This IIR filter generates then a shift-orthogonal pulse, namely the Nyquist pulse defined in (4.1). Hence the Löwdin orthogonalization (4.35) provides an approximation to our OOPPM design. In the following we will investigate its stability, i.e. its convergence property.

4.3.3 Stability and Approximation

In this section we investigate the limit of the Löwdin orthogonalization in (4.35) for translates (time-shifts) of the optimized pulse p with time duration $T_p < \infty$ where we further assume that p is a bounded stable generator. If we set $K := \lfloor T_p \rfloor$ then certainly $p \in L^2_K$. In this case the auto–correlation of p

$$r_p(t) := (p * \overline{p}_-)(t) = \int_{\mathbb{R}} p(\tau)\overline{p(\tau - t)} d\tau, \tag{4.37}$$

with the time reversal $p_-(t) := p(-t)$ is a compactly supported bounded function on $[-K, K]$. Due to the Poisson summation formula we can represent Φ_p almost everywhere by the Fourier series (4.15) ($T_0 = 1$) of the samples $\{r_p(n)\}$

$$\Phi_p(\nu) = \sum_{n=-\infty}^{\infty} r_p(n) e^{-2\pi i n \nu} = \sum_{n=-K}^{K} (\mathbf{G})_{n0} e^{-2\pi i n \nu}, \tag{4.38}$$

which is the symbol of the Toeplitz matrix $\mathbf{G}$, since we have from (4.33) and (4.37) that $r_p(n - m) = (\mathbf{G})_{nm}$. Moreover the symbol is continuous since the sum is finite due to the compactness of r_p.

On the other hand the initial pulse p is a Wiener function[4] [Grö01, Def. 6.1.1] so that Φ_p defines a continuous function and condition (4.31) holds pointwise [AST01, Prop.1]. Since both sides in (4.38) are identical a.e. they are identical everywhere by continuity (see also [Grö01, p.105]). Thus, the spectrum of $\mathbf{G}$ is continuous, strictly positive and bounded by the Riesz bounds. Hence the inverse square-roots of $\mathbf{G}$ and $\mathbf{G}_M$ exists s.t. for any $M \in \mathbb{N}$ (by Cauchy's interlace theorem, [BG05, Th. 9.19])

$$A\mathbf{1}_M \leq \mathbf{G}_M \leq B\mathbf{1}_M \quad \text{and} \quad \frac{1}{B}\mathbf{1}_M \leq \mathbf{G}_M^{-1} \leq \frac{1}{A}\mathbf{1}_M, \tag{4.39}$$

where $\mathbf{1}_M$ denotes the identity on ℓ^2_M. Now we can approximate the Gram matrix by STRANG's circulant preconditioner [Str86], s.t. the diagonalization is given by a discrete Fourier transform (DFT) [Dav79]. To get a continuous formulation of the approximated Löwdin pulses we use the ZAK transform [Jan88], given for a continuous function f as

$$(\mathrm{Z}f)(t, \nu) := \sum_{n \in \mathbb{Z}} f(t - n) e^{2\pi i n \nu} \quad \text{for} \quad t, \nu \in \mathbb{R}. \tag{4.40}$$

Our main result, already presented in [WJT10a; WJT10b] but now with full proof, is the following theorem.

Theorem 12. *Let $K \in \mathbb{N}$ and $p \in L^2_K$ be a continuous stable generator for $\mathscr{V}(p)$. Then we can approximate the limit set of the Löwdin pulses $\{p^\circ_m\}$ by a sequence of finite function sets*

[4] Wiener functions are locally bounded in L^∞ and globally in ℓ^1.

$\{\tilde{p}_m^{\circ,M}\}_{m=-M}^{M}$, *which are approximate Löwdin orthogonal (ALO). The functions* $\tilde{p}_m^{\circ,M}$ *are given pointwise for* $M \geq K$ *and* $m \in \{-M,\dots,M\}$ *by the Zak transform as*

$$\tilde{p}_m^{\circ,M}(t) := \begin{cases} \frac{1}{2M+1} \sum_{l=0}^{2M} \frac{e^{\frac{-2\pi i m l}{2M+1}} (\mathrm{Z}p)(t, \frac{l}{2M+1})}{\sqrt{(\mathrm{Z}r_p)(0, \frac{l}{2M+1})}} & |t| \leq M - \frac{K}{2} \\ 0 & \textit{else} \end{cases}, \tag{4.41}$$

such that for each $m \in \mathbb{Z}$

$$\tilde{p}_m^{\circ}(t) = \lim_{M\to\infty} \tilde{p}_m^{\circ,M}(t) \tag{4.42}$$

converges pointwise on compact sets. The limit in (4.42) *can be stated as*

$$\hat{\tilde{p}}^{\circ}(\nu) = \hat{p}(\nu) \cdot (\Phi_p(\nu))^{-\frac{1}{2}} \tag{4.43}$$

for $\nu \in \mathbb{R}$ *in the frequency-domain. Hence the Löwdin generator* $\tilde{p}^{\circ} := \tilde{p}_0^{\circ}$ *is an orthonormal generator for* $\mathscr{V}(p)$.

Proof. The proof consists of two parts. In the first part we derive an straightforward finite construction in the time domain to obtain time-limited pulses (4.41) being approximations to the Löwdin pulses. Using Strang's circulant preconditioner the ALO pulses can be easily derived in terms of DFTs. In the second part we will then show that this finite construction is indeed a stable approximation to the Nyquist pulse. Here we need pointwise convergence, i.e. convergence in ℓ^{∞} (the set of bounded sequences). Finally, to establish the shift-orthogonality we use properties of the Zak transform.

Since the inverse square-root of a $N \times N$ Toeplitz matrix is hard to compute, we approximate for any $M \geq K$ the Gram matrix $\mathbf{G}_M$ by using Strang's circulant preconditioner $\tilde{\mathbf{G}}_M$ [Str86; CJ07]. Moreover, the Gram matrix is hermitian and banded such that we can define the elements of the first row by [Gra06, (4.19)] as

$$(\tilde{\mathbf{G}}_M)_{0n} := \begin{cases} r_p(n) & n \in [0, K] \\ r_p(N-n) & n \in [N-K, N-1] \\ 0 & \text{else} \end{cases}. \tag{4.44}$$

Here we abbreviate $N := 2M + 1$. The crucial property of Strang's preconditioner $\tilde{\mathbf{G}}_M$ is the fact that the eigenvalues $\lambda_l(\tilde{\mathbf{G}}_M)$ are sample values of the symbol Φ_p in (4.30). This special property is in general not valid for other circulant preconditioners [CJ07]. To see this, we derive the eigenvalues by [Gra06, Theorem 7] as

$$\tilde{\lambda}_l^M := \lambda_l(\tilde{\mathbf{G}}_M) = \sum_{n=0}^{K} r_p(n) e^{-2\pi i n \frac{l}{N}} + \sum_{n=N-K}^{N-1} r_p(N-n) e^{-2\pi i n \frac{l}{N}} \quad \text{for} \quad l \in [0, 2M] \tag{4.45}$$

by inserting the first row of $\tilde{\mathbf{G}}_M$ given in (4.44). If we set in the second sum $n' = n - N$ we get from (4.38)

$$\tilde{\lambda}_l^M = \sum_{n=-K}^{K} r_p(n) e^{-2\pi i l \frac{n}{N}} = \Phi_p\left(\frac{l}{2M+1}\right). \tag{4.46}$$

Since p is compactly supported the symbol Φ_p is continuous and the second equality in (4.46) holds pointwise. Moreover, the Riesz bounds (4.31) of p guarantee that $\tilde{\mathbf{G}}_M$ is strictly positive and invertible for any M. Now we are able to define the ALO pulses in matrix notation by setting[5] in $\mathbf{p}^M(t) := \{p(t-n)\}_{n=-M}^{M} = (p(t+M), \dots, p(t-M))^T$ for any $t \in \mathbb{R}$

$$\tilde{\mathbf{p}}^{\circ,M}(t) := \tilde{\mathbf{G}}_M^{-\frac{1}{2}} \mathbf{p}^M(t) = \mathbf{F}_M \tilde{\mathbf{D}}_M^{-\frac{1}{2}} \mathbf{F}_M^* \mathbf{p}^M(t), \tag{4.47}$$

since the circulant matrix $\tilde{\mathbf{G}}_M = \mathbf{F}_M \tilde{\mathbf{D}}_M \mathbf{F}_M^*$ can be written by the unitary $N \times N$ DFT matrix $\mathbf{F}_M$, with

$$[\mathbf{F}_M]_{nm} := \frac{1}{\sqrt{N}} e^{-2\pi i \frac{nm}{N}} \quad \text{with} \quad n, m \in \{0, \dots, 2M\} \tag{4.48}$$

and the diagonal matrix $\tilde{\mathbf{D}}_M$ of the eigenvalues of $\tilde{\mathbf{G}}_M$. Let us start in (4.47) from the right by applying the IDFT matrix $\mathbf{F}_M^*$, then we get for any kth component with $k \in \{0, \dots, 2M\}$

$$[\mathbf{F}_M^* \mathbf{p}^M(t)]_k = \frac{1}{\sqrt{N}} \left(\sum_{n=0}^{M} e^{2\pi i \cdot \frac{n}{N} k} \cdot p(t+M-n) + \sum_{n=M+1}^{2M} e^{2\pi i \cdot \frac{n}{N} k} \cdot p(t+M-n) \right) \tag{4.49}$$

$$\stackrel{\substack{j=n+M \\ \downarrow}}{=} \frac{1}{\sqrt{N}} \sum_{j=-M}^{M} e^{2\pi i \cdot \frac{j+M}{N} k} \cdot p(t-j). \tag{4.50}$$

Next we multiply with the components $[\tilde{\mathbf{D}}_M^{-\frac{1}{2}}]_{kl} = \delta_{kl} / \sqrt{\tilde{\lambda}_l^M}$ of the inverse square-root of the diagonal matrix $\tilde{\mathbf{D}}_M$

$$\left[\tilde{\mathbf{D}}_M^{-\frac{1}{2}} \mathbf{F}_M^* \mathbf{p}^M(t)\right]_l = \frac{1}{\sqrt{N}} \sum_{j,k} \left(\frac{\delta_{kl}}{\sqrt{\tilde{\lambda}_l^M}} e^{2\pi i \frac{j+M}{N} k} \cdot p(t-j) \right) = \frac{1}{\sqrt{N}} \left(\sum_{j=-M}^{M} \frac{e^{2\pi i \frac{j+M}{N} l} \cdot p(t-j)}{\sqrt{\tilde{\lambda}_l^M}} \right). \tag{4.51}$$

In the last step we evaluate the DFT at $m \in \{0, \dots, 2M\}$

$$\tilde{p}_{m-M}^{\circ,M}(t) = [\tilde{\mathbf{p}}^{\circ,M}(t)]_m = \frac{1}{N} \sum_{l=0}^{2M} \left(\frac{e^{-2\pi i m \cdot \frac{l}{N}}}{\sqrt{\tilde{\lambda}_l^M}} \sum_{j=-M}^{M} p(t-j) e^{2\pi i l \frac{j+M}{N}} \right) \tag{4.52}$$

$$= \frac{1}{N} \sum_{l=0}^{2M} e^{-2\pi i l \frac{m-M}{N}} \frac{\sum_{j=M}^{M} p(t-j) e^{2\pi i l \cdot \frac{j}{N}}}{\sqrt{(\mathrm{Z}r_p)(0, \frac{l}{N})}}. \tag{4.53}$$

[5]By slight abuse of our notation we understand in this section any matrix as an $N \times N$ matrix and $\mathbf{p}^M(t), \tilde{\mathbf{p}}^{\circ,M}(t)$ as N-dim. vectors for any $t \in \mathbb{R}$.

where we used the Zak transform (4.40) of r_p to express the eigenvalues $\tilde{\lambda}_l^M$. In the next step we extend the DFT sum of the numerator in (4.53) to an infinite sum. This is possible since $p(\cdot - k)$ always has the same support length K for each $k \in \mathbb{Z}$. Thus, for all $|t| > M - \frac{K}{2}$ the non-zero sample values are shifted in the kern of $\mathbf{P}_M$; hence $\mathbf{p}^M(t) = 0$. On the other hand for $|t| \leq M - \frac{K}{2}$ any shift $|j| > M$ results in $p(t-j) = 0$. If we also set for each M the index $k := m - M$ in (4.53) then the continuous ALO pulses can be written as

$$\tilde{p}_k^{\circ,M}(t) = \begin{cases} \frac{1}{N} \sum_l e^{-2\pi i l \frac{k}{N}} \frac{(\mathrm{Z}p)(t,\frac{l}{N})}{\sqrt{(\mathrm{Z}r_p)(0,\frac{l}{N})}} & |t| \leq M - \frac{K}{2} \\ 0 & \text{else} \end{cases}. \tag{4.54}$$

This defines for each k the operation $\tilde{B}_k^M$ by $\tilde{p}_k^{\circ,M} = \tilde{B}_k^M p$. For each $t \in \mathbb{R}$ the term $\frac{(\mathrm{Z}p)(t,\nu)}{\sqrt{(\mathrm{Z}\tilde{r})(0,\nu)}}$ is a continuous function in ν since the Zak transforms are finite sums of continuous functions and the nominator vanishes nowhere. This is guaranteed by the positivity and continuity of Φ_p due to (4.39). Hence the ALO pulses are continuous as well.

The second part of the proof shows the convergence of our finite construction to a Nyquist pulse. Therefore we use the *finite section method* for the Gram matrix. Gray showed [Gra06, Lemma 7] that

$$\left\| \tilde{\mathbf{G}}_M - \mathbf{G}_M \right\|_w \to 0 \tag{4.55}$$

as $M \to \infty$ in the weak norm $\|A\|_w^2 := 1/N \sum_j \lambda_j^2(A)$ implying weak convergence of the operators. Since $\tilde{\mathbf{G}}_M$ is strictly positive for each $M \in \mathbb{N}$ we get by [GGC08]

$$\left\| \tilde{\mathbf{G}}_M^{-\frac{1}{2}} - \mathbf{G}_M^{-\frac{1}{2}} \right\|_w \to 0. \tag{4.56}$$

Unfortunately this does not provides a strong convergence, which is necessary to state convergence in ℓ^2

$$\tilde{\mathbf{G}}_M^{-\frac{1}{2}} \mathbf{P}_M \mathbf{c} \to \mathbf{G}^{-\frac{1}{2}} \mathbf{c} \quad \text{for any } \mathbf{c} \in \ell^2. \tag{4.57}$$

However, from [SJB03] *finite strong convergence* can be ensured, i.e. convergence of (4.57) for all $\mathbf{c} \in \ell_{M'}^2$ for each $M' \in \mathbb{N}$. But for any $t \in \mathbb{R}$ there exists an M' sufficiently large, due to the compact support property of p, such that $\mathbf{c} = \mathbf{p}(t) := \{p(t-n)\} \in \ell_{M'}^2$.This is in fact sufficient, since it implies pointwise convergence in $\ell_{M'}^\infty$ of (4.57), i.e. component-wise convergence for each $t \in \mathbb{R}$. Let us take for each $t \in \mathbb{R}$ the number $M' \in \mathbb{N}$ such that $\max\{|t|, K\} \leq M'$. Then we can define the limit of the kth component as

$$\tilde{p}_k^\circ(t) := \lim_{M \to \infty} \tilde{p}_k^{\circ,M}(t). \tag{4.58}$$

If we define $\Delta\nu = \frac{1}{2M+1}$ and $\nu_l = l\Delta\nu$ we can write for (4.58) by inserting (4.54)

$$\lim_{M \to \infty} \tilde{p}_k^{\circ,M}(t) = \lim_{M \to \infty} \sum_{l=0}^{2M} \frac{e^{-2\pi i k \nu_l} (\mathrm{Z}p)(t,\nu_l)}{\sqrt{(\mathrm{Z}r_p)(0,\nu_l)}} \Delta\nu. \tag{4.59}$$

Using the quasi-periodicity [Jan88, (2.18),(2.19)] of the Zak transform for $k \in \mathbb{Z}$ we have for any $t, \nu \in \mathbb{R}$

$$(\mathrm{Z}p(\cdot - k))(t,\nu) = (\mathrm{Z}p)(t-k,\nu) = e^{-2\pi i \nu k}(\mathrm{Z}p)(t,\nu). \tag{4.60}$$

We can express the partial sum on the right hand side of (4.59) in the limit as a Riemann integral

$$\tilde{p}_k^\circ(t) := \lim_{M\to\infty} \tilde{p}_k^{\circ,M}(t) = \int_0^1 \frac{(\mathrm{Z}p(\cdot - k))(t,\nu)}{\sqrt{(\mathrm{Z}r_p)(0,\nu)}} d\nu. \tag{4.61}$$

This shows that $\tilde{B}_k^M p$ converge pointwise for $M \to \infty$ to $\tilde{B}_k p = \tilde{B}_0 p(\cdot - k) = \tilde{p}_k^\circ$ for each k. The sequence $\{\tilde{p}_k^\circ\}$ is then generated by shifts of $\tilde{p}^\circ := \tilde{p}_0^\circ$ since the shift operation commutes with $\tilde{B} := \tilde{B}_0$. This in turn commutes with the Zak transformation, i.e. we have

$$\begin{aligned}(\tilde{B}p)(t-k) &= \int \frac{(\mathrm{Z}p)(t-k,\nu)}{\sqrt{(\mathrm{Z}r_p)(0,\nu)}} d\nu = \int \frac{(\mathrm{Z}p(\cdot-k))(t,\nu)}{\sqrt{(\mathrm{Z}r_p)(0,\nu)}} d\nu & (4.62)\\ &= (\tilde{B}p(\cdot-k))(t). & (4.63)\end{aligned}$$

From (4.61) it is now easy to show that $\tilde{p}^\circ$ is an orthonormal generator. We write the left hand side of (4.61) in the Zak domain, by applying the Zak transformation[6] to $\tilde{p}^\circ$

$$(\mathrm{Z}\tilde{p}^\circ)(t,\nu) = \frac{(\mathrm{Z}p)(t,\nu)}{\sqrt{(\mathrm{Z}r_p)(0,\nu)}}. \tag{4.64}$$

If we multiple (4.64) by the exponential and integrate over the time we yield for every $\nu \in \mathbb{R}$

$$\int_0^1 e^{-2\pi i \nu t}(\mathrm{Z}\tilde{p}^\circ)(t,\nu) dt = \int_0^1 \frac{e^{-2\pi i \nu t} \cdot (\mathrm{Z}p)(t,\nu)}{\sqrt{\mathrm{Z}r_p(0,\nu)}} dt. \tag{4.65}$$

Since $\Phi_p = (\mathrm{Z}r_p)(0,\cdot)$ is time-independent we get the "orthogonalization trick" (4.1) by using in (4.65) the inversion formula [Jan88, (2.30)] of the Zak transform

$$\hat{\tilde{p}}^\circ(\nu) = \hat{p}(\nu) \cdot (\Phi_p(\nu))^{-\frac{1}{2}} = \hat{p}^\circ(\nu). \tag{4.66}$$

Again, this is also defined pointwise since the right hand side is continuous in ν. It can now be easily verified that $\tilde{p}^\circ$ fulfills the Nyquist condition (4.32), which shows that $\tilde{p}^\circ$ is an orthonormal generator for $\mathscr{V}(p)$. □

Remark. Note, that relation (4.60) induces a time-shift. To apply this to the ALO pulses in (4.54) the time domain has to be restricted further. Hence the ALO pulses do not have global

[6] A similar result is also known in the context of Gabor frames, see also [Grö01, p. 8.3].

shift character for finite $M \in \mathbb{N}$, but locally, i.e. $\tilde{p}_k^{\circ,M}$ shifted back to the center matches $\tilde{p}^{\circ,M}$ for $t \in [-M + \frac{K}{2} + |k|, M - \frac{K}{2} - |k|]$:

$$\tilde{p}_k^{\circ,M}(t+k) = \frac{1}{N} \sum_l e^{-2\pi i \frac{l}{N} k} \frac{\sum_{n=-M}^{M} p(t+k-n) e^{2\pi i \frac{l}{N} n}}{\sqrt{\tilde{\lambda}_l^M}} \tag{4.67}$$

$$= \frac{1}{N} \sum_l \frac{\sum_{n=-M}^{M} p(t+k-n) e^{2\pi i \frac{l}{N}(n-k)}}{\sqrt{\tilde{\lambda}_l^M}} \tag{4.68}$$

$$= \frac{1}{N} \sum_l \frac{\sum_{n'=-M-k}^{M-k} p(t-n') e^{2\pi i \frac{l}{N} n'}}{\sqrt{\tilde{\lambda}_l^M}}. \tag{4.69}$$

Since $p(t+M+|k|) = 0$ and $p(t-M-|k|) = 0$ for $|k| < M$ and $|t| \le M - \frac{K}{2} - |k|$, we end up with

$$\tilde{p}_k^{\circ,M}(t+k) = \frac{1}{N} \sum_l \frac{\sum_{n=-M}^{M} p(t-n) e^{2\pi i \frac{l}{N} n}}{\sqrt{\tilde{\lambda}_l^M}} = \tilde{p}^{\circ,M}(t). \tag{4.70}$$

For all $|k| < M$ the ALO pulses have the same shape in the window $|t| \le M - \frac{K}{2} - |k|$ if we shift them back to the origin.

Moreover, the ALO pulses are all continuous on the real line, since they are a finite sum of continuous functions by definition (4.47). Hence each ALO pulse goes continuously to zero at the support boundaries. So far it is not clear whenever $\tilde{p}^\circ$ is continuous or not. Nevertheless its spectrum $\hat{\tilde{p}}^\circ$ is continuous and so we can state $\tilde{p}^\circ = p^\circ$ almost everywhere. Hence the orthogonalization trick defines the Nyquist pulse p° only in an L^2 sense.

4.4 Discussion of the Results

In this section we will discuss now the properties of our OOPPM design for UWB, i.e. the optimization and orthogonalization, which can be completely described by an IIR filtering process. First we will relate the Löwdin orthogonalization to the canonical tight frame construction. Afterwards we will show in Section 4.4.2 that the Löwdin transform yields the orthogonal generator with the minimal L^2-difference to the initial optimized pulse. This is the same optimality property as for canonical tight frames [JS02]. But such an energy optimality does not guarantee FCC compliance. So we will discuss in Section 4.4.3 the influence of a perfect orthogonalization to the FCC optimization. Finally, we will discuss the implementation of a perfect orthogonalization by FIR filtering.

4.4.1 Relation Between Tight Frames and ONBs

Any Riesz basis $\{p_k\}$ for a Hilbert space $\mathscr{H}$ is also a *exact frame* for $\mathscr{H}$ with the frame operator S defined by

$$\mathrm{S} : \mathscr{H} \to \mathscr{H}, \quad f \mapsto \mathrm{S}f = \sum_k \langle f, p_k \rangle_{\mathscr{H}} p_k, \tag{4.71}$$

where the frame bounds are given by the Riesz bounds $0 < A \le B < \infty$ of $\{p_k\}$ [Chr03, Th. 5.4.1, 6.1.1], i.e.

$$A \, \|f\|^2_{\mathscr{H}} \le \langle Sf, f \rangle_{\mathscr{H}} \le B \, \|f\|^2_{\mathscr{H}} \quad \text{for any} \quad f \in \mathscr{H}. \tag{4.72}$$

Here $\langle \cdot, \cdot \rangle_{\mathscr{H}}$ denotes the inner product in $\mathscr{H}$ and $\|\cdot\|_{\mathscr{H}}$ the induced norm. Since S is bounded and invertible, i.e. the inverse operator exists and is bounded [Chr03], we can write each $f \in \mathscr{H}$ as

$$f = \mathrm{S}\mathrm{S}^{-1} f = \sum_k \langle \mathrm{S}^{-1} f, p_k \rangle_{\mathscr{H}} p_k. \tag{4.73}$$

In this case the Löwdin orthonormalization corresponds to the *canonical tight frame construction*.

Lemma 6. *Let the sequence $\{p_k\}$ be a Riesz basis for the Hilbert space $\mathscr{H} := \overline{\mathrm{span}\{p_k\}}$ and $\mathbf{G}$ its Gram matrix. Then the canonical tight frame $\{p_k^\circ\}$ is given for each $k \in \mathbb{Z}$ by:*

$$p_k^\circ := S^{-\frac{1}{2}} p_k = \sum_l (\mathbf{G}^{-\frac{1}{2}})_{kl} \, p_l \tag{4.74}$$

in an L^2-sense.[7]

Proof. Let $a, b \in \mathbb{R}$ with $a + b = -1$. Then S^a and S^b, defined by the spectral theorem, are also positive and self-adjoint on $\mathscr{H}$. Moreover for each f we have the following unique representation $f = \sum_k c_k p_k$ with $\mathbf{c} \in l^2(\mathbb{Z})$ due to the Riesz basis property. For $f = p_l$ in (4.73) we get

$$p_l = \sum_k \langle S^{-1} p_l, p_k \rangle_{\mathscr{H}} = \sum_k \langle S^a p_l, S^b p_k \rangle_{\mathscr{H}} p_k \tag{4.75}$$

since $S^a p_l, S^b p_k \in \mathscr{H}$ there exist unique sequences $\mathbf{c}_l, \mathbf{d}_k$ s.t. $S^a p_l = \sum_\alpha c_{l\alpha} p_\alpha, S^b p_k = \sum_\beta d_{k\beta} p_\beta$. Hence we get

$$p_l = \sum_k \left(\sum_\alpha c_{l\alpha} p_\alpha, \sum_\beta d_{k\beta} p_\beta \right) p_k \tag{4.76}$$

$$= \sum_k \sum_{\alpha,\beta} c_{l\alpha} \bar{d}_{k\beta} \langle p_\alpha, p_\beta \rangle_{\mathscr{H}} p_k \tag{4.77}$$

$$= \sum_k \sum_{\alpha,\beta} (\mathbf{C})_{l\alpha} (\mathbf{G})_{\alpha\beta} (\mathbf{D}^*)_{\beta k} p_k = \sum_k (\mathbf{C}\mathbf{G}\mathbf{D}^*)_{lk} p_k \tag{4.78}$$

[7] This statement was already given without further explanation by Y. Meyer in [Mey86] equation (3.3). Note that Y. Meyer used condition (3.1) and (3.2) in [Mey86] which are equivalent to the Riesz basis condition.

where $c_{l\alpha}$ and $d_{\beta k}$ are the coefficients of the biinfinite matrices $\mathbf{C}$ resp. $\mathbf{D}$. Since for each $l \in \mathbb{Z}$ we have $\sum_k \delta_{lk} p_k = p_l$ and $\{(\mathbf{CGD}^*)_{lk}\}, \{\delta_{lk}\} \in \ell^2$, we get $0 = \sum_k ((\mathbf{CGD}^*)_{lk} - \delta_{lk}) p_k$ for each $l \in \mathbb{Z}$. So by [Chr03, Th.6.1.1(vii)] we can conclude that $(\mathbf{CGD}^*)_{lk} = \delta_{lk}$ for all $l, k \in \mathbb{Z}$ and get

$$\mathbf{CGD}^* = \mathbf{1} \Leftrightarrow \mathbf{G} = \mathbf{C}^{-1}(\mathbf{D}^*)^{-1} \Leftrightarrow \mathbf{G}^{-1} = \mathbf{D}^*\mathbf{C}. \tag{4.79}$$

Obviously $\mathbf{D}$ and $\mathbf{C}$ are not an unique decomposition of S^{-1}, since a and b are not. If $a = b = -\frac{1}{2}$, we have $\mathbf{D}^* = \mathbf{C}^*$ and hence $\mathbf{G}^{-\frac{1}{2}} = \mathbf{C} = \mathbf{D}$. This establishes (4.74) in an L^2-sense. □

If we now set $p_n := p(\cdot - n) \in L^2$ the Riesz-basis is generated by shifts of a stable generator and $\mathscr{H} = \mathscr{V}(p)$ becomes a principal shift-invariant (PSI) space, which is a separable Hilbert subspace of L^2 as discussed in Section 4.3.1. The canonical tight frame construction then generates a shift-orthonormal basis, i.e. an orthonormal generator. The reason is that shift-invariant frames and Riesz bases are the same in regular shift-invariant spaces [CCK01, Th.2.4]. So any frame becomes a Riesz basis (exact frame) and any tight frame an ONB (exact tight frame). Hence for regular PSI spaces there exists no redundancy for frames. This generalize the Löwdin transform for generating a Nyquist pulse to any stable generator p.

From Meyer [Mey86] we know that (4.74) can be written in frequency domain as the orthogonalization trick. Therefore the limit of the Löwdin transformation $\tilde{B} = B \colon \mathscr{V}(p) \to \mathscr{V}(p)$

$$f \mapsto Bf = \int_0^1 \frac{(\mathrm{Z}f)(\cdot, \nu)}{\sqrt{\mathrm{Z}r_p)(0, \nu)}} d\nu \tag{4.80}$$

equals the inverse square-root of the frame operator in (4.74).

4.4.2 Optimality of the Löwdin Orthogonalization

Janssen and Strohmer have shown in [JS02] that the canonical tight-frame construction of Gabor frames for L^2 is via Ron-Shen duality equivalent to an ONB construction on the adjoint time-frequency lattice. Furthermore they have shown that among all tight Gabor frames, the canonical construction yields this particular generator with minimal L^2-distance to the original one. However, for SI spaces this optimality of the Löwdin orthogonalization has to be proved otherwise. To prove this we use the structure of regular PSI spaces.

Theorem 13. *The unique orthonormal generator with the minimal L^2 distance to the normalized stable generator $p \in L^2$ for $\mathscr{V}(p)$ is given by the Löwdin generator p°.*

Proof. Let us first note that $\mathscr{V}(p)$ is a regular SI space since p is a stable generator. This has as consequence that frames are Riesz bases for $\mathscr{V}(p)$ [RS94, Th. 2.2.7 (e)]. So any element $f \in \mathscr{V}(p) = p *' \mathbf{c}$ is uniquely determined by an ℓ^2 sequence $\mathbf{c}$. By the Riesz–Fischer Theorem this sequence $\mathbf{c}$ defines by its Fourier series a unique $L^2([0,1])$-function $\tau = \hat{\mathbf{c}}$. Hence, the Fourier transform of any $f \in \mathscr{V}(p)$ is represented uniquely by τ as $\hat{f} = \tau\hat{p}$, see also [BDR94, Th.2.10(d)]. On the other hand f is an orthonormal generator if and only if $\Phi_f = 1$ a.e.. By using the periodicity of τ we get

$$\Phi_f = \sum_k |\hat{p}(\cdot - k)|^2 \, |\tau(\cdot - k)|^2 \tag{4.81}$$

$$= |\tau|^2 \sum_k |\hat{p}(\cdot - k)|^2 = |\tau|^2 \cdot \Phi_p = 1. \tag{4.82}$$

Thus, we have $|\tau| = 1/\sqrt{\Phi_p}$ almost everywhere. Let us set $\tilde{\tau} := 1/\sqrt{\Phi_p}$ a.e. and a complex periodic phase function $\phi := e^{i\alpha(\cdot)} : \mathbb{R} \to \{ z \in \mathbb{C} \mid |z| = 1\}$ with $\alpha : \mathbb{R} \to [0, 2\pi]$ 1-periodic and measurable. Then any function $\tau \in L^2([0,1])$ which satisfy (4.82) a.e. is given by $\tau = \tilde{\tau} \cdot \phi$ a.e.. The L^2-distance is then given by

$$\|p - f\|_{L^2}^2 = \left\|\hat{p} - \hat{f}\right\|_{L^2}^2 = \|\hat{p} - \tau\hat{p}\|_{L^2}^2 = \|p\|_{L^2}^2 + \|\tau\hat{p}\|_{L^2}^2 - \int_{\mathbb{R}} \tau \, |\hat{p}|^2 - \int_{\mathbb{R}} \overline{\tau} \, |\hat{p}|^2 \tag{4.83}$$

$$= 2 - \int_{\mathbb{R}} (\tau + \overline{\tau}) \, |\hat{p}|^2 = 2 - 2 \int_{\mathbb{R}} \cos(\alpha) \; |\hat{p}|^2 \, \Phi_p^{-\frac{1}{2}} \tag{4.84}$$

$$\geq 2 - 2 \int_{\mathbb{R}} |\hat{p}|^2 \Phi_p^{-\frac{1}{2}} = 2 \, (1 - \langle p, p^\circ \rangle_{\mathscr{H}}) \, . \tag{4.85}$$

Since $|\hat{p}|^2$ is positive and Φ_p is bounded and strictly positive a.e. the distance is minimized if and only if $\alpha(\nu) = 0$ a.e. in $\mathbb{R}$, i.e. if we have equality in (4.85). Hence $\phi = 1$ a.e. and so $\tilde{\tau} = \tau$ a.e., which corresponds hence to the unique orthonormal Löwdin generator $f = p^\circ$ with an L^2-distance to p given in (4.85). □

Remark. Note, that in fact the phase function ϕ has no influence on the power spectrum

$$\left|\hat{p}^\circ_\phi\right|^2 = |\phi\tilde{\tau}\hat{p}|^2 = |\tilde{\tau}\hat{p}|^2 = |\hat{p}^\circ|^2 \, . \tag{4.86}$$

Nevertheless, We have to rescale the orthonormal generator p° to respect the FCC mask, see Section 4.5. For this the maximal difference of the power spectrum[8] of the (normalized) optimal designed pulse and the orthonormalized pulse is of interest, i.e.

$$\left\| |\hat{p}|^2 - |\hat{p}^\circ|^2 \right\|_{L^\infty} = \operatorname*{ess\,sup}_{\nu \in \mathbb{R}} \left\{ \left| \frac{\Phi_p(\nu) - 1}{\Phi_p(\nu)} \right| \cdot |\hat{p}(\nu)|^2 \right\} \leq \left\| \frac{\Phi_p - 1}{\Phi_p} \right\|_{L^\infty} \cdot \|\hat{p}\|_{L^\infty}^2 \, . \tag{4.87}$$

This shows again that this L^∞ distortion is also determined by the spectral properties of the optimal designed pulse p and its Riesz bounds. Unfortunately it is very hard to control the optimization and orthogonalization filter simultaneously as will be shown in the next section.

[8] In fact the L^∞-distance of the FCC mask S_{FCC} and $|\hat{p}^\circ|^2$ in F is relevant, assumed $|\hat{p}^\circ|^2$ is bounded by S_{FCC}.

4.4.3 Interdependence of Orthogonalization and Optimization

The causal FIR optimization process in (4.12) of a fixed initial pulse q of odd order L with clock rate $1/T_0$ can be also written in the time-symmetric form as a real semi-discrete convolution

$$p = q *'_{T_0} \mathbf{g}^L \quad \text{for} \quad \mathbf{g}^L \in \ell^2_{\tilde{L}}(\mathbb{R}) \tag{4.88}$$

with $\tilde{L} := (L-1)/2$. In this section we investigate the interdependence of the IIR filter $\mathbf{h}$ and the FIR filter $\mathbf{g}^L$, i.e. the interdependence of the orthogonalization filtering in Section 4.3.2 and the FCC optimization filtering in Section 4.2 for arbitrary clock rates. So far we have first optimized spectrally and afterwards performed the orthonormalization. In this order for a chosen q, the orthogonalization filter $\mathbf{h}$ depends on $\mathbf{g}^L$, hence we write $\mathbf{h} = \mathbf{h}_g$. Moreover the clock rates of the filters differ, hence we stick the time-shifts as index in the semi-discrete convolutions. For the T-shift-orthogonal pulse we get then

$$p^{T,\circ} = (q *'_{T_0} \mathbf{g}^L) *'_T \mathbf{h}_g. \tag{4.89}$$

Let us set $T = \Delta T_0$ and $T_q = N_q T_0$ for $N_q \in \mathbb{N}, \Delta > 0$. Since the filter clock rate of $\hat{\mathbf{g}}^L$ is fixed to $1/T_0$ to ensure full FCC–range control, the variation is expressed in Δ. To get rid of T_0 we scale the time t to $t' = t/T_0$ such that the time–shift of $\mathbf{g}^L$ is $T_0' = 1$. Then (4.89) becomes

$$p^{\Delta,\circ} = (q *' \mathbf{g}^L) *'_\Delta \mathbf{h}_g = p *'_\Delta \mathbf{h}_g. \tag{4.90}$$

We observe the following effects

1. If $\frac{1}{\Delta} \in \mathbb{N}$ then $p^{\Delta,\circ} = q^{\Delta,\circ}$.
2. If $\Delta \in \mathbb{N}$ then the distortion of $\hat{\mathbf{h}}_g$ is limited periodically to the interval $[-\frac{1}{2\Delta}, \frac{1}{2\Delta}]$.

To explain point 1), let us first orthogonalize q by $\mathbf{h}_q$ and ask for the filter $\tilde{\mathbf{g}}_{\Delta,\mathbf{g}^L}$ which preserves the Δ-orthogonalization in the presence of $\mathbf{g}^L$. Hence we aim at

$$p^{\Delta,\circ} = p *'_\Delta \mathbf{h}_g = q^{\Delta,\circ} *' \tilde{\mathbf{g}}_{\Delta,\mathbf{g}^L}. \tag{4.91}$$

But from (4.43) we know how $\mathbf{h}_g$ acts in the frequency-domain:

$$\left|\hat{p}^{\Delta,\circ}(\nu)\right|^2 = \frac{|\hat{p}(\nu)|^2}{\frac{1}{\Delta}\sum_k \left|\hat{p}(\nu - \frac{k}{\Delta})\right|^2} = \left|\hat{q}^{\Delta,\circ}(\nu) \cdot \hat{\tilde{\mathbf{g}}}_{\Delta,\mathbf{g}^L}(\nu)\right|^2. \tag{4.92}$$

With $|\hat{p}(\nu)|^2 = \left|\hat{\mathbf{g}}^L(\nu) \cdot \hat{q}(\nu)\right|^2$ we get by periodicity of $\left|\hat{\mathbf{g}}^L\right|^2$ for $\frac{1}{\Delta} \in \mathbb{N}$

$$\left|\hat{p}^{\Delta,\circ}(\nu)\right|^2 = \frac{|\hat{\mathbf{g}}^L(\nu)|^2 \cdot |\hat{q}(\nu)|^2}{\frac{1}{\Delta}\sum_k |\hat{\mathbf{g}}^L(\nu - \frac{k}{\Delta})|^2 \cdot \left|\hat{q}(\nu - \frac{k}{\Delta})\right|^2} = \frac{|\hat{\mathbf{g}}^L(\nu)|^2 \cdot |\hat{q}(\nu)|^2}{\frac{1}{\Delta}|\hat{\mathbf{g}}^L(\nu)|^2 \cdot \sum_k \left|\hat{q}(\nu - \frac{k}{\Delta})^2\right|} \tag{4.93}$$

$$= \frac{|\hat{q}(\nu)|^2}{\frac{1}{\Delta}\sum_k \left|\hat{q}(\nu - \frac{k}{\Delta})\right|^2} = \left|\hat{q}^{\Delta,\circ}(\nu)\right|^2, \tag{4.94}$$

s.t. $\tilde{g}_{\Delta,\mathbf{g}^L}(k) = \delta_{k0}$, which shows point 1). Hence, the price of orthogonalization is the loss of a frequency control, since the frequency property is now completely given by the basic pulse q and Δ. In Fig. 4.10 the effect is plotted for $\Delta \in [1,2]$ and $L = 25$. For small Δ the distortion is increase by the orthogonalization. This also shows that a perfect orthogonalization and optimization with the same clock rates is not possible.
In point 2) a perfect orthogonalization does not completely undo the optimization, since $T = \Delta$ is greater than the optimization shift $T_0 = 1$. For $\Delta = 2$ we can describe the filter by using the addition theorem in $|\hat{\mathbf{g}}^L(\nu+1/2)|^2 = \hat{\mathbf{r}}_{\mathbf{g}^L}(\nu+1/2) = 2r_{\mathbf{g}^L,0} - \hat{\mathbf{r}}_{\mathbf{g}^L}(\nu)$ by

$$\sum_k \left|\hat{p}\left(\nu - \frac{k}{2}\right)\right|^2 = \hat{\mathbf{r}}_{\mathbf{g}^L}(\nu)\left[\sum_k \left|\hat{q}\left(\nu + \frac{2k}{2}\right)\right|^2 + \quad \frac{2r_{\mathbf{g}^L,0} - \hat{\mathbf{r}}_{\mathbf{g}^L}(\nu)}{\hat{\mathbf{r}}_{\mathbf{g}^L}(\nu)} \underbrace{\sum_k \left|\hat{q}\left(\nu + \frac{2k+1}{2}\right)\right|^2}_{=:\Phi'_q(\nu)}\right]$$

$$= \hat{\mathbf{r}}_{\mathbf{g}^L}(\nu)\left[\Phi_q(\nu) - 2\frac{\hat{\mathbf{r}}_{\mathbf{g}^L}(\nu) - r_{\mathbf{g}^L,0}}{\hat{\mathbf{r}}_{\mathbf{g}^L}(\nu)}\Phi'_q(\nu)\right] \tag{4.95}$$

which results in the filter power spectrum (4.92)

$$\left|\hat{\tilde{\mathbf{g}}}_{2,\mathbf{g}^L}(\nu)\right|^2 = \left(1 - 2\frac{\hat{\mathbf{r}}_{\mathbf{g}^L}(\nu) - r_{\mathbf{g}^L,0}}{\hat{\mathbf{r}}_{\mathbf{g}^L}(\nu)} \cdot \frac{\Phi'_q(\nu)}{\Phi_q(\nu)}\right)^{-1}. \tag{4.96}$$

But since we fixed $\Delta = 2$ and q we can calculate Φ_q, Φ'_q and $\hat{q}^{\Delta,\circ}$. This provides a separation of the filter power spectrum $\hat{\mathbf{r}}_{\mathbf{g}^L}(\nu) = |\hat{\mathbf{g}}^L(\nu)|^2$ and the orthogonalization. Unfortunately, this does not yield linear constraints for $\mathbf{r}$.
Finally, note that the time-shifts and hence the filter clock rates, have to be chosen such that an overlap of the basic pulse occurs. Otherwise a frequency shaping is not possible.

Summarizing, the discussion above shows that joint optimization and orthogonalization is a complicated problem and only in specific situations a closed-form solution seems to be possible.

4.4.4 Compactly Supported Orthogonal Generators

For PPM transmission a time-limited shift-orthogonal pulse is necessary to guarantee an ISI free modulation in a finite time. Such a PPM pulse is a compactly supported orthogonal (CSO) generator. In PPM this is simply realized by avoiding the overlap of translates.

To apply our OOPPM design it is hence necessary to guarantee a compact support of the Löwdin generator p° given in Theorem 12. In this section we will therefore investigate the support properties of orthogonal generators. In our previous paper [WJT10b] we could show numerically an approximately shift-orthogonal behaviour. The existence of a CSO generator (with overlap) was already shown by Daubechies in [Dau90]. Unfortunately, she could not derive a closed-form for such an CSO generator. Moreover, to obtain a realizable construction of a CSO generator this construction has to be performed in a finite time. So our Löwdin construction should be obtained by a FIR filter.

PSI spaces of compactly supported (CS) generators, were characterized in detail by de Boor et al. in [BDR94] and called *local PSI spaces*. If the generator is also stable, as in Theorem 12, then there exists a sequence $\mathbf{c} \in \ell^2$ such that $p *' \mathbf{c}$ is an CSO generator. Moreover, any CSO generator is of this form. To investigate compactness, de Boor et.al. introduced the concept of *linear independent* shifts for CS generators. The linear independence property of a CS generator p is equivalent by [BDR94, Res. 2.24] to

$$\{(\mathrm{L}p)(z-n)\}_{n \in \mathbb{Z}} \neq 0 \quad \text{for all} \quad z \in \mathbb{C} \tag{4.97}$$

where $(\mathrm{L}p)$ denotes the Laplace transform of p. This means $(\mathrm{L}p)$ do not have periodic zero points. Note that this definition of independence is stronger than finitely independence, see definition in [BDR94]. If we additionally demand linear independence of p in our Theorem 12, then this CS generator is unique up to shifts and scalar multiplies. Furthermore, a negative result is shown in [BDR94], which excludes the existence of a CSO generator if p itself is not already orthogonal. But if p is already orthogonal, then p is unique up to shifts and scalar multiplies and then the Löwdin construction becomes a scaled identity (normalizing of p). The statement is the following:

Theorem 14 (Th. 2.29 in [BDR94])**.** *Let $p \in L^2$ be a linear independent generator for $\mathscr{S}(p)$ which is not orthogonal, then there does not exists a compactly supported orthogonal generator p° for $\mathscr{S}(p)$, i.e. there exists no filter $\mathbf{h} \in \ell^2$ such that $p^\circ = p *' \mathbf{h}$.*

If p is a linear independent generator which is not orthogonal, then the Löwdin generator p° is not compactly supported. We extend this together with the existence and uniqueness of a linear independent generator for a local PSI space $\mathscr{S}(p)$:

Corrolary 5. *Let $p \in L^2$ be compactly supported. If there exists a compactly supported orthogonal generator p° for $\mathscr{S}(p)$, then it is unique.*

Proof. Any CSO generator $p^\circ \in L^2$ is a linear independent generator by [BDR94, Prop. 2.25(c)]. Since the linear independent generator is unique by [BDR94, Th. 2.28(b)], the CSO generator is as well. □

Remark. In any case there exists an orthogonal generator for a local PSI space. For a stable CS generator p our Theorem 12 gives an explicit construction and approximation for an orthogonal generator by an IIR filtering of p. If the Löwdin generator is not CS, it is the unique orthogonal generator with the minimal L^2-distance to the original stable CS generator by Theorem 13. So far it is not clear whether there exists a IIR filter $\mathbf{c} \in \ell^2$ which generates a CSO generator from a stable CS generator or not. What we can say is that if the inverse square-root of the Gram matrix is banded, then the rows corresponds to FIR filters which produce CSO generators, since the semi-discrete convolution reduces to a finite linear combination of CS generators. So this is a sufficient condition for the Löwdin generator to be CSO, but not a necessary one.

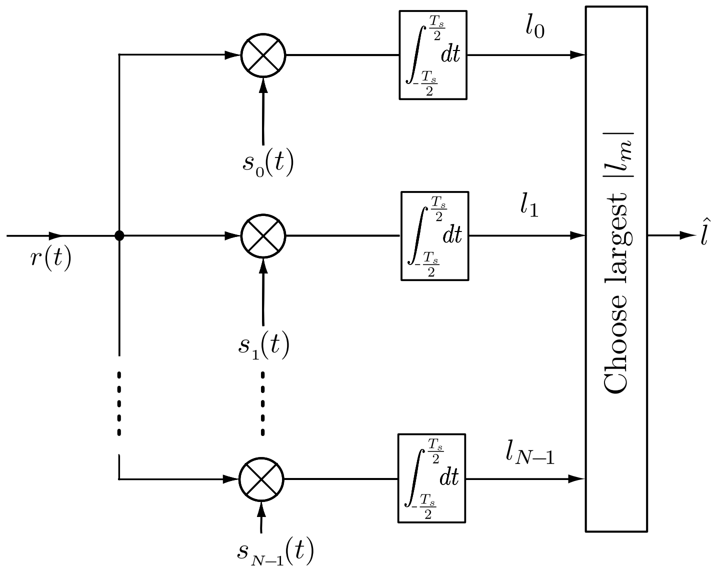


Figure 4.3: Optimal Receiver for N-ary orthogonal modulation with scaled Löwdin pulses $s_i := \alpha^* p_{i-M}^{T,\circ,M}$.

4.5 Application of the Design

Here we give some exemplary applications of our filter designs developed in Section 4.2 and Section 4.3 for UWB-IR.

FIR filter realized by a distributed transversal filter The FIR filter is completely realized in an analog fashion. It consists of time-delay lines and multiplication of the input with the filter constants. Note also that these filter values are real-valued. An application to UWB was already considered in [ZZMW09].

Transmitter and receiver designs Our channel model is an AWGN channel, i.e. the received signal $r(t)$ is the transmitted UWB signal $u(t)$ given in (4.7) plus white Gaussian noise. For simplicity of the discussion we omitted the time-hopping sequence in (4.4). We propose now three $N - ary$ waveform modulations for our pulse design. All modulations are linear and performed in the baseband. Hence the signals (pulses) are all real-valued.

(a) A pulse shape modulation (PSM) with the Löwdin pulses $\{p_m^{T,\circ,M}\}_{m=-M}^{M}$, which corresponds to a N-ary orthogonal waveform modulation. The receiver is realized with N correlators using the Löwdin pulses as templates. The absolute value is taken from the correlators output l_m due to the random amplitude flip by a_n in (4.7). For the optimal receiver see Fig. 4.3.

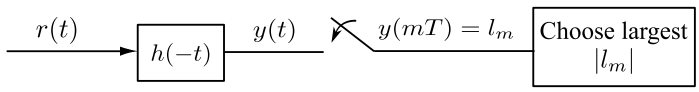


Figure 4.4: Matched filter receiver for an OPPM scheme with centered ALO or LO pulse.

(b) The centered ALO and LO pulse $\tilde{p}^{T,\circ,M}$ resp. $p^{T,\circ,M}$ for an OPPM design are in fact a non-orthogonal modulation scheme with a matched filter at the receiver, see Fig. 4.4.

(c) The limiting OOPPM design with the Löwdin pulse p° is not practically feasible, since we have to implement an IIR filter. Hence we only refer to this setup for the theoretical limit.

Scaling with respect to the FCC mask The operations B^M and $\tilde{B}^M$ generate pulses which are normalized in energy but do not respect anymore the FCC mask. So we have to find for the mth pulse its maximal scaling factor $\alpha_m > 0$ s.t.

$$\left|\alpha_m \cdot \hat{p}_m^{\circ,M}(\nu)\right|^2 \leq S_{\text{FCC}}(\nu) \tag{4.98}$$

is still valid for any $\nu \in F$. This problem is solved by

$$\alpha_m^* = \left\| \frac{\left|\hat{p}_m^{\circ,M}\right|^2}{S_{\text{FCC}}} \right\|_{L^\infty([0,F])}^{-\frac{1}{2}} = \left\| \frac{\sqrt{S_{\text{FCC}}}}{\hat{p}_m^{\circ,M}} \right\|_{L^\infty([0,F])}. \tag{4.99}$$

For the scaled Löwdin pulses we can easily obtain the following upper bound for the NESP value (4.13)

$$\eta(\alpha_m^* p_m^{\circ,M}) = \frac{\int_F \left|\alpha_m^* \hat{p}_m^{\circ,M}\right|^2}{\int_F S_{\text{FCC}}} \leq \frac{\left\|\hat{p}_m^{\circ,M}\right\|_{L^2}^2}{\mathcal{E}_{F_p}} \cdot \left\| \frac{\sqrt{S_{\text{FCC}}}}{\hat{p}_m^{\circ,M}} \right\|_{L^\infty([0,F]}^2 \tag{4.100}$$

$$= \frac{1}{\mathcal{E}_{F_p}} \cdot \left\| \frac{\sqrt{S_{\text{FCC}}}}{\hat{p}_m^{\circ,M}} \right\|_{L^\infty([0,F]}^2 = \frac{1}{\mathcal{E}_{F_p}} \cdot \left\| \frac{S_{\text{FCC}}}{\left|\hat{p}_m^{\circ,M}\right|^2} \right\|_{L^\infty([0,F]} \tag{4.101}$$

where we denoted with $\mathcal{E}_{F_p} = \int_{F_p} S_{\text{FCC}}$ the allowed energy of the FCC mask in the passband F_p. Thus, the maximization of the symbol energy under the FCC mask is the maximization of α_m^* in (4.99), i.e. a maximization of the L^∞-norm in the frequency domain.

4.5.1 Performance of the Proposed Designs

For a given transmission design, consisting of a modulation scheme and a receiver, the average bit error probability P_e over E_b/N_0 is usually considered as the performance criterion. We consider real-valued signals in the baseband with finite symbol duration T_s. The optimal receiver for a non-orthogonal N-ary waveform transmission is the correlation receiver with M correlators, see Fig. 4.3 with maximum likelihood decision.

N-ary orthogonal PSM for scheme (a) above The average (symbol) error probability for N-ary orthogonal pulses with equal energy $\mathcal{E}$ can be upper bounded by [Tre68]

$$P_e \le (N-1)\,\mathrm{erfc}\left(\sqrt{\frac{\mathcal{E}}{N_0}}\right) \tag{4.102}$$

Note, that this error probability is the same as for an orthogonal PPM modulation (4.10). To obtain equal energy symbols and FCC compliance we have to scale each Löwdin pulse with $\sqrt{\mathcal{E}} = \alpha^* = \min_m\{\alpha_m^*\}$.

N-ary overlapping PPM for scheme (b) above Here we can substitute the N correlations by one matched filter $h = p^{T,\circ,M}$ resp. $\tilde{h} = \tilde{p}^{T,\circ,M}$ and obtain the statistics $|l_m|$ by sampling the output. The average error probability per symbol P_e given exactly in [Tre68, Prob. 4.2.11] for equal energy signals and can be computed numerically. The energy is given by $\sqrt{\mathcal{E}} = \alpha^* = \alpha_0^*$ and $\sqrt{\tilde{\mathcal{E}}} = \tilde{\alpha}_0^*$ calculated in (4.99) for $p^{T,\circ,M}$ resp. $\tilde{p}^{T,\circ,M}$. Upper bounds obtained in [Jac67] can be used for the ALO resp. LO average error probability

$$P_e \le \frac{1}{2}\sum_{j=2}^{N}\mathrm{erfc}\left(\sqrt{\frac{\mathcal{E}}{2N_0}(1-\rho_{1j})}\right) \quad \text{and} \quad \tilde{P}_e \le \frac{1}{2}\sum_{j=2}^{N}\mathrm{erfc}\left(\sqrt{\frac{\tilde{\mathcal{E}}}{2N_0}(1-\tilde{\rho}_{1j})}\right) \tag{4.103}$$

with $\rho_{1j} = \mathcal{E}r_{p^{T,\circ,M}}(jT)$ and $\tilde{\rho}_{1j} = \tilde{\mathcal{E}}r_{\tilde{p}^{T,\circ,M}}(jT)$, since the symbols are given by $s_j = \sqrt{\mathcal{E}}p^{T,\circ,M}(\cdot - jT)$ resp. $\tilde{s}_j = \sqrt{\tilde{\mathcal{E}}}\tilde{p}^{T,\circ,M}(\cdot - jT)$ for $j = 1,\ldots,N$. The error probabilities depend on the pulse energy and on the decay of the sampled auto-correlation defined in (4.37).

4.5.2 Simulation Results

The most common basic pulse for an UWB-IR transmission is the Gaussian monocycle: $q(t) \simeq t \cdot \exp\left(-t^2/\sigma^2\right)$ where σ is chosen such that the maximum of $|\hat{q}(f)|^2$ is reached at the center frequency $f_c = 6.85$GHz of the passband [WJT10a]. Since we need compact support and continuity for our construction, we mask q with a unit triangle window Λ instead of a simple truncation. Also any other continuous window function which goes continuously to zero (e.g. the Hann window) can be used, as long as the lower Riesz bound $A > 0$ can be ensured, see Theorem 12. We have used an algorithm in [WJT10a] to compute A and B numerically. Note that for any continuous compactly supported function we have a finite upper Riesz bound B, see [JM91, Th.2.1]. The width (window length) is chosen to $T_q = T_\Lambda = N_qT_0 \approx 0.21428$ns, such that at least 99.99% of the energy of q is contained in the window $[-T_q/2, T_q/2]$, see Fig. 4.5. We express all time instants as integer multiples of T_0. Also, in Fig. 4.5 we plot the optimal pulse obtained by a FIR filter of order $L = 25$ which results in a time-duration $T_p = 30T_0 = 5T_q$ of p. In our simulation we choose $T_q := 6T_0 = N_qT_0$ and $L = 25$ as the filter order of the FCC-optimization. Hence, the optimized pulses have a total time duration of $T_p = (L-1)T_0 + T_q = 30T_0 = N_pT_0$.

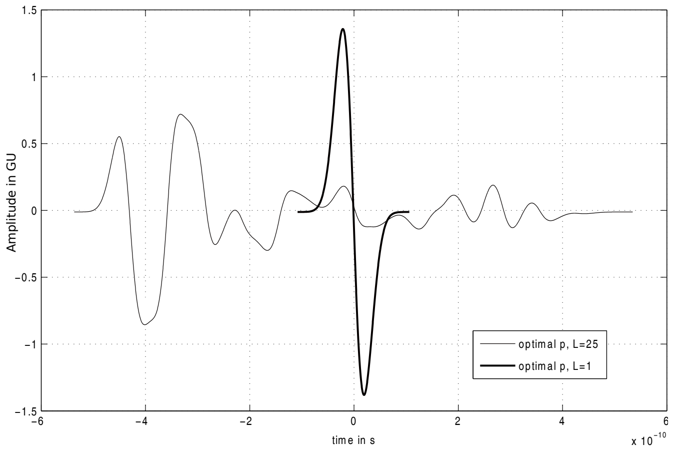


Figure 4.5: Optimal pulse p for $L=25$ and basic pulse $q=p$ for $L=1$ in time-domain in generic units (GU).

The Riesz condition (4.31) has been already verified in [WJT10a] for this particular setup. Theorem 12 uses the normalization $T' = 1$. Translating between different support lengths $T_p' = K$ is done by setting $t := t'T_p/K$. Now the support of $p(t')$ is $[-K/2, K/2]$ with fixed $T' = 1$. To obtain good shift-orthogonality, we have to choose $M > K$. This we control with an integer multiple $m = 2$, i.e. $M = mK = 2K$. The support length T_s of all the LO (ALO) pulses is then given as

$$\begin{aligned} T_{\tilde{p}^{T,\circ,M}} &= (N-1)T + T_p = (2mK)T_p/K + N_p T_0 \\ &= T_{p^{T,\circ,M}} = (2m+1)N_p T_0 = 150 T_0. \end{aligned} \tag{4.104}$$

Now the time slot $[-T_{p^{T,\circ,M}}/2, T_{p^{T,\circ,M}}/2]$ exactly contains N mutually orthogonal pulses $\{p_m^{T,\circ,M}\}$, i.e. N orthogonal symbols with symbol duration $T_s = T_{p^{T,\circ,M}}$ having all the same energy and respecting strictly the FCC mask. This is a N-ary orthogonal signal design, which requires high complexity at receiver and transmitter, since we need a filter bank of N different filters.
Our proposal goes one step further. If we only consider one filter, which generates at the output the centered Löwdin orthogonal pulse $p^{T,\circ,M}$, we can use this as a approximated Nyquist pulse with a PPM shift of $T = T_p/K$ to enable N-ary OPPM transmission by obtaining almost orthogonality.

Advantages of the proposed design are: a low-complexity at transmitter and receiver, a combining of $\mathbf{g}^L$ and $\mathbf{h}$ into a single filter operating with clock rate $1/T_0$ and q as input if $T = T_p/(T_0 K) \in \mathbb{N}$, a signal processing "On the fly" and finally a much higher bit rate compared to a binary-PPM. The only precondition for all this, is a perfect synchronisation between transmitter and receiver. In fact, we have to sample equidistantly at rate of $1/T$. The output of the matched filter $h(-t) = p^{T,\circ,M}(t)$ is given by

$$y(t) = \int_{-\infty}^{\infty} r(\tau) p^{T,\circ,M}(\tau - t) d\tau \tag{4.105}$$

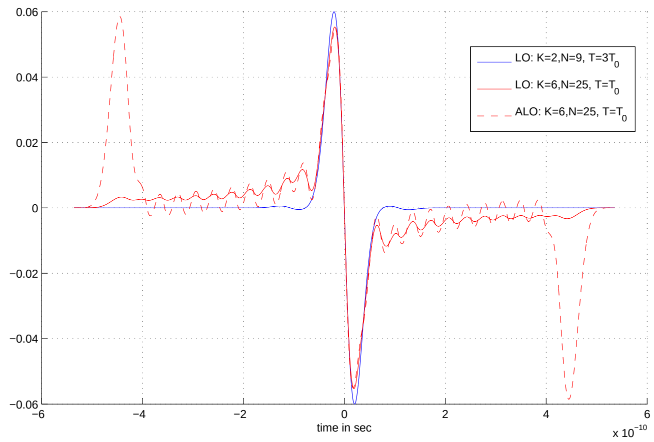


Figure 4.6: ALO and LO pulses for Gaussian monocycle in time, $L = 1$ and $M = 2K$.

and recovers the mth symbol. The statistic $l_m = y(mT)$ is the correlation of the received signal with the symbol s_m, see Fig. 4.4. Note that the shifts have support in the window $[-1.5T_{p^{T,\circ,M}}, 1.5T_{p^{T,\circ,M}}]$, but are almost orthogonal outside the symbol window $[-T_{p^{T,\circ,M}}/2, T_{p^{T,\circ,M}}/2]$ due to the compactness and approximate shift-orthogonal character of the Löwdin pulse $p^{T,\circ,M}$.

In Fig. 4.6 the centered orthogonal pulses for $T = 5T_0 = 5T_q/6$ match almost everywhere the original masked Gaussian monocycle, since the translates are almost non-overlapping, hence they are already almost orthogonal. For $T = T_0 = T_q/6$ the overlap results in a distortion of the centered orthogonal pulses, where the ALO pulses have high energy concentration at the boundary (circulant extension of the Gram Matrix).

In Fig. 4.7 the pulse shapes in time for the centered Löwdin orthogonal pulses are plotted. The ALO pulses are matching the LO pulses for $T > 2T_0$ almost perfectly such that we did not plot them, since the resolution of the plot is to small to see any mismatch. Only for the critical shift $T = T_0$ a visible distortion is obtained at the boundary. In the next Fig. 4.8 we plotted therefore the ALO and LO pulse for $T = 1.5T_0$ to show that the ALO pulses indeed converge very fast to the LO pulse if $T >> T_0$. The reason is that for small time shifts of p the Riesz bounds and so the clustering behaviour of $\mathbf{G}_M$ and $\mathbf{G}_M^{-\frac{1}{2}}$ decreases. Hence the approximation quality of $\mathbf{G}_M^{-\frac{1}{2}}$ with $\tilde{\mathbf{G}}_M^{-\frac{1}{2}}$ decreases, which results in a shape difference.

To study the shift-orthogonal character of the ALO and LO pulses for various T, we have plotted the auto-correlations in Fig. 4.9. As can be seen, the samples $r_{p^{T,\circ,M}}(mT) = \rho_{1m} \approx \delta_{m0}$, i.e. they vanish at almost each sample point except in the origin. The NESP performance for various values of T is shown in Fig. 4.10. Approaching $T = T_0$ cancels the FIR prefilter optimization of $\mathbf{g}^L$, i.e. the spectrum becomes flat. Finally, in Fig. 4.11 the gain of our orthogonalization

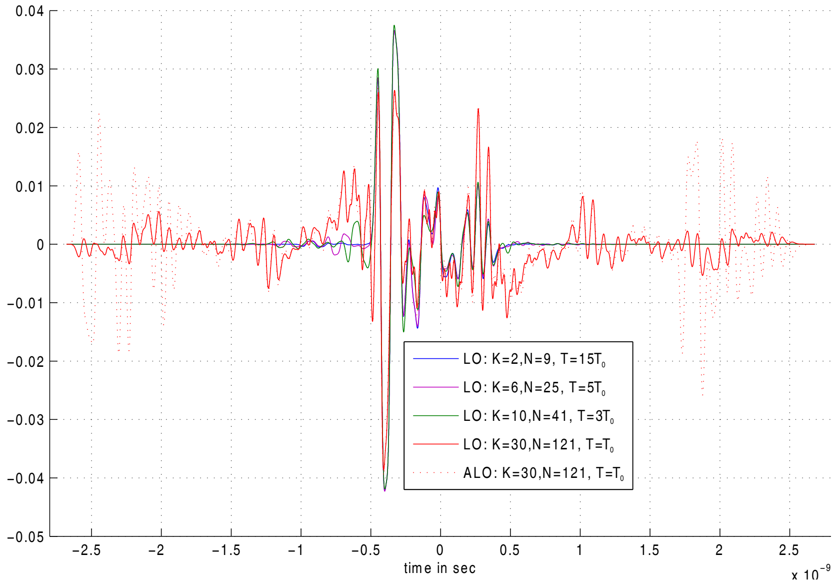


Figure 4.7: Orthogonal pulses $p^{T,\circ,M}$ with $L = 25$, $M = 2K$ and various $T = T_p/K$.

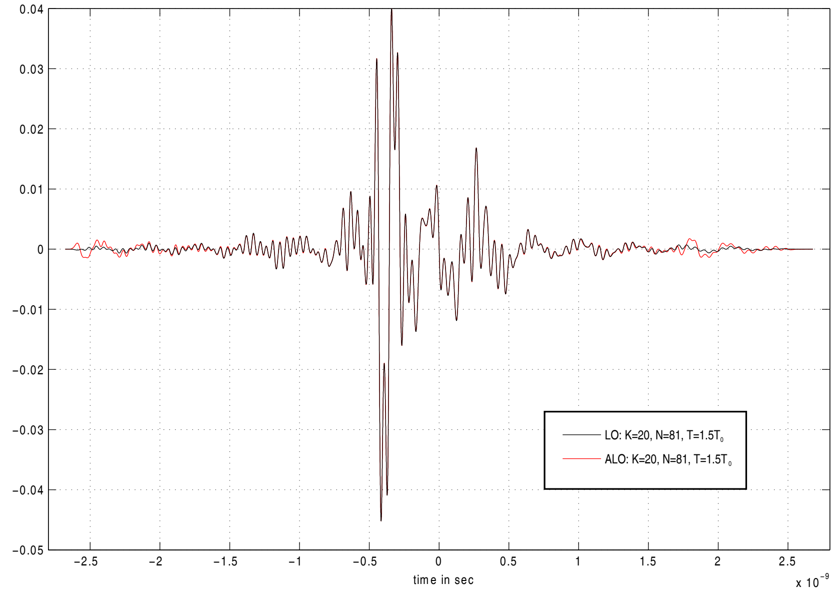


Figure 4.8: Centered ALO and LO pulse for $T = 1.5T_0$ with $L = 25$, $M = 2K = 40$.

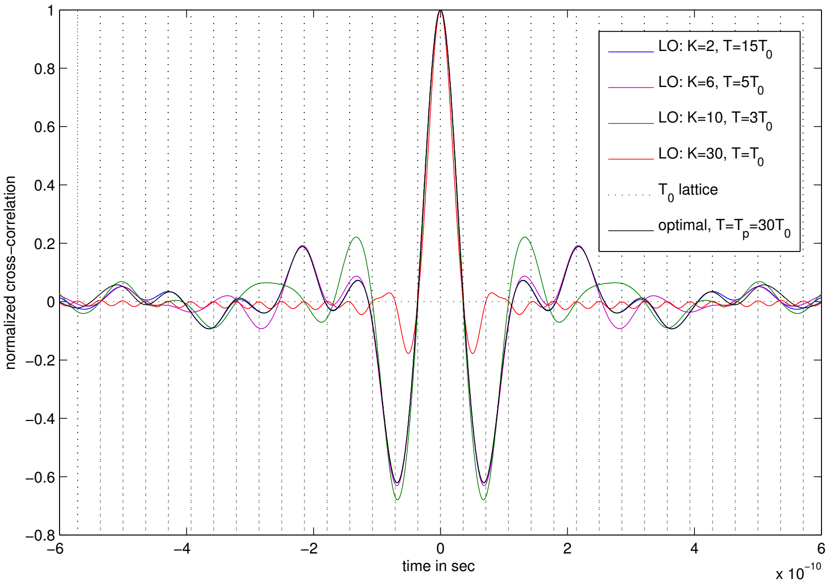


Figure 4.9: Auto-correlation of pulses for $L = 25, T = T_p/K$.

strategy can be seen. In both cases, the N-ary OOPPM design transmit at a bit rate of

$$R_b(K) = \frac{\log(N)}{T_s} = \frac{\log(4K+1)}{150T_0}. \tag{4.106}$$

Fig. 4.11 shows the NESP value over the bit rate for fixed $L = 25$ and $m = 2$. Hence, decreasing $T = T_p/K$ results in more overlap and increases the amount N of symbols in T_s but slightly decreases η, see Fig. 4.10.

Summarizing, a triplication of the bit rate from 0.18Gbit/s to 0.6Gbit/s is possible without loosing much spectral efficiency. Let us note the fact that the bit data rate is an uncoded rate. Obviously, the unshaped Gaussian monocycle then yields the highest data rate, since (4.106) behaves logarithmically in the number of symbols, as seen in Fig. 4.11. But this has practically zero SNR when respecting the FCC regulation. On the other hand, a longer symbol duration, allows in (4.3) a higher energy $\mathcal{E}$ and hence a lower error rate in (4.10).

Hence, the increased symbol duration used for FCC optimal filtering of the masked Gaussian monocycle q can be more than compensated by the proposed OOPPM technique.

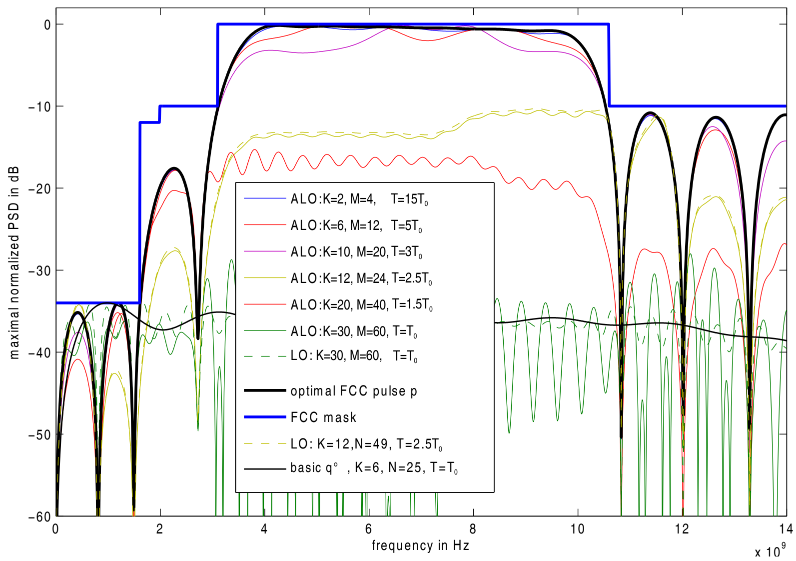


Figure 4.10: PSD of pulses for $L = 25, T = T_p/K$.

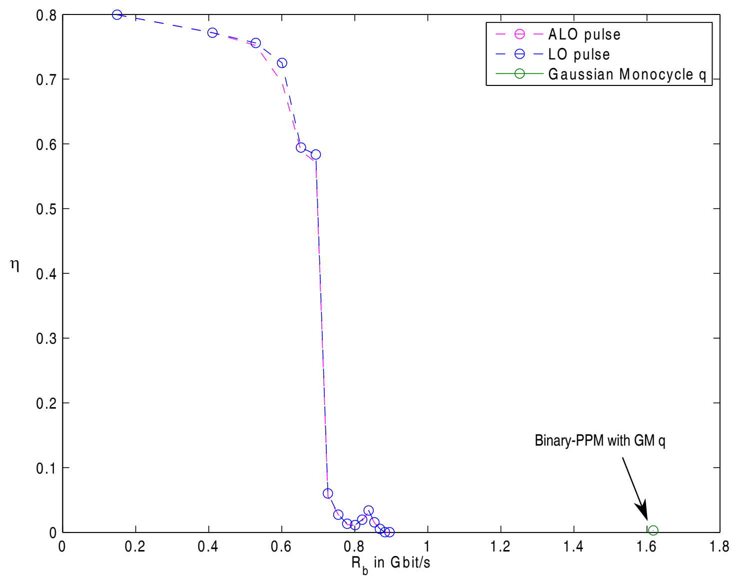


Figure 4.11: Orthogonalization results for $L = 25$ in $T_s = 150T_0$ over various T resulting in various bit rates R_b.

5 Conclusion

Support properties of signals becoming more and more relevant in modern signal processing. It could be show in this dissertation how support restrictions of time-discrete signals can be used to significantly reduce the sampling rate of convolution outputs in a stable way. These stable embeddings offers a way for many applications: for example in phase retrieval, linear time invariant systems and to even more general systems, which can be formulated as bilinear mappings. A general condition for such bilinear systems could be formalized, the restricted norm multiplicativity (RNMP), which guarantees the RIP with exponential probability on the image of these bilinear maps from measurements scaling only additive in the dimension of the input signals. The RNMP could be established for the discrete convolution with sparse signals and allows a low-dimensional stable embedding, which therefore enables lower sample rates. Moreover, the result could possibly enabling a compressed phase retrieval, which will be more investigated by the author in the near future.

For time-continuous signals a compact support can also be used to shift-orthogonalize in a stable and efficient way time-limited pulses, which can be applied in UWB systems. Here, we have proposed a new pulse design method for UWB-IR which provides high spectral efficiency under FCC constraints and allowing an N-ary OPPM transmission with finite transmission and receiving time by keeping almost orthogonality. In fact, the correlation parameters can be keep below the noise level by using small time-shifts $T < T_p$. As a result, this provides much higher data rates as compared to BPSK or BPPM. Furthermore, our pulse design provides a N-ary orthogonal PPM transmission by getting a lower bit error rate at the price of a higher complexity. Simultaneous orthogonalization and spectral frequency shaping is a challenging and hard problem. For certain shifts, being integer multiples of T_0, a numerical solver might be helpfully to directly solve the combined problem as discusses in Section 4.4.3. It should be highlighted the broad application of the OOPPM design, not only being limited to UWB systems but rather applicable to any communication system demanding a local frequency constraint.

A Appendix A

A.1 Discrete Uncertainty Principle

Here, we use results on the discrete uncertainty principle of DONOHO-STARK [DS89] and a refined version of TAO [Tao05] and CHEBOTARËV [SL96] to motivate our result for groups of prime order without an extra zero padding. Donoho and Stark could show in [DS89] the following *discrete uncertainty principle* for any $\mathbf{x} \in \mathbb{R}^n$

$$\|\mathbf{x}\|_0 \|\mathbf{F}\mathbf{x}\|_0 \geq n \tag{A.1}$$

$$\|\mathbf{x}\|_0 + \|\mathbf{F}\mathbf{x}\|_0 \geq 2\sqrt{n} \tag{A.2}$$

Since $\operatorname{diag}(\mathbf{F}\mathbf{x})\mathbf{F}\mathbf{y} = \mathbf{F}\mathbf{x} \odot \mathbf{F}\mathbf{y}$, where $\odot$ denotes the pointwise product, we can see by definition (1.68), that

$$\|\mathbf{x} \circledast \mathbf{y}\| = \sqrt{n} \|\mathbf{F}\mathbf{x} \odot \mathbf{F}\mathbf{y}\| . \tag{A.3}$$

Hence, we get the following implication

$$\begin{aligned} (\mathbf{x} \neq \mathbf{0} \neq \mathbf{y}, \|\mathbf{F}\mathbf{x}\|_0 + \|\mathbf{F}\mathbf{y}\|_0 > n) &\Rightarrow (\mathbf{F}\mathbf{x} \odot \mathbf{F}\mathbf{y} \neq 0) \\ &\Leftrightarrow (\mathbf{0} \neq \mathbf{x} \circledast \mathbf{y}) \\ &\Leftrightarrow (0 \neq \|\mathbf{x} \circledast \mathbf{y}\|) \end{aligned} \tag{A.4}$$

By the discrete uncertainty principle (A.1) and with $\|\mathbf{x}\|_0 \leq s$, $\|\mathbf{y}\|_0 \leq f$ in Theorem 2 we get

$$\|\mathbf{F}\mathbf{x}\|_0 + \|\mathbf{F}\mathbf{y}\|_0 \geq \frac{n}{s} + \frac{n}{f} = n\frac{s+f}{sf} \overset{!}{>} n. \tag{A.5}$$

This is only possible if $s + f > sf$, which holds if and only if $s = 1$ or $f = 1$. But this is trivial. So Donoho-Stark can not provide the existence of $a > 0$ in Theorem 2 for all cases. In fact, if n is prime, the TAO inequality [Tao05]

$$\|\mathbf{x}\|_0 + \|\mathbf{F}\mathbf{x}\|_0 \geq n + 1 \Leftrightarrow \|\mathbf{F}\mathbf{x}\|_0 \geq n + 1 - s \tag{A.6}$$

yields with the assumption $n \geq s + f - 1$

$$\|\mathbf{F}\mathbf{x}\|_0 + \|\mathbf{F}\mathbf{y}\|_0 \geq 2n + 2 - s - f \geq n + 1 > n. \tag{A.7}$$

Hence, whenever $1 \leq s + f - 1 \leq n$ and n is prime, we have for all $\mathbf{x} \in \Sigma_s^n, \mathbf{y} \in \Sigma_f^n$ that $\mathbf{x} \circledast \mathbf{y} \neq \mathbf{0} \Leftrightarrow \mathbf{x} \neq \mathbf{0} \neq \mathbf{y}$. Due to the upper bound in (2.31), which was shown in Lemma 2, the map $\|\mathbf{x} \circledast \mathbf{y}\|$ is continuous and hence the infimum a is attained and is strictly larger than zero.

Bibliography

[ALN08] S. Arora, J. R. Lee, and A. Naor. “Euclidean distortion and the sparsest cut”. In: *J. Amer. Math. Soc.* **21**: (2008), pp. 1–21.

[AMR02] R. Abraham, J. E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis, and Applications*. Springer, 2002.

[And05] J. B. Anderson. *Digital transmission engineering*. Wiley-IEEE Press, 2005.

[AS02] A. Aldroubi and Q. Sun. “Locally Finite Dimensional Shift-Invariant Spaces in $\mathbb{R}^d$”. In: *Proceedings of the American Mathematical Society* **130**:9 (2002), pp. 2641–2654.

[AST01] A. Aldroubi, Q. Sun, and W.-S. Tang. “p-Frames and Shift Invariant Subspaces of L^p”. In: *Journal of Fourier Analysis and Applications* **7**: (2001), pp. 1–22.

[AU94] A. Aldroubi and M. Unser. “Sampling procedures in function spaces and asymptotic equivalence with shannon’s sampling theory”. In: *Journal Numerical Functional Analysis and Optimization* **15**: (1994), pp. 1 –21.

[BBCE07] R. Balan, B. G. Bodmann, P. Casazza, and D. Edidin. “Fast Algorithms for Signal Reconstruction without Phase”. In: *Proc. SPIE 6701*. 2007.

[BCE06] R. Balan, P. Casazza, and D. Edidin. “On signal reconstruction without phase”. In: *Applied and Computational Harmonic Analysis* **20**: (2006), pp. 345–356.

[BCMN13] A. S. Bandeira, J. Cahill, D. G. Mixon, and A. A. Nelson. “Saving phase: Injectivity and stability for phase retrieval”. In: *EURASIP Journal on Applied Signal Processing* (2013).

[BCW10] R. Baraniuk, V. Cevher, and M. Wakin. “Low-Dimensional Models for Dimensionality Reduction and Signal Recovery: A Geometric Perspective”. In: *Proceedings of the IEEE* **98**: (2010), pp. 959–971.

[BD73] F. F. Bonsall and J. Duncan. *Complete normed algebras*. Springer, 1973.

[BDDW08] R. G. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. “A Simple Proof of the Restricted Isometry Property for Random Matrices”. In: *Constructive Approximation* **28**:3 (2008), pp. 253–263.

[BDK84] I. Bar-David and G. Kaplan. “Information rates of photon-limited overlapping pulse position modulation channels”. In: *IEEE Trans. Inf. Theory* **30**: (1984), pp. 455–463.

[BDR94] C. d. de Boor, R. A. DeVore, and A. Ron. “The Structure of Finitely Generated Shift-Invariant Spaces in $L^2(\mathbb{R}^d)$”. In: *Journal of Functional Analysis* **119**: (1994), pp. 37–78.

[BEJJ06] C. R. Berger, M. Eisenacher, H. Jäkel, and F. Jondral. “Pulseshaping in UWB Systems Using Semidefinite Programming with Non-Constant Upper Bounds”. In: *IEEE Int. Symp. on Personal, Indoor and Mobile Radio Communications*. 2006.

[BEZJ07] C. Berger, M. Eisenacher, S. Zhou, and F. Jondral. “Improving the UWB Pulseshaper Design Using Nonconstant Upper Bounds in Semidefinite Programming”. In: *IEEE Journal of Selected Topics in Signal Processing* **1**:3 (2007), pp. 396–404.

[BG05] A. Böttcher and S. M. Grudsky. *Spectral Properties of Banded Toeplitz Matrices*. SIAM, 2005.

[Bil99] Y. Bilu. “Structure of Sets with Small Sumset”. In: *Asterisque* **258**: (1999), pp. 77–108.

[BL98] J. J. Benedetto and S. Li. “The Theory of Multiresolution Analysis Frames and Applications to Filter Banks”. In: *Applied and Computational Harmonic Analysis* **5**: (1998), pp. 389–427.

[BM13] A. S. Bandeira and D. G. Mixon. “Near-Optimal Phase Retrieval of Sparse Vectors”. In: *Wavelets and Sparsity XV, Proceedings of SPIE Optics+Photonics*. 2013.

[Boo87] C. d. de Boor. “The polynomials in the linear span of integer translates of a compactly supported function”. In: *Constructive Approximation* **3**:1 (1987), pp. 199–208.

[BW09] R. Baraniuk and M. Wakin. “Random Projections of Smooth Manifolds”. In: *Foundations of Computational Mathematics* **9**: (2009), pp. 51–77.

[CCK01] P. Casazza, O. Christensen, and N. J. Kalton. “Frames Of Translates”. In: *Collect. Math.* **52**: (2001), pp. 35–54.

[Cha02] M.-C. Chang. “A polynomial bound in Freiman’s theorem”. In: *Duke Math. J.* **113**: (2002), pp. 399–419.

[Chr03] O. Christensen. *An introduction to frames and Riesz bases*. Birkhauser, 2003.

[Chu08] E. Chu. *Discrete and continuous Fourier transforms: analysis, applications and fast algorithms*. CRC Press, 2008.

[CJ07] R. H.-F. Chan and X.-Q. Jin. *An Introduction to Iterative Toeplitz Solvers*. Ed. by N. J. Higham. SIAM, 2007.

[CSV12] E. J. Candes, T. Strohmer, and V. Voroninski. “PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming”. In: *Communications on Pure and Applied Mathematics* **66**: (2012), pp. 1241–1274.

[Dau90] I. Daubechies. “The wavelet transform, time-frequency localisation and signal analysis”. In: *IEEE Trans. Inf. Theory* **36**:5 (1990), pp. 961–1005.

[Dav79] P. J. Davis. *Circulant matrices*. Wiley, 1979.

[DDEK12] M. Davenport, M. F. Duarte, Y. Eldar, and G. Kutyniok. "Introduction of Compressed Sensing". In: Cambridge University Press, 2012. Chap. 1. Introduction of Compressed Sensing, pp. 1–64.

[DE11] M. F. Duarte and Y. C. Eldar. "Structured Compressed Sensing: From Theory to Applications". In: *IEEE Trans. Signal Process.* **59**: (2011), pp. 4053 –4085.

[Dim04] D. K. Dimitrov. *Approximation Theory: A volume dedicated to Blagovest Sendov*. Sofia, 2004.

[DK07] I. Dotlic and R. Kohno. "Design of the Family of Orthogonal and Spectrally Efficient UWB Waveforms". In: *IEEE Journal of Selected Topics in Signal Processing* **1**:1 (2007), pp. 21–30.

[DLS02] T. Davidson, Z.-Q. Luo, and J. F. Sturm. "Linear matrix inequality formulation of spectral mask constraints with applications to FIR filter design". In: *IEEE Trans. Signal Process.* **50**:11 (2002), pp. 2702–2715.

[DS89] D. Donoho and P. B. Stark. "Uncertainty Principles and Signal Recovery". In: *SIAM J. Appl. Math.* **49**: (1989), pp. 906–931.

[EFS13] M. Ehler, M. Fornasier, and J. Siegl. "Quasi-linear compressed sensing". In: *SIAM: Multiscale Modeling and Simulation* **submitted**: (2013).

[EK12] Y. Eldar and G. Kutyniok. *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[EM12] Y. Eldar and S. Mendelson. "Phase Retrieval: Stability and Recovery Guarantees". In: *arxiv* (2012), pp. 1–39.

[FCC02] FCC. *Revision of part 15 of the commission's rules regarding ultra-wideband transmission systems*. ET Docket 98-153 FCC 02-48. First Report and Order, 2002.

[Fie78] J. R. Fienup. "Reconstruction of an object from the modulus of its Fourier transform using a support constraint". In: *JOSA A* **3**: (1978), pp. 27–29.

[FR11] M. Fornasier and H. Rauhut. "Handbook of Mathematical Methods in Imaging". In: ed. by O. Scherzer. Springer, 2011. Chap. Compressive Sensing, pp. 187–228.

[Gar07] D. Garling. *Inequalities: A Journey into Linear Analysis*. Cambridge University Press, 2007.

[GGC08] J. Gutierrez-Gutierrez and P. M. Crespo. "Asymptotically Equivalent Sequences of Matrices and Hermitian Block Toeplitz Matrices With Continuous Symbols: Applications to MIMO Systems". In: *IEEE Transactions on Information Theory* **54**:12 (2008), pp. 5671–5680.

[GK08] A. C. Gordillo and R. Kohno. "Design of Spectrally Efficient Hermite Pulses for PSM UWB Communications". In: *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* **8**: (2008), pp. 2016–2024.

[Gra06] R. M. Gray. *Toeplitz And Circulant Matrices*. Vol. 2. Now Publishers Inc, 2006, pp. 155–239.

[Gre04] B. Green. “Applied and Numerical Harmonic Analysis”. In: Springer, 2004. Chap. Spectral Structure of Sets of Integers, pp. 83–96.

[Gre67] W. Greub. *Multilinear algebra*. Springer, 1967.

[Grö01] K. Gröchenig. *Foundations of time-frequency analysis*. Springer Verlag, 2001, p. 359.

[GS72] R. W. Gerchberg and W. O. Saxton. “A practical algorithm for the determination of phase from image and diffraction plane pictures”. In: *Optik* **35**: (1972), 237–246.

[GT06] B. Green and T. Tao. “Compressions, convex geometry and the Freiman-Bilu Theorem”. In: *Q J Math* **57**: (2006), pp. 495–504.

[HB11] C. Hegde and R. G. Baraniuk. “Sampling and Recovery of Pulse Streams”. In: *IEEE Trans. Signal Process.* **59**:4 (2011), pp. 1505–1517.

[HJ90] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

[HK99] Hunt and Kaloshin. “Regularity of embeddings of infinite-dimensional fractal sets into finite-dimensional spaces”. In: *Nonlinearity* **12**: (1999), pp. 1263–1275.

[IM04] P. Indyk and J. Matoušek. “Handbook of Discrete and Computational Geometry”. In: ed. by (2nd edition). CRC Press, 2004. Chap. Low Distortion Embeddings of Finite Metric Spaces, pp. 177–197.

[Jac67] I. Jacobs. “Comparisson of M-ary modulation systems”. In: *Bell Syst. Tech. J* **46**:5 (1967), 843–864.

[Jan88] A. J. E. M. Janssen. “The Zak transform: A signal transform for sam- pled time-continuous signals”. In: *Philips J. Res.* **43**: (1988), pp. 23–69.

[Jia97] R.-Q. Jia. “Shift-invariant spaces on the real line”. In: *Proc. Amer. Math. Soc.* **125**: (1997), pp. 785–793.

[JM91] R.-Q. Jia and C. A. Micchelli. “Curves and surfaces”. In: San Diego, 1991. Chap. Using the refinement equations for the construction of Pre-Wavelets II: powers and two.

[JS02] A. Janssen and T. Strohmer. “Characterization and Computation of Canonical Tight Windows for Gabor Frames”. In: *J. Fourier. Anal. Appl.* **8(1)**: (2002), pp. 1–28.

[Jun07] P. Jung. “Weyl–Heisenberg Representations in Communication Theory”. PhD thesis. Technical University Berlin, 2007.

[Kan08] E. Kaniuth. *A Course in Commutative Banach Algebras*. Sp, 2008.

[KL00] S. Konyagin and V. Lev. “Combinatorics and linear algebra of Freiman’s isomorphism”. In: *Mathematika* **47**: (2000), pp. 39–51.

[Lan36] R. Landshoff. “Quantenmechanische Berechnung des Verlaufes der Gitterenergie des Na-Cl-Gitters in Abhaengigkeit vom Gitterabstand”. In: *Zeitschrift für Physik A* **102**:3-4 (1936), pp. 201–228.

[Lee13] M. J. Lee. *Introduction to Smooth Manifolds*. Springer, 2013.

[LL01] E. H. Lieb and M. Loss. *Analysis*. Ed. by L. Small, J. E. Humphreys, J. L. Shaneson, and D. Sattinger. Graduate studies in mathematics. Vol. 14. AMS, 2001.

[LNQY09] C. Ling, J. Nie, L. Qi, and Y. Ye. "Biquadratic optimization over unit spheres and semidefinite programming relaxations". In: *SIAM J. Optim.* **20**: (2009), pp. 1286–1310.

[Lue69] D. Luenberger. *Optimization by Vector Space Methode*. Mathematics in Science and Engineering. New York: John Wiley and Sons, Inc., 1969.

[LV11] Y. Lu and M. Vetterli. "Sparse spectral factorization: Unicity and reconstruction algorithms". In: *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*. Prague, 2011.

[LW08] Y. Liu and Q. Wan. "Designing Optimal UWB Pulse Waveform Directly by FIR Filter". In: *International Conference on Wireless Communications, Networking and Mobile Computing*. 2008.

[LYG03] X. Luo, L. Yang, and G. B. Giannakis. "Designing Optimal Pulse-Shapers for Ultra-Wideband Radios". In: *J. Commun. Netw.* **5**:4 (2003), pp. 344–353.

[Löw50] P.-O. Löwdin. "On the Nonorthogonality Problem Connected with the use of Atomic Wave Functions in the Theory of Molecules and Crystals". In: *J. Chem. Phys.* **18**: (1950), pp. 367–370.

[Löw70] P.-O. Löwdin. "On the Nonorthogonality Problem". In: *Advances in Quantum Chemistry* **5**: (1970), pp. 185–199.

[Mat02] J. Matoušek. *Lectures on Discrete Geometry*. Springer, 2002.

[Mey86] Y. Meyer. "Ondelettes et fonctions splines." In: *Séminaire Équations aux dérivées partielles* **6**: (1986), pp. 1–18.

[Mid96] D. Middleton. *An introduction to statistical communication theory*. Ed. by J. B. Anderson. third. International Series in pure and applied physics. New York: IEEE Press, 1996, p. 1140.

[Nat96] M. B. Nathanson. *Additive Number Theory: Inverse Problems and the Geometry of Sumsets*. Springer, 1996.

[NM03] Y.-P. Nakache and A. F. Molisch. "Spectral Shape of UWB Signal Influence of Modulation Format, Multiple Access Scheme and Pulse Shape". In: *Proceedings of the IEEE Vehicular Technology Conference*. 2003.

[NM06] Y.-P. Nakache and A. F. Molisch. "Spectral shaping of UWB signals for time-hopping impulse radio". In: *Selected Areas in Communications, IEEE Journal on* **24**: (2006), pp. 738 –744.

[OSB99] A. V. Oppenheim, R. W. Schafer, and J. A. Buck. *Discrete-time signal processing*. Prentice Hall, 1999.

[OYDS12] H. Ohlsson, A. Y. Yang, R. Dong, and S. S. Sastry. "Compressive Phase Retrieval from Squared Output Measurements via Semidefinite Programming". In: *16th IFAC Symposium on System Identification* (2012).

[Pal04] A. R. Palacios. "Absolute-Valued Algebras, and Absolute-Valuable Banach Spaces". In: (2004), pp. 99–155.

[Pal94] W. T. Palmer. *Banach Algebras and The General Theory of *-Algebras - Volume I: Algebras and Banach Algebras*. Cambridge University Press, 1994.

[PCWD03] B. Parr, B. Cho, K. Wallace, and Z. Ding. "A Novel Ultra-Wideband Pulse Design Algorithm". In: *IEEE Commun. Lett.* **7**:5 (2003), pp. 219–222.

[Pis89] G. Pisier. *The volume of convex bodies and Banach space geometry*. Cambridge University Press, 1989, p. 250.

[Pro01] J. G. Proakis. *Digital Communications*. Singapore: McGraw Hill, 2001.

[RCLV13] J. Ranieri, A. Chebira, Y. Lu, and M. Vetterli. "Phase Retrieval for Sparse Signals: Uniqueness Conditions". In: *IEEE Transactions on Information Theory* **submitted**: (2013).

[RFP10] B. Recht, M. Fazel, and P. A. Parrilo. "Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization". In: *SIAM Journal on Applied Mathematics* **52**: (2010), pp. 471–501.

[RM07] F. Ramirez-Mireles. "Performance of UWB N-Orthogonal PPM in AWGN and Multipath Channels". In: *IEEE Transactions on Vehicular Technology*. 2007, pp. 1272–1285.

[RMS98] F. Ramirez-Mireles and R. A. Scholtz. "Multiple-access with time hopping and block waveform PPM modulation". In: *IEEE Int. Conf. Communications*. 1998.

[Rob09] J. C. Robinson. "Linear embeddings of finite-dimensional subsets of Banach spaces into Euclidean spaces". In: *Nonlinearity* **22**: (2009), pp. 7111–728.

[RS94] A. Ron and Z. Shen. *Frames and Stable Bases for Shift-Invariant Subspaces of $L^2(\mathbb{R}^d)$*. Tech. rep. Wisconsin University Madison, 1994.

[Rud60] W. Rudin. "Trigonometric Series with Gaps". In: (1960).

[Rud62] W. Rudin. *Fourier Analysis on Groups*. Interscience, 1962.

[Rus91] I. Rusza. "Arithmetic progressions in sumsets". In: *Acta Arith.* **60**: (1991), pp. 191–202.

[SBE13] Y. Shechtman, A. Beck, and Y. Eldar. "GESPAR: Efficient Phase Retrieval of Sparse Signals". In: *arxiv* (2013).

[Sch93] R. A. Scholtz. "Multiple access with time-hopping impulse modulation". In: *IEEE Military Communications Conference*. Vol. 2. Boston, MA, 1993, pp. 447–450.

[SHL94] M. K. Simon, S. M. Hinedi, and W. C. Lindsey. *Digital communication techniques: signal design and detection*. Prentice Hall, 1994, 912p.

[SJB03] F.-W. Sun, Y. Jiang, and J. Baras. "On the convergence of the inverses of Toeplitz matrices and its applications". In: *IEEE Transactions on Information Theory* **49**:1 (2003), pp. 180–190.

[SL96] S. Shamai and R. Laroia. "The intersymbol Interference Channel: Lower bounds on capacity andchannel precoding loss". In: *Information Theory, IEEE Transactions on* **42**: (1996), pp. 1388–1404.

[SS03] E. M. Stein and R. Shakarchi. *Fourier Analysis*. Princeton University Press, 2003.

[Str86] G. Strang. "A Proposal for Toeplitz Matrix Calculations". In: *Stud. Apply. Math.* **74**: (1986), pp. 171–176.

[Stu99] J. F. Sturm. "Using SeDuMi 1.02, A MATLAB toolbox for optimization over symmetric cones". In: *Optimization Methods and Software* **11**:1 (1999), pp. 625–653.

[SW71] E. M. Stein and G. Weiss. *Fourier Analysis on Euclidean Spaces*. Princeton, New Jersey: Princeton University Press, 1971.

[Tao05] T. Tao. "An uncertainty principle for cyclic groups of prime order". In: *Math. Research Letters.* **12**: (2005), pp. 121–127.

[Tia+06] Z. Tian et al. *Ultra wideband wireless communication:ultra wideband pulse shaper design*. Wiley, 2006.

[Tre68] H. L. van Trees. *Detection, Estimation, and Modulation Theory*. Wiley, 1968.

[TV06] T. Tao and V. Vu. *Additive Combinatorics*. Cambridge University Press, 2006.

[Vai93] P. Vaidyanathan. *Multirate Systems and Filter Banks*. Prentice-Hall Series in Signal Processing, 1993.

[Ver11] R. Vershynin. *Lectures in Geometric Functional Analysis*. 2011.

[VRE05] S. Villarreal-Reyes and R. Edwards. "Spectrum shaping of PPM TH-IR based ultra wideband systems by means of the PPM modulation index and pulse doublets". In: *Antennas and Propagation Society International Symposium, 2005 IEEE*. 2005.

[Wan13] Y. Wang. "Minimal Frames for Phase Retrieval". In: *Workshop of phaseless recovery*. 2013.

[WBV99] S.-P. Wu, S. Boyd, and L. Vandenberghe. *Applied and Computational Control, Signals and Circuits: 5. FIR filter design via spectral factorization and convex optimization*. Ed. by B. Datta. Birkhauser, 1999, pp. 215–245.

[Web94] R. Webster. *Convexity*. Oxford University Press, 1994.

[Wer02] D. Werner. *Funktionalanalysis*. German. 4th ed. Springer, 2002, p. 503.

[Win02] M. Z. Win. "Spectral density of random UWB signals". In: *IEEE Commun. Lett.* **6**:12 (2002), pp. 526–528.

[Win99] M. Z. Win. “Spectral density of random time-hopping spread-spectrum UWB signals with uniform timing jitter”. In: *IEEE Military Communications Conference*. Vol. 2. 1999, pp. 1196–1200.

[WJ12a] P. Walk and P. Jung. “Approximation of Löwdin Orthogonalization to a Spectrally Efficient Orthogonal Overlapping PPM Design for UWB Impulse Radio”. In: *EURASIP Journal on Applied Signal Processing* **92**: (2012), pp. 649–666.

[WJ12b] P. Walk and P. Jung. “Compressed sensing on the image of bilinear maps”. In: *IEEE International Symposium on Information Theory*. 2012, pp. 1291 –1295.

[WJ13a] P. Walk and P. Jung. “On a Reverse ℓ_2–inequality for Sparse Circular Convolutions”. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. 2013, pp. 4638 –4642.

[WJ13b] D. Wieruch and P. Jung. “Compressive Gray Space Detection for Interweaved Cognitie Radio Systems”. In: *ICWS*. 2013.

[WJT10a] P. Walk, P. Jung, and J. Timmermann. “Löwdin Transform on FCC Optimized UWB Pulses”. In: *IEEE WCNC*. 2010.

[WJT10b] P. Walk, P. Jung, and J. Timmermann. “Löwdin’s approach for orthogonal pulses for UWB impulse radio”. In: *IEEE Workshop on Signal Processing Advances in Wireless Communications*. 2010.

[WS98] M. Z. Win and R. A. Scholtz. “Impulse radio: how it works”. In: *IEEE Commun. Lett.* **2**:2 (1998), pp. 36–38.

[WTDG04] X. Wu, Z. Tian, T. N. Davidson, and G. B. Giannakis. “Orthogonal Waveform Design for UWB Radios”. In: *Proceedings of the IEEE Signal Processing Workshop on Advances in Wireless Communications*. 2004, pp. 11–14.

[WTDG06] X. Wu, Z. Tian, T. Davidson, and G. Giannakis. “Optimal waveform design for UWB radios”. In: *IEEE Trans. Signal Process.* **54**: (2006), pp. 2009–2021.

[WX13] Y. Wang and Z. Xu. “Phase Retrieval for Sparse Signals”. In: *arxiv* (2013).

[WZK08] S.-l. Wang, J. Zhou, and H. Kikuchi. “Performance of Overlapped and Orthogonal TH-BPPM UWB Systems”. In: *IEEE Conference on Wireless Communications, Networking and Mobile Computing*. 2008.

[Young12] W. H. Young. “On the multiplication of successions of Fourier constants”. In: *Proc. Roy. Soc. Lond. Series A* **87**: (1912), 331—339.

[Young13] W. H. Young. “On the determination of the summability of a function by means of its Fourier constants”. In: *Proc. London Math. Soc.* **s2-12**:1 (1913), pp. 71–88.

[You80] R. M. Young. *An Introduction to Nonharmonic Fourier Series*. Academic Press, 1980.

[ZMSF01] S. Zeisberg, C. Müller, J. Siemes, and A. Finger. “PPM based UWB system throughput optimisation”. In: *IST Mobile Communications Summit*. 2001.

[ZZMW09] Y. Zhu, J. Zuegel, J. Marciante, and H. Wu; “Distributed Waveform Generator: A New Circuit Technique for Ultra-Wideband Pulse Generation, Shaping and Modulation”. In: *IEEE Journal of Solid-State Circuits* **44**:3 (2009), pp. 808 –823.