



universität
wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation / Title of the Doctoral Thesis

„Topics in time-frequency analysis“

verfasst von / submitted by

Irina Shafkulovska, B.Sc. RWTH M.Sc. RWTH

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Doktorin der Naturwissenschaften (Dr. rer. nat)

Wien, 2024 / Vienna 2024

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on the student
record sheet:

A 796 605 405

Dissertationsgebiet lt. Studienblatt /
field of study as it appears on the student record sheet:

Mathematik

Betreut von / Supervisor:

Univ.-Prof. Mag. Dr. Karlheinz Gröchenig

Mitbetreut von / Co-Supervisor:

Dr. Markus Faulhuber, BSc MSc

Abstract

This thesis consists of six papers in the field of time-frequency analysis and sampling theory. The common thread is the so-called short-time Fourier transform (STFT), an integral transform that arises as the representation coefficient of the Schrödinger representation of the Heisenberg group.

We study four types of problems, namely, (i) sampling and the existence of new Gabor frames, (ii) optimization of point configurations, (iii) phase retrieval of the short-time Fourier transform, and (iv) uncertainty principles for joint time-frequency representations.

We construct new Gabor frames for two types of windows: Hermite functions and periodic exponential B-splines. For the Hermite functions, we rely on the Janssen representation of the Gabor frame operator, specifically, the Janssen test, to establish new Gabor frames for Hermite functions. For the periodic exponential B-splines, we derive a near-optimal result on sampling with derivatives in shift-invariant spaces generated by such windows. As a direct consequence, we also obtain new Gabor frames for periodic exponential B-splines and separated sets with density arbitrarily close to the critical density.

We also investigate the frame bound behaviour of Gabor frames on rectangular lattices. In particular, we show that if the window is the hyperbolic secant, an eigenfunction of the Fourier transform, then among all rectangular lattices of integer density, the square lattice is the unique optimizer of both bounds, and thus has the best condition number. We employ similar techniques to provide new insight into the open problem of universal optimality in two dimensions. In particular, we exclude that the honeycomb structure, the natural non-lattice contender of the conjectured optimal hexagonal lattice, is universally optimal.

We continue with the phase retrieval problem for the short-time Fourier transform: a nonlinear, non-stable problem that deals with the uniqueness of phaseless samples of the short-time Fourier transform. We focus solely on the recovery of compactly supported functions from phaseless samples of their short-time Fourier transform with respect to particularly chosen holomorphic windows, not necessarily the Gaussian window.

We conclude with uncertainty principles for metaplectic time-frequency representations of square-integrable functions in the spirit of Benedicks's theorem. While such an uncertainty principle already exists for a subclass of metaplectic time-frequency representations, including the classical time-frequency representations, there was a lack of an understanding of the underlying group structure which would include all existing results as special cases in a unified framework for metaplectic time-frequency representations. We characterize the structure of metaplectic operators whose associated time-frequency representation allows for a Benedicks-type of an uncertainty principle, both in the sesquilinear and the quadratic version.

Zusammenfassung

Diese Arbeit umfasst sechs Artikel in dem Bereich der Zeit-Frequenz-Analyse und Abtasttheorie. Der rote Faden ist die Kurzzeit-Fourier-Transformation (englisch: *Short-time Fourier transform*, kurz *STFT*), eine Integraltransformation, die eng mit der Schrödinger-Darstellung der Heisenberggruppe verbunden ist.

Wir beschäftigen uns mit vier Hauptthemen, nämlich: (i) Abtasten und die Existenz neuer Gabor-Frames, (ii) Optimierung von Punktkonfigurationen, (iii) Phase-Retrieval der Kurzzeit-Fourier-Transformation und (iv) Unschärferelationen für simultane Zeit-Frequenz-Darstellungen.

Wir konstruieren neue Gabor-Frames für zwei Arten von Fensterfunktionen: Hermite-Funktionen und periodische exponentielle B-Splines. Für die Hermite-Funktionen verwenden wir den Janssen-Test, um neue Gabor-Frames für Hermite-Funktionen zu erzeugen. Für die periodischen exponentiellen B-Splines erzielen wir fast optimale Ergebnisse bei der Auswahl von Abtastmengen mit Multiplizitäten in translationsinvarianten Räumen, die von solchen Fensterfunktionen generiert werden. Als Folgerung erhalten wir neue Gabor-Frames für periodische exponentielle B-Splines und Mengen mit Dichte beliebig nahe an der kritischen Dichte.

Wir analysieren auch das Verhalten der Frame-Schranken von Gabor-Frames auf Rechtecksgittern. Insbesondere zeigen wir für den Sekans Hyperbolicus, eine Eigenfunktion der Fouriertransformation, dass unter allen Rechtecksgittern fester ganzzahliger Dichte das quadratische Gitter die besten Frame-Schranken aufweist. Daher besitzt das quadratische Gitter die minimale Konditionszahl. Wir wenden ähnliche Methoden an, um das Verständnis des offenen Problems der universellen Optimalität in zwei Dimensionen zu erweitern und schließen aus, dass die Honeycomb-Struktur universell optimal ist, obwohl sie eine natürliche Konkurrenz des hexagonalen Gitters darstellt.

Des Weiteren untersuchen wir das Phase-Retrieval-Problem für die Kurzzeit-Fourier-Transformation. Dies ist ein nicht-lineares, instabiles Problem, das sich mit der Kurzzeit-Fourier-Transformation befasst. Wir konzentrieren uns auf die Bestimmung kompakt getragener Funktionen von phasenlosen Abtastungen deren Kurzzeit-Fourier-Transformation bezüglich bestimmter holomorpher Fensterfunktionen, die nicht unbedingt die Gaussfunktion sind.

Zum Abschluss diskutieren wir Unschärferelationen für Zeit-Frequenz-Darstellungen im Sinne des Satzes von Benedicks. Obwohl solche Unschärferelationen für eine Unterklasse der metaplektischen Zeit-Frequenz-Darstellungen, einschließlich der klassischen, bereits vorhanden sind, fehlte bisher eine umfassende Beschreibung der zugrundeliegenden Gruppenstruktur, die alle existierenden Resultate in einem einheitlichen Rahmen zusammenführt. Wir charakterisieren die Struktur der metaplektischen Operatoren, deren Zeit-Frequenz-Darstellung einer Unschärferelation analog zu der Benedicks Unschärferelation unterliegt, sowohl in der sesquilinearen als auch in der quadratischen Version.

Acknowledgements

I would like to begin by thanking my supervisor Karlheinz Gröchenig for his invaluable advice, expertise, and endless supply of time during my PhD study. Thank you for your patience and guidance through the daunting first steps in my academic career.

This thesis would not have come to be without my co-supervisor Markus Faulhuber. Thank you for giving me the chance to come to Vienna, for the gentlest of introductions to a researcher's day-to-day, sharing your mathematical knowledge with me, and being the kindest project leader one could wish for.

My sincere gratitude goes to Hartmut Führ, Philipp Grohs, Anupam Gumber, Lukas Liehr, and Ilia Zlotnikov, my other co-authors of the articles included here and those beyond the scope of this thesis. Thank you for sharing new engaging problems with me and teaching me different important facets of the art of collaboration.

I would like to thank all of the members of the Numerical Harmonic Analysis Group at the University of Vienna. I will always appreciate the many helpful ears, discussions in the office and the seminar, and the great time at conferences. Thank you for turning from colleagues into friends.

I would like to thank Peter Balazs and Nicki Holighaus, and the Acoustics Research Institute at the Austrian Academy of Sciences. Thank you for the chance to experience a research-wise vastly diverse working environment, and the valuable lessons on interdisciplinary work that I could only learn from you.

This acknowledgement is not complete without all of my friends who were available at all times and tolerated my scattered thoughts and somewhat scarce phone calls. Thank you for listening to the impromptu maths lectures and reminding me of everything beyond mathematics.

Last but not least, I would like to thank my family, especially my parents and my sister, for their love, support, and belief in me. Thank you for keeping me strong enough to keep going.

Contents

1	Preamble	11
1.1	Introduction	13
1.2	Sampling problems	16
1.3	Frame theory	17
1.4	Shift-invariant spaces	19
1.5	Phase retrieval of the short-time Fourier transform	20
1.6	Metaplectic time-frequency representations	21
1.7	Uncertainty principles	23
1.8	Universal optimality	24
2	Discussion of the main results	29
2.1	Sampling theorems with derivatives in shift-invariant spaces generated by periodic exponential B-splines	32
2.2	On the frame property of Hermite functions and analytic extensions of their frame sets	36
2.3	Gabor frame bound optimizations	38
2.4	A note on energy minimization in dimension 2	40
2.5	From completeness of discrete translates to phaseless sampling of the short-time Fourier transform	42
2.6	Benedicks-type uncertainty principle for metaplectic time-frequency representations	46
	References	50
3	Publications	61
3.1	Sampling theorems with derivatives in shift-invariant spaces generated by periodic exponential B-splines	63
3.2	On the frame property of Hermite functions and analytic extensions of their frame sets	105
3.3	Gabor frame bound optimizations	133
3.4	A note on energy minimization in dimension 2	195
3.5	From completeness of discrete translates to phaseless sampling of the short-time Fourier transform	211
3.6	Benedicks-type uncertainty principle for metaplectic time-frequency representations	249

1 Preamble

1.1 Introduction

We start by introducing the elementary objects that we will encounter throughout the thesis. We denote by $L^2(\mathbb{R}^d)$ the space of square-integrable functions on $\mathbb{R}^d$, endowed with the scalar product and the norm

$$\langle f, g \rangle = \int_{\mathbb{R}^d} f(t) \overline{g(t)} dt, \quad \|f\|_2 = \sqrt{\langle f, f \rangle}, \quad \text{respectively.}$$

The Euclidean scalar product of vectors $x, \omega \in \mathbb{R}^d$ is denoted with $x \cdot \omega := x^t \omega$, and the Euclidean norm is denoted with $|x| := \sqrt{x \cdot x}$.

We denote by $\mathcal{F}f = \hat{f}$ the Fourier transform of a function $f \in L^1(\mathbb{R}^d)$;

$$\mathcal{F}f(\omega) = \hat{f}(\omega) = \int_{\mathbb{R}^d} f(t) e^{-2\pi i \omega \cdot t} dt, \quad \omega \in \mathbb{R}^d.$$

The Fourier transform extends from the dense subset $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d) \subseteq L^2(\mathbb{R}^d)$ to a unitary operator on $L^2(\mathbb{R}^d)$ [53, 130]. Throughout the thesis, we denote this extension with $\mathcal{F}$.

Let $x, \omega \in \mathbb{R}^d$ and let $f \in L^2(\mathbb{R}^d)$ be a square-integrable function. The translation (time shift) by x and the modulation (frequency shift) by ω are the unitary operators T_x and M_ω defined as

$$T_x f(t) = f(t - x) \quad \text{and} \quad M_\omega f(t) = e^{2\pi i t \cdot \omega} f(t)$$

respectively. Their composition

$$\pi(x, \omega) = M_\omega T_x$$

is called a time-frequency shift. These elementary operators are the key to an efficient analysis of the time-frequency content of square-integrable functions, and have a simple interaction with the Fourier transform $\mathcal{F}$:

$$\mathcal{F}T_x = M_{-x}\mathcal{F}, \quad \mathcal{F}M_\omega = T_\omega\mathcal{F}.$$

They satisfy the commutation relation

$$M_\omega T_x = e^{2\pi i x \cdot \omega} T_x M_\omega,$$

i.e., π is a *projective representation* of $\mathbb{R}^{2d} \cong \mathbb{R}^d \times \mathbb{R}^d$, and in the following, we will investigate its representation coefficients.

Let $g \in L^2(\mathbb{R}^d)$ be a non-zero function. The *short-time Fourier transform* of $f \in L^2(\mathbb{R}^d)$ with respect to the *window* g is given by

$$V_g f(x, \omega) = \int_{\mathbb{R}^d} f(t) \overline{g(t - x)} e^{-2\pi i t \cdot \omega} dt, \quad (x, \omega) \in \mathbb{R}^{2d}.$$

Equivalently, $V_g f$ is the representation coefficient

$$V_g f(x, \omega) = \langle f, \pi(x, \omega)g \rangle, \quad (x, \omega) \in \mathbb{R}^{2d}.$$

In this context, we call $\mathbb{R}^{2d}$ the *phase space*. The operator $V_g : L^2(\mathbb{R}^d) \rightarrow L^2(\mathbb{R}^{2d})$ is (up to scaling) an isometry and its range is a reproducing kernel Hilbert space. In particular, whenever $g, h \in L^2(\mathbb{R}^d)$ are non-zero and not orthogonal to each other, they satisfy the Calderón-type reproducing formula

$$f(t) = \frac{1}{\langle h, g \rangle} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} V_g f(x, \omega) \pi(x, \omega) h(t) dx d\omega, \quad (1.1)$$

where the integral is to be understood in the weak sense [58].

Other prominent joint time-frequency representations are the (cross-) ambiguity function of two square-integrable functions $f, g \in L^2(\mathbb{R}^d)$

$$A(f, g)(x, \omega) := \int_{\mathbb{R}^d} f(t + \frac{x}{2}) \overline{g(t - \frac{x}{2})} e^{-2\pi i \omega t} dt, \quad (x, \omega) \in \mathbb{R}^{2d}$$

and the (cross-) Wigner distribution

$$W(f, g)(x, \omega) := \int_{\mathbb{R}^d} f(x + \frac{t}{2}) \overline{g(x - \frac{t}{2})} e^{-2\pi i \omega t} dt, \quad (x, \omega) \in \mathbb{R}^{2d}.$$

The short-time Fourier transform, the ambiguity function and the Wigner distribution have been instrumental in defining new function spaces [38, 53, 58], operator theory [17, 38, 58] as well as applications in quantum mechanics [44], radar signal processing [33, 122, 142] and optics [21].

This thesis explores several problems in time-frequency analysis tied to these joint time-frequency representations (see Figure 1). With the exception of Paper 6, the contributions in this manuscript solve problems which stem from or are in close relation to *Gabor analysis*.

The main focus of Gabor analysis is (stable) discretizations of (1.1), i.e., we investigate discrete point configurations $\Lambda \subseteq \mathbb{R}^{2d}$ and families of functions $(h_\lambda)_{\lambda \in \Lambda} \subseteq L^2(\mathbb{R}^d)$ such that any $f \in L^2(\mathbb{R}^d)$ can be represented as

$$f = \sum_{\lambda \in \Lambda} V_g f(\lambda) h_\lambda = \sum_{\lambda \in \Lambda} \langle f, \pi(\lambda)g \rangle h_\lambda, \quad (1.2)$$

with unconditional convergence of the series in the L^2 -norm. We adopt a twofold view of (1.2): as a sampling problem for the short-time Fourier transform (first series, see Subsection 1.2) and as a Gabor frame problem (second series, see Subsection 1.3). The minor difference is primarily evident in the first two Papers 1 and 2, both contributing with new discretizations of (1.1), where the first and the second point of view are reflected respectively. The remaining papers address different problems in the area and the distinction is not particularly present.

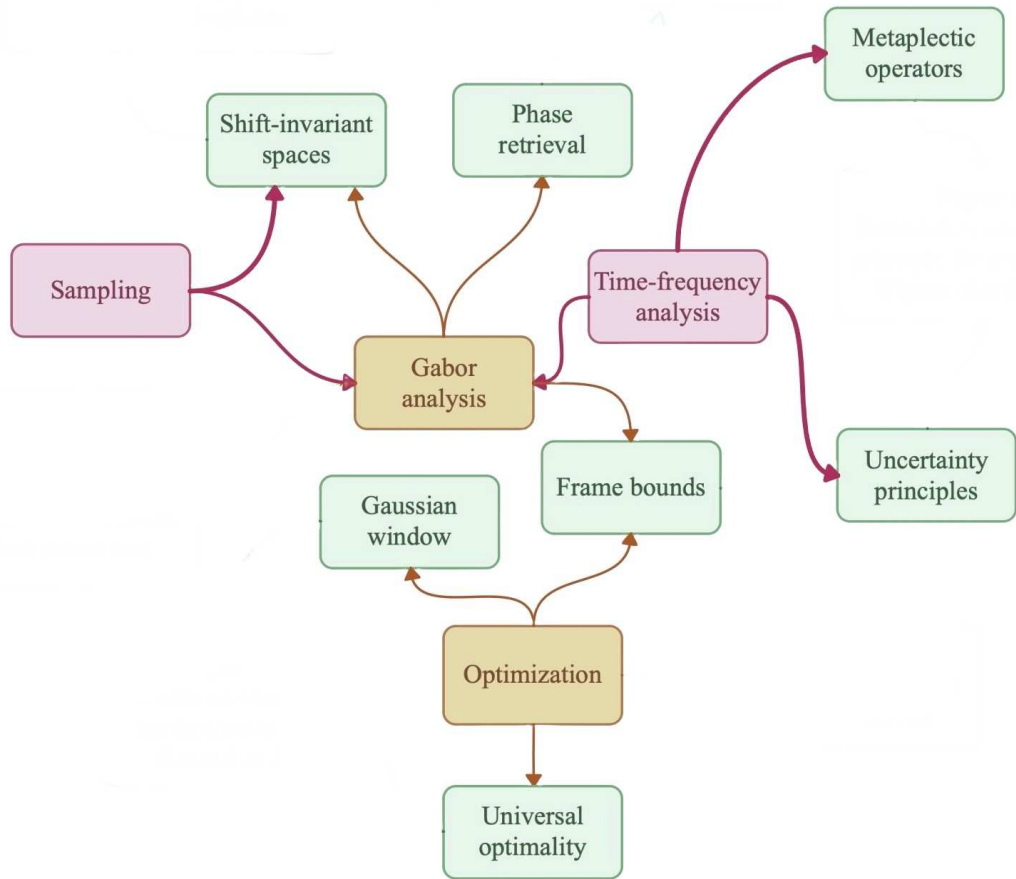


Figure 1: An overview of the areas of research included in this thesis. The two central themes of the thesis are *sampling problems* and *time-frequency analysis* (red), which branch out to topics (yellow) and subtopics (green).

1.2 Sampling problems

Sampling is a natural problem stemming from a vast spectrum of applications. In general, a continuous function f , often interpreted as a signal, is represented by samples of f on a discrete set. The natural question is whether and how well these samples represent the original function. Quoting verbatim from [10],

... the sampling problem consists of two main parts:

- (1) *Given a class of functions $\mathcal{H}$ on $\mathbb{R}^d$, find conditions on sampling sets $X = \{x_j \in \mathbb{R}^d : j \in J\}$, where J is a countable index set, under which a function $f \in \mathcal{H}$ can be reconstructed uniquely and stably from its samples $\{f(x_j) : x_j \in X\}$.*
- (2) *Find efficient and fast numerical algorithms which recover any function $f \in \mathcal{H}$ from its samples on X .*

We distinguish two types of sets. We call a countable set $X = \{x_j \in \mathbb{R}^d : j \in J\}$, $X \subseteq \mathbb{R}^d$, a *uniqueness set* for a class of functions $\mathcal{H}$ on $\mathbb{R}^d$ if the restriction operator

$$\mathcal{H} \rightarrow \mathbb{C}^X, \quad f \mapsto f|_X$$

is injective. If $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ is a Banach space and $(\ell(J), \|\cdot\|_{\ell(J)})$ is a Banach space of sequences indexed by a countable set J , then we call X a *sampling set* for $\mathcal{H}$ if there also exist constants $0 < A \leq B < \infty$ such that

$$A \|f\|_{\mathcal{H}} \leq \left\| \left(f(x_j) \right)_{j \in J} \right\|_{\ell(J)} \leq B \|f\|_{\mathcal{H}}, \quad f \in \mathcal{H}.$$

This holds if and only if the restriction operator (in this case also called the sampling operator) is injective, bounded and has a bounded inverse. Typically, $\mathcal{H}$ is a reproducing kernel Hilbert space and $\ell(J) = \ell^2(J)$ is the Hilbert space of square-summable sequences on J . In this case, the samples are given by

$$f(x_j) = \langle f, k_{x_j} \rangle, \quad j \in J,$$

where $(k_x)_{x \in X}$ are the reproducing vectors, and X is a sampling set if and only if the family $(k_{x_j})_{j \in J}$ is a frame for $\mathcal{H}$ as defined in the next subsection.

As stated in the introduction, the main type of sampling in this thesis is the sampling of the short-time Fourier transform. In this context, the reproducing kernel Hilbert space is the range $V_g(L^2(\mathbb{R}^d))$ of the short-time Fourier transform with respect to a window $g \in L^2(\mathbb{R}^d)$, $g \not\equiv 0$, and we speak of (stable) sampling of the short-time Fourier transform.

Closely related as well as of independent interest is sampling in shift-invariant spaces. In this case, the space $\mathcal{H}$ is given by

$$V^p(g) = \left\{ f \in L^p(\mathbb{R}) : f = \sum_{\ell \in \mathbb{Z}} c_{\ell} T_{\ell} g, \ (c_{\ell})_{\ell \in \mathbb{Z}} \in \ell^p(\mathbb{Z}) \right\} \subseteq L^p(\mathbb{R})$$

for a suitably chosen $g \in L^p(\mathbb{R}^d)$, and we study the stability with respect to $\ell^p(\mathbb{Z})$, $1 \leq p \leq \infty$. For $p = 2$, the space $V^2(g)$ is also a reproducing kernel Hilbert space.

The third case where we consider discrete evaluations is phase retrieval for the short-time Fourier transform, in which case our main objective is to study uniqueness sets.

1.3 Frame theory

Let $\mathcal{H}$ be a Hilbert space and $G = (g_\lambda)_{\lambda \in \Lambda} \subseteq \mathcal{H}$ a countable sequence in $\mathcal{H}$. We call G a *frame* for $\mathcal{H}$ if there exist constants $0 < A \leq B < \infty$ such that for all $f \in \mathcal{H}$

$$A \|f\|_{\mathcal{H}}^2 \leq \sum_{\lambda \in \Lambda} |\langle f, g_\lambda \rangle|^2 \leq B \|f\|_{\mathcal{H}}^2, \quad (1.3)$$

with unconditional convergence of the series. If the chain of inequalities (1.3) holds, we call the optimal A and B the lower and upper frame bound, respectively. The frame bounds exist if and only if the associated frame operator S_G , given by

$$S_G f = \sum_{\lambda \in \Lambda} \langle f, g_\lambda \rangle g_\lambda, \quad f \in \mathcal{H},$$

is bounded, with a bounded inverse. The operator S_G is self-adjoint and positive semi-definite, and the frame bounds are precisely the bounds of its spectrum.

Frames were first introduced by Duffin and Schaeffer [50] to treat a sampling problem for band-limited functions. We refer to [29] for an extensive overview of frame theory.

If G is a frame, then there exists a sequence $H = (h_\lambda)_{\lambda \in \Lambda} \subseteq \mathcal{H}$, called the *dual frame* of G , such that

$$f = \sum_{\lambda \in \Lambda} \langle f, g_\lambda \rangle h_\lambda = \sum_{\lambda \in \Lambda} \langle f, h_\lambda \rangle g_\lambda, \quad f \in \mathcal{H}. \quad (1.4)$$

The dual frame is a frame itself. It is not necessarily unique, but a distinguished dual frame is the canonical dual frame given by $(S_G^{-1} g_\lambda)_{\lambda \in \Lambda}$. If the dual frame is unique, then G is a *Riesz basis* of $\mathcal{H}$, and $\mathcal{H}$ is isomorphic to $\ell^2(\Lambda)$ via the *analysis operator*

$$C_G : \mathcal{H} \rightarrow \ell^2(\Lambda), \quad f \mapsto (\langle f, g_\lambda \rangle)_{\lambda \in \Lambda}$$

as well as the *synthesis operator*

$$C_G^* : \ell^2(\Lambda) \rightarrow \mathcal{H}, \quad (c_\lambda)_{\lambda \in \Lambda} \mapsto \sum_{\lambda \in \Lambda} c_\lambda g_\lambda.$$

In general, by (1.4), the left inverse of the analysis operator C_G is given by the synthesis operator of a dual frame, in particular, by the synthesis operator C_H^* of its canonical dual frame $H = (S_G^{-1} g_\lambda)_{\lambda \in \Lambda}$. Furthermore, their operator norms satisfy

$$\|C_G\|_{L^2 \rightarrow \ell^2} = \sqrt{B}, \quad \|C_H^*\|_{\ell^2 \rightarrow L^2} = \frac{1}{\sqrt{A}}.$$

Thus, the *condition number* of the reconstruction of $f \in \mathcal{H}$ from the coefficients $C_G f$ is given by the ratio $\sqrt{\kappa} = \frac{\sqrt{B}}{\sqrt{A}}$. Moreover, the convergence rate of the reconstruction of f from $S_G f$ via a Neumann series is $\frac{B-A}{B+A} = \frac{\kappa-1}{\kappa+1}$ [50]. Therefore, frames used in practice are typically chosen among those with a small condition number κ .

1.3.1 Gabor frames

In Gabor analysis, the Hilbert space is $\mathcal{H} = L^2(\mathbb{R}^d)$. Moreover, we consider a separated set $\Lambda \subseteq \mathbb{R}^{2d}$, i.e., Λ satisfies

$$\inf\{|\lambda - \gamma| : \lambda, \gamma \in \Lambda, \lambda \neq \gamma\} > 0,$$

and define the *Gabor system* of a window $g \in L^2(\mathbb{R}^d)$ along Λ as the collection of time-frequency shifts of g given by

$$\mathcal{G}(g, \Lambda) = \{\pi(\lambda)g : \lambda \in \Lambda\}.$$

We call a Gabor system a *Gabor frame* if it is a frame for $L^2(\mathbb{R}^d)$. In this case, we obtain precisely the desired discretization (1.2) and the inequality

$$A \|f\|_{L^2(\mathbb{R}^d)}^2 \leq \sum_{\lambda \in \Lambda} |\langle f, \pi(\lambda)g \rangle|^2 \leq B \|f\|_{L^2(\mathbb{R}^d)}^2.$$

Equivalently, the set Λ is a sampling set for the range $V_g(L^2(\mathbb{R}^d))$ of the short-time Fourier transform with respect to the window $g \in L^2(\mathbb{R}^d)$.

Determining whether a particular Gabor system $\mathcal{G}(g, \Lambda)$ is a frame is a rather difficult problem, and the only fully resolved case is the one of the one-dimensional Gaussian window $g(t) = 2^{1/4} e^{-\pi t^2}$ [104, 125, 126].

The well-known necessary density condition on the lower Beurling density $D^-(\Lambda)$ states that $D^-(\Lambda) \geq 1$ [30, 115], where the lower Beurling density is defined as

$$D^-(\Lambda) := \liminf_{R \rightarrow \infty} \inf_{z \in \mathbb{R}^{2d}} \frac{\#(\Lambda \cap B_R(z))}{\text{vol}(B_R(0))}, \quad \text{vol}(B_R(0)) = \frac{\pi^d R^{2d}}{d!}.$$

In general, most results deal primarily with lattices $\Lambda = A\mathbb{Z}^{2d}$, $A \in \text{GL}(2d, \mathbb{R})$, in which case a duality theory has been developed [42, 87, 121, 137].

Stepping away from discrete subgroups in phase space, the first relaxation of the lattice assumption is to consider semi-regular sets, for instance, $\Lambda = X \times \mathbb{Z}$, for some $X \subseteq \mathbb{R}$. In this setting, the problem of sampling the short-time Fourier transform is equivalent to sampling in an associated *shift-invariant space* [74].

We refer to the survey article [61] as well as the textbooks [44, 53, 58] for details on various approaches to the problem.

1.4 Shift-invariant spaces

Given a generator g subject to some mild decay conditions, the shift-invariant space generated by g in $L^p(\mathbb{R})$ is defined to be

$$V^p(g) := \left\{ f \in L^p(\mathbb{R}) : f = \sum_{\ell \in \mathbb{Z}} c_\ell T_\ell g, (c_\ell)_{\ell \in \mathbb{Z}} \in \ell^p(\mathbb{Z}) \right\} \subseteq L^p(\mathbb{R}).$$

If $g(x) = \frac{\sin \pi x}{\pi x}$, then $V^2(g) = \{f \in L^2(\mathbb{R}) : \text{supp } \hat{f} \subseteq [-\frac{1}{2}, \frac{1}{2}]\}$ is the Paley-Wiener space $PW^2(\mathbb{R})$ of band-limited functions. In this sense, shift-invariant spaces are a generalization of the space of band-limited functions. The benefit is the freedom to better incorporate particular features of the modelled problem, such as decay, convexity, or some local behaviour [64]. Throughout, we will assume that g has *stable shifts*, i.e., there exist constants $0 < A_p \leq B_p < \infty$ such that

$$A_p \|c\|_p^p \leq \left\| \sum_{\ell \in \mathbb{Z}} c_\ell T_\ell g \right\|_p^p \leq B_p \|c\|_p^p, \quad c = (c_\ell)_{\ell \in \mathbb{Z}} \in \ell^p(\mathbb{Z}),$$

with the usual modification for $p = \infty$. In particular, if g has stable shifts, then $V^p(g)$ is a Banach space when endowed with the L^p -norm. Furthermore, if $p = 2$, then $(T_\ell g)_{\ell \in \mathbb{Z}}$ is a Riesz basis for $V^2(g)$, and $V^2(g)$ is a reproducing kernel Hilbert space [9]. The upper bound holds under minimal conditions, but the lower bound is quite sensitive to the properties of the generator. We refer to Ron's survey [120], and results in [9, 10, 75, 91, 106] for further details.

Note that, if the generator g has stable translates, then a set X is sampling for $V^p(g)$, cf. Subsection 1.2, if there exist constants $0 < A_p \leq B_p < \infty$ such that

$$A_p \|c\|_p^p \leq \left\| \sum_{x \in X} \sum_{\ell \in \mathbb{Z}} c_\ell T_\ell g(x) \right\|_p^p \leq B_p \|c\|_p^p, \quad c = (c_\ell)_{\ell \in \mathbb{Z}} \in \ell^p(\mathbb{Z}),$$

with the usual modification for $p = \infty$. Furthermore, the stability of the translates enables the following connection to Gabor frames.

Theorem 1.1 (Gröchenig, Romero, Stöckler [75]). *Assume that $g \in C(\mathbb{R})$ decays as $|\varphi(x)| \leq C(1 + |x|)^{-1-\varepsilon}$ for $C, \varepsilon > 0$, and has stable integer shifts. Let $X \subseteq \mathbb{R}$ be a separated set. Then the following are equivalent:*

- (i) *The family $\mathcal{G}(g, (-X) \times \mathbb{Z})$ is a frame for $L^2(\mathbb{R})$.*
- (ii) *X is a sampling set for $V^p(g)$ for some $p \in [1, \infty]$.*
- (iii) *X is a sampling set for $V^p(g)$ for all $p \in [1, \infty]$.*

This connection between Gabor frames and shift-invariant spaces has already been exploited for totally positive functions and with tools from complex analysis, e.g., in [63, 75]. The main issue remains to be choosing a generator where the sufficient density would be close to known necessary conditions [30, 115].

1.5 Phase retrieval of the short-time Fourier transform

The term *phase retrieval problem* stands for the problem of recovery of a function $f : \Omega \rightarrow \mathbb{C}$ which belongs to a class of functions $\mathcal{H}$, typically a reproducing kernel Hilbert space, from its phaseless samples $|f(x)|$, $x \in X$, for some subset $X \subseteq \Omega$. Clearly, if f is not uniquely determined from the samples $f(x)$, $x \in X$, then it cannot be determined from the phaseless samples. Otherwise, to recover f is equivalent to recovering the samples

$$f(x) = |f(x)|e^{2\pi i\psi(x)}, \quad x \in X,$$

i.e., it suffices to determine the phase function ψ , hence the recovery problem is called a *phase retrieval* problem.

One of the first phase retrieval problems goes back to Pauli [113], who posed the question of whether a function is uniquely determined by $|f|$ and its phaseless Fourier transform $|\hat{f}|$. Although the answer is negative [34, 81, 84], it opened a wide range of phase retrieval problems in related settings, such as the phase retrieval problem for the short-time Fourier transform.

While classical Gabor analysis focuses on recovering a function f from $V_g f$, the STFT phase retrieval problem asks for the reconstruction of $f \in \mathcal{H} \subseteq L^2(\mathbb{R}^d)$ from the phaseless evaluations of $|V_g f|$, or equivalently, from the evaluations of the spectrogram $\text{SPEC}_g f = |V_g f|^2$. Since V_g is a linear operator, for all $g, f \in L^2(\mathbb{R}^d)$ and $\tau \in \mathbb{T}$ holds

$$|V_g(\tau f)| = |V_g f|.$$

Therefore, a function can only be recovered up to a *global phase*, i.e., at best, one recovers from $|V_g f|$ a representative τf for some $\tau \in \mathbb{T}$. Formally, we introduce the equivalence relation

$$f \sim h \quad :\Leftrightarrow \quad \exists \tau \in \mathbb{T} : h = \tau f.$$

The phase retrieval problem for the short-time Fourier transform on $\mathcal{H}$ asks whether

$$\mathcal{H}/\sim \rightarrow [0, \infty)^\Lambda, \quad f \mapsto \left(|V_g f(\lambda)|\right)_{\lambda \in \Lambda}$$

is injective on the quotient space $\mathcal{H}/\sim$. If this is the case, we call the set $\Lambda \subseteq \mathbb{R}^{2d}$ a *uniqueness set for the STFT phase retrieval problem* on $\mathcal{H}$.

The problem has received a lot of attention due to many applications [1, 41, 92, 145], and both theoretical [7, 25, 73] as well as numerical aspects [55, 107] have been studied. We refer to the surveys [51, 67] for an overview.

The phase retrieval problem for the short-time Fourier transform, unlike many classical problems in time-frequency analysis, is a non-linear problem which seems to defy most pre-existing methods. Firstly, the reconstruction is in general highly unstable [7, 25]. Secondly, for all windows g sufficiently well-concentrated in the time-frequency plane, for all sufficiently dense sets Λ , the Gabor system $\mathcal{G}(g, \Lambda)$ is a frame [58]. In particular, two functions f, h whose short-time Fourier samples agree

on Λ must be identical. In contrast, initially shown for the Gaussian window [8], phaseless sampling poses the following discretization barrier for lattices [68].

Theorem 1.2 (Grohs, Liehr [68]). *Let $g \in L^2(\mathbb{R})$ and let $\Lambda = A\mathbb{Z}^2$, $A \in \text{GL}(2, \mathbb{R})$, be a lattice. Then there exist functions $f, h \in L^2(\mathbb{R})$, such that $f \not\sim h$, but*

$$|V_g f(\lambda)| = |V_g h(\lambda)|, \quad \text{for all } \lambda \in \Lambda.$$

This result was later generalized to the multivariate setting [70]. There are several ways to overcome the non-uniqueness issue posed in Theorem 1.2. For instance, one could sample the spectrogram on the union of a lattice Λ and non-trivial perturbations of Λ [72], modify the setting to a multi-window phase retrieval problem [71], or study approximate solutions [28, 82, 86]. Alternatively, one could restrict to a subclass of functions such as a shift-invariant space [26, 27, 59, 119], in particular, the Paley-Wiener space $PW^2(\mathbb{R}^d)$ [69, 136]. Note that, in the latter case, this is equivalent to recovering compactly supported functions.

1.6 Metaplectic time-frequency representations

The classical joint time-frequency representations $V_g f$, $A(f, g)$ and $W(f, g)$ share many features, e.g., the format of the reconstruction formulas, orthogonality relations, and decay properties [58], because they all follow the same pattern of construction. To see this, let $\mathcal{F}_2$ denote the partial Fourier transform in the last d variables. The short-time Fourier transform, the Wigner distribution, and the ambiguity function can all be written as the partial Fourier transform (in the last d variables) of a dilation of the elementary tensor $f \otimes \bar{g} \in L^2(\mathbb{R}^{2d})$, i.e., in each case, there exists a matrix $L \in \text{GL}(4d, \mathbb{R})$ such that the joint time-frequency representation of f and g can be written as

$$\mathcal{F}_2((f \otimes \bar{g})(L^{-1} \cdot)).$$

Later on, we will see that this is a special case of a metaplectic time-frequency representation as introduced by Cordero and Rodino in [39].

1.6.1 The symplectic group

Just like the classical time-frequency shifts, the symmetric time-frequency shifts

$$\rho(x, \omega) := T_{x/2} M_\omega T_{x/2} = M_{\omega/2} T_x M_{\omega/2}$$

are also a projective representation of $\mathbb{R}^{2d}$, satisfying

$$\rho(\lambda + \gamma) = e^{\pi i \lambda^t \mathcal{J} \gamma} \rho(\lambda) \rho(\gamma) = e^{\pi i \sigma(\lambda, \gamma)} \rho(\lambda) \rho(\gamma), \quad \lambda, \gamma \in \mathbb{R}^{2d}. \quad (1.5)$$

Hereby, $\mathcal{J}$ is the standard skew-symmetric matrix

$$\mathcal{J} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \in \text{GL}(2d, \mathbb{R}), \quad (1.6)$$

and $\sigma(\lambda, \gamma) = \lambda^t \mathcal{J} \gamma$ is its associated bilinear form.

The symplectic group $\mathrm{Sp}(2d, \mathbb{R}) \leq \mathrm{SL}(2d, \mathbb{R})$ consists of the matrices $\mathcal{A} \in \mathbb{R}^{2d \times 2d}$ which preserve the phase factor in (1.5), i.e.,

$$\sigma(\mathcal{A}\lambda, \mathcal{A}\gamma) = \sigma(\lambda, \gamma), \quad \lambda, \gamma \in \mathbb{R}^d.$$

Equivalently,

$$\mathrm{Sp}(2d, \mathbb{R}) = \{ \mathcal{A} \in \mathrm{GL}(2d, \mathbb{R}) : \mathcal{A}^t \mathcal{J} \mathcal{A} = \mathcal{J} \}.$$

The projective representation ρ can be lifted to a unitary representation of the reduced Heisenberg group $\mathbb{H}$, which is topologically the product space of $\mathbb{R}^{2d}$ and the torus $\mathbb{T}$. The extension of ρ on $\mathbb{H}$ acts as

$$\rho(x, \omega, e^{i\alpha}) = e^{i\alpha} T_{x/2} M_\omega T_{x/2}.$$

From this point of view, the symplectic group is precisely the subgroup of matrices $\mathcal{A} \in \mathrm{GL}(2d, \mathbb{R})$ whose trivial extension on the Heisenberg group

$$(\lambda, e^{\pi i \alpha}) \mapsto (\mathcal{A}\lambda, e^{\pi i \alpha}), \quad \lambda \in \mathbb{R}^{2d}, e^{\pi i \alpha} \in \mathbb{T},$$

is a group automorphism on the Heisenberg group. The elementary examples of unitary matrices are the standard skew-symmetric matrix $\mathcal{J}$ as well as

$$V_Q := \begin{pmatrix} I & 0 \\ Q & I \end{pmatrix}, \quad D_L := \begin{pmatrix} L & 0 \\ 0 & L^{-t} \end{pmatrix}, \quad Q = Q^t \in \mathbb{R}^{d \times d}, L \in \mathrm{GL}(d, \mathbb{R}).$$

The symplectic group is a connected, non-compact Lie subgroup of the special linear group $\mathrm{SL}(2d, \mathbb{R})$ [43, 96]. Its double cover is called the *metaplectic group*, denoted by $\mathrm{Mp}(2d, \mathbb{R})$.

1.6.2 Metaplectic operators

By the Stone-von Neumann theorem [53, 58], for any symplectic matrix $\mathcal{A} \in \mathrm{Sp}(2d, \mathbb{R})$ there exists a unitary operator $\mu(\mathcal{A})$ with

$$\rho(\mathcal{A}\lambda) = \mu(\mathcal{A})\rho(\lambda)\mu(\mathcal{A})^{-1},$$

i.e., $\mu(\mathcal{A})$ arises as an intertwining operator. The operator is unique up to multiplication with a scalar of modulus one. We consider some standard examples:

$$\begin{aligned} \mu(\mathcal{D}_L)f(x) &= |\det L|^{-1/2} f(L^{-1}x), \\ \mu(\mathcal{V}_Q)f(x) &= e^{\pi i x^t Q x} f(x), \\ \mu(\mathcal{J}) &= \mathcal{F}. \end{aligned} \tag{1.7}$$

It turns out that it is not possible to make a particular choice of operators such that $\mathcal{A} \mapsto \mu(\mathcal{A})$ is a group isomorphism. However, it is possible to pick *two* intertwining operators, $\mu(\mathcal{A})$ and $-\mu(\mathcal{A})$, such that the projection

$$\mu(\mathcal{A}) \mapsto \mathcal{A}, \quad -\mu(\mathcal{A}) \mapsto \mathcal{A}$$

is a group epimorphism. This is often referred to as the Weil representation of the metaplectic group [135] since this allows us to identify the double cover $\text{Mp}(2d, \mathbb{R})$ with a group of unitary operators acting on $L^2(\mathbb{R}^d)$. Note that the projection preserves the group structure. Thus, to understand the action of the metaplectic operator, it suffices to know the metaplectic operators which project to generators of the symplectic group, as well as the interaction between these distinguished operators. The fact that many standard operators, such as the Fourier transform, the dilations and the fractional Fourier transforms are metaplectic operators allows us to study them under a unified framework [133].

As already remarked, the classical joint time-frequency representations can be written as a partial Fourier transform of a suitable dilation $\mu(\mathcal{D}_L)(f \otimes \bar{g})$ of an elementary tensor $f \otimes \bar{g} \in L^2(\mathbb{R}^{2d})$. Since this is the action of a *composition of two metaplectic operators* on $f \otimes \bar{g}$, thus the action of a *metaplectic operator* on $f \otimes \bar{g}$, we can write each of the joint-time frequency representations (up to a multiplicative constant of modulus one) as $\mu(\mathcal{A})(f \otimes \bar{g})$ for a suitable matrix $\mathcal{A} \in \text{Sp}(4d, \mathbb{R})$ (note the doubling of the dimension). This motivates the following definition.

Definition 1.3. Let $\mu(\mathcal{A}) \in \text{Mp}(4d, \mathbb{R})$ be a metaplectic operator and $f, g \in L^2(\mathbb{R}^d)$. The metaplectic time-frequency representation $W_{\mathcal{A}}(f, g)$ of f and g is defined as

$$W_{\mathcal{A}}(f, g)(\lambda) := \mu(\mathcal{A})(f \otimes \bar{g})(\lambda), \quad \lambda \in \mathbb{R}^{2d}.$$

These joint representations are clearly a generalization of the classical ones. They also appear in the literature as (generalized) $\mathcal{A}$ -Wigner distributions [39], and provide new families of frames [35] and an alternative tool for the analysis of particular common operators in time-frequency analysis [36, 37].

1.7 Uncertainty principles

A meta-theorem in time-frequency analysis says that

a function cannot be concentrated well both in time and in frequency.

This is clearly a very vague formulation and there is a wide range of options to measure the time-frequency content of a function [14, 16, 40, 49, 54, 78, 109].

At one end of the spectrum, there are support constraints such as Benedicks's [11, 16] and Donoho-Stark's uncertainty principle [49].

Theorem 1.4 (Benedicks's uncertainty principle). Let $f \in L^2(\mathbb{R}^d)$. Then

$$|\text{supp}(f)| \cdot |\text{supp}(\mathcal{F}f)| < \infty$$

holds if and only if $f \equiv 0$.

On the other end, uncertainty principles such as Hardy's uncertainty principle [78] restrict the maximal simultaneous decay of f and $\mathcal{F}f$.

Theorem 1.5 (Hardy's uncertainty principle). *Let $f : \mathbb{R} \rightarrow \mathbb{C}$ be a measurable function such that $f(t) \leq C_1 e^{-\pi\alpha|t|^2}$ and $\mathcal{F}f(\omega) \leq C_2 e^{-\pi|\omega|^2/\alpha}$ for some constants $C_1, C_2, \alpha > 0$. Then there exists a constant $C > 0$ such that $f(t) = C e^{-\pi\alpha|t|^2}$.*

In both cases, the concentration of f and its Fourier transform can be reformulated as the concentration of the elementary tensor $f \otimes \mathcal{F}f \in L^2(\mathbb{R}^{2d})$, so the research naturally progressed to considering joint representations of a function f on the time-frequency plane $\mathbb{R}^{2d}$. For instance, Benedicks's uncertainty principle for $L^2(\mathbb{R}^{2d})$ implies the following uncertainty principle [66, 83, 89, 141].

Theorem 1.6. *Let $f, g \in L^2(\mathbb{R}^d)$. Then the following are equivalent:*

- (i) $|\text{supp } V_g f| < \infty$;
- (ii) $|\text{supp } A(f, g)| < \infty$;
- (iii) $|\text{supp } W(f, g)| < \infty$;
- (iv) $f \equiv 0$ or $g \equiv 0$.

Further uncertainty principles have also been generalized for the short-time Fourier transform [20, 47, 66, 76, 118, 129]. Some progress has also been made for metaplectic time-frequency representations $W_{\mathcal{A}}(f, g)$ [13, 48, 128] with various conditions on the generating symplectic matrix $\mathcal{A} \in \text{Sp}(4d, \mathbb{R})$.

1.8 Universal optimality

Universal optimality is a problem which originates in solid-state physics and is linked to sphere packing problems. The problem of universal optimality is the existence of optimal point configurations which minimize certain potential energies. The Gaussian function $\varphi(t) = 2^{1/4} e^{-\pi|t|^2}$ plays a prominent role in this area, and its striking properties in time-frequency analysis have proven themselves as particularly useful in this context [31, 32].

We first recall a result by Janssen [88] that determines the frame bounds of the Gabor frame $\mathcal{G}(g, \Lambda)$ for lattices $\Lambda = A\mathbb{Z}^2$ of even density. Recall that the matrix $\mathcal{J}$ denotes the standard skew-symmetric matrix introduced in (1.6), and $\sigma(\cdot, \cdot)$ is the associated skew-symmetric form. The *adjoint lattice* Λ° of Λ is given by

$$\Lambda^\circ := \{\lambda^\circ \in \mathbb{R}^2 : \pi(\lambda)\pi(\lambda^\circ) = \pi(\lambda^\circ)\pi(\lambda) \ \forall \lambda \in \Lambda\} = \mathcal{J}A^{-t}\mathbb{Z}^2.$$

Theorem 1.7 (Janssen [88]). *Let $g(t) = 2^{1/4} e^{-\pi t^2}$ denote the Gaussian window and $\Lambda = A\mathbb{Z}^2$, $A \in \text{GL}(2d, \mathbb{R})$, be a lattice of density $|\det A|^{-1} = 2n$, $n \in \mathbb{N}$. Then the associated lower and upper frame bounds of the Gabor frame $\mathcal{G}(g, \Lambda)$ are given by*

$$\begin{aligned} A(\Lambda) &= 2n \min_{z \in \mathbb{R}^2} \sum_{\lambda^\circ \in \Lambda^\circ} e^{-\pi|\lambda^\circ|^2/2} e^{2\pi i \sigma(\lambda^\circ, z)} = \min_{z \in \mathbb{R}^2} \sum_{\lambda \in \Lambda} e^{-2\pi|\lambda+z|^2}, \\ B(\Lambda) &= 2n \max_{z \in \mathbb{R}^2} \sum_{\lambda^\circ \in \Lambda^\circ} e^{-\pi|\lambda^\circ|^2/2} e^{2\pi i \sigma(\lambda^\circ, z)} = \max_{z \in \mathbb{R}^2} \sum_{\lambda \in \Lambda} e^{-2\pi|\lambda+z|^2}. \end{aligned} \quad (1.8)$$

The appearance of the skew-symmetric form above and the adjoint lattice Λ° is expected due to the symplectic nature of the time-frequency plane [53], while the two-dimensional Gaussian function appears as the short-time Fourier transform of the Gaussian window with respect to itself. Poisson's summation formula implies the second series with respect to the lattice Λ [58]. This established the link between the frame bounds of the one-dimensional Gaussian window and the problem of universal optimality in two dimensions described below.

Definition 1.8. *Let $p : \mathbb{R} \rightarrow \mathbb{R}_+$ be a non-negative function (the potential). Let $\Gamma \subseteq \mathbb{R}^d$ be a periodic point configuration. We call the quantity*

$$E_p(\Gamma) = \lim_{R \rightarrow \infty} \frac{1}{\#(\Gamma \cap B_R^d(0))} \sum_{\substack{\gamma, \gamma' \in \Gamma \cap B_R^d(0) \\ \gamma \neq \gamma'}} p(|\gamma - \gamma'|)$$

the energy of Γ with respect to the potential p .

Throughout, we assume that the density

$$\rho(\Gamma) = \lim_{R \rightarrow \infty} \frac{\#(\Gamma \cap B_R^d(0))}{|B_R^d(0)|}.$$

is fixed. Note that any periodic structure with positive density is the finite union of shifts of a lattice, i.e., there exists a lattice $\Lambda = A\mathbb{Z}^d$ and points $x_1, \dots, x_J \in \mathbb{R}^d$ with $x_j - x_k \notin \Lambda$ for $j \neq k$, such that

$$\Gamma = \bigcup_{j=1}^J (x_j + \Lambda).$$

This allows us to rewrite the energy as

$$E_p(\Gamma) = \frac{1}{J} \sum_{j,k=1}^J \sum_{\lambda \in \Lambda \setminus \{x_j - x_k\}} p(|\lambda + x_j - x_k|).$$

Additionally, the density of Γ is given by $\rho(\Gamma) = J \cdot \rho(\Lambda)$.

The main problem is finding a point configuration that minimizes the energy with respect to the potential p among all periodic configurations of a fixed density. The problem of universal optimality centers around finding the optimal configuration for *completely monotone potentials*. We call a potential function $p : \mathbb{R} \rightarrow \mathbb{R}_+$ *completely monotone* if and only if

$$(-1)^k p^{(k)}(r) \geq 0 \quad \text{for all } k \in \mathbb{N}.$$

Completely monotone functions were first introduced by Bernstein in [18]. They appear in several problem areas besides universal optimality and sphere packing, for instance, summation formulas and Hausdorff's moment problem [4, 79, 80].

Definition 1.9. Among all periodic configurations $\Gamma \subset \mathbb{R}^d$ of fixed density δ , we say Γ_0 is universal if it minimizes the potential energy among all configurations of density δ for every completely monotone potential p ;

$$E_p(\Gamma_0) \leq E_p(\Gamma), \quad \begin{array}{l} p \text{ completely monotone,} \\ \Gamma \subseteq \mathbb{R}^d \text{ periodic, } \rho(\Gamma) = \delta. \end{array}$$

Cohn and Kumar also call a universally optimal point configuration a ground state for the completely monotone potentials. While initially it might not be clear why a single point configuration should minimize the energy across a wide family of potentials, the Bernstein-Widder theorem [19, 138] simplifies the matter to the Gaussian potential. Namely, any completely monotone function can also be written as the Laplace transform of a non-negative Borel measure μ .

Theorem 1.10 (Bernstein-Widder theorem). *Let p be completely monotone. Then there exists a non-negative Borel measure μ_p such that*

$$p(r) = \int_0^\infty e^{-\alpha r} d\mu_p(\alpha).$$

Due to this characterization, the problem of determining a universally optimal periodic set Γ_0 is reduced to only determining whether a periodic point configuration minimizes the two-sided exponential potential, or equivalently (after the reparametrization $r \mapsto \sqrt{r}$), whether a periodic point configuration minimizes the Gaussian potential energy.

Lemma 1.11. *Among all periodic configurations $\Gamma \subset \mathbb{R}^d$ of fixed density δ , Γ_0 is universal if it minimizes the potential energy among all configurations of density δ for the Gaussian potential energy $\varphi_\alpha(x) = e^{-\alpha\|x\|_2^2}$ for every $\alpha > 0$;*

$$E_{\varphi_\alpha}(\Gamma_0) \leq E_{\varphi_\alpha}(\Gamma), \quad \Gamma \subseteq \mathbb{R}^d \text{ periodic, } \rho(\Gamma) = \delta.$$

The problem has been solved in dimensions 1 by Cohn and Kumar [31] and by Cohn, Kumar, Miller, Radchenko, and Viazovska in dimensions 8 and 24 [32] by finding a so-called 'magic function' with special interpolating properties.

A celebrated result by Montgomery [108] shows that the hexagonal lattice is a *unique minimizer* of the energy with respect to the Gaussian window for any density *among the class of lattices*. As observed above, the two-dimensional case is naturally close to optimization problems which occur in Gabor analysis. Faulhuber combined Montgomery's result [108] with Jannsen's computations of the Gabor frame bounds for the Gaussian window function [88] and showed that the scaled hexagonal lattice $\frac{1}{\sqrt{2n}}\Lambda_2$, where Λ_2 is given by

$$\Lambda_2 = \frac{\sqrt{2}}{\sqrt[4]{3}} \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} \end{pmatrix} \cdot \mathbb{Z}^2,$$

is the unique minimizer of the upper frame bound $B(\Lambda)$ in (1.8) among all lattices of fixed density $2n$ [52]. Recently, Bétermin, Faulhuber, and Steinerberger [24] showed that the hexagonal lattice of density $2n$ is also the unique maximizer of the lower frame bound $A(\Lambda)$ in (1.8) among all lattices of fixed density $2n$.

It remains a conjecture that the hexagonal lattice is a universally optimal point configuration among all periodic point configurations as well as the optimizer of the Gaussian's frame bounds. However, despite the similarities, the problem of optimizing the Gaussian frame bounds is not equivalent to universal optimality. Universally optimal sets are invariant under scalings, i.e., a set Γ is universally optimal among all periodic sets of density $\rho(\Gamma)$ if and only if $r\Gamma$ is universally optimal among all periodic sets of density $r^{-2}\rho(\Gamma)$. This is due to the fact that the density can be interchanged with the parameter $\alpha > 0$ in Lemma 1.11. However, it is still not obvious whether a universally optimal set exists. Contrary to this, the density is the deciding factor whether or not Λ is a sampling set for the short-time Fourier transform with respect to the Gaussian window [104, 125, 126] and scaling invariance does not hold in this case.

2 Discussion of the main results

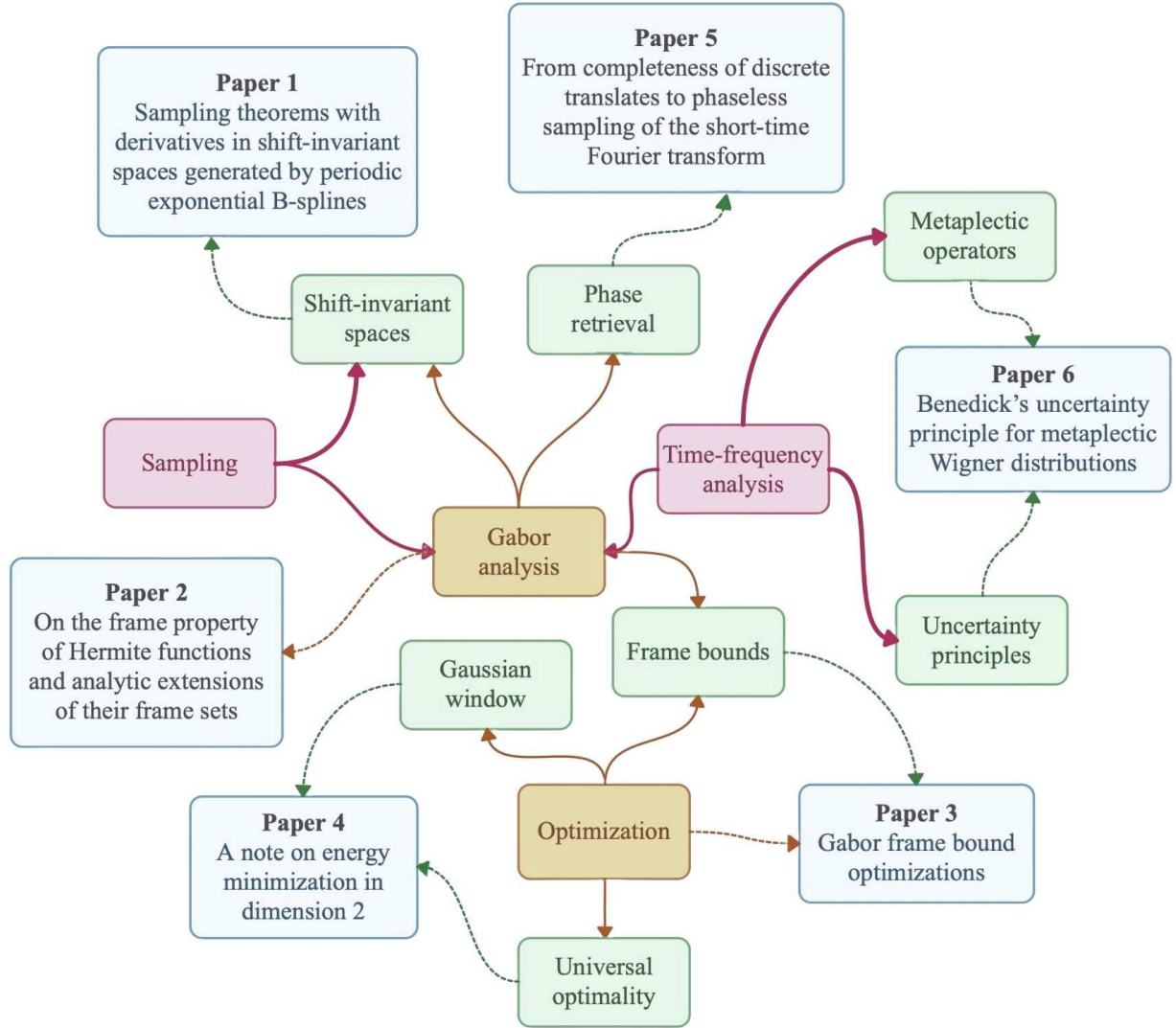


Figure 2: Update of Figure 1 with the papers (blue) included in this thesis.

The results of this thesis resolve problems primarily connected with sampling (Paper 1 and Paper 2), Gabor frame bounds and energy minimization (Paper 3 and Paper 4), STFT phase retrieval (Paper 5) and uncertainty principles for joint time-frequency representations (Paper 6). The main results and proof ideas for each paper are discussed separately below.

The notation is not fully consistent throughout the papers, hence some of the results are not cited verbatim and are stated in the notation established in the introduction.

2.1 Sampling theorems with derivatives in shift-invariant spaces generated by periodic exponential B-splines

The main contribution of Paper 1 is determining a near-optimal result on sampling with derivatives in shift-invariant spaces $V^p(g)$, where the generator g is a *periodic exponential B-spline*, i.e., we investigate the following.

Under which conditions on a separated set $X \subseteq \mathbb{R}$ and bounded multiplicity function $\mu_X : X \rightarrow \{0, \dots, m-1\}$ can we stably recover $f \in V^p(g)$ from the data $f^{(s)}(x)$, $0 \leq s \leq \mu(x)$, $x \in X$?

We call a pair (X, μ_X) of a discrete set $X \subseteq \mathbb{R}$ and a bounded multiplicity function $\mu_X : X \rightarrow \{0, \dots, S\}$ a sampling pair for $V^p(g)$ if there exist constants $0 < A_p \leq B_p < \infty$ such that

$$A_p \|f\|_p^p \leq \sum_{x \in X} \sum_{s=0}^{\mu_X(x)} |f^{(s)}(x)|^p \leq B_p \|f\|_p^p, \quad f \in V^p(g),$$

with the usual modification for $p = \infty$. A pair (X, μ_X) is a uniqueness pair for $V^p(g)$ if any function $f \in V^p(g)$ is uniquely determined by the data $f^{(s)}(x)$, $0 \leq s \leq \mu_X(x)$, $x \in X$. The classical sampling is the case $\mu_X \equiv 0$.

Sampling the derivatives in addition to the standard samples allows for better capturing of local trends and convexity properties as well as sparser sampling schemes since larger gaps between sampling locations can be compensated with samples of higher derivatives. This sampling problem has been studied for the Paley-Wiener space [2, 57, 116, 117], spaces generated by totally positive functions of Gaussian type [63], and spaces generated by PEB-splines [127]. In contrast to the complex analysis setup in [63, 116, 117] and the available results on zero sets, Paper 1 relies instead on spline theory to recover the functions $f \in V^p(g)$.

Periodic exponential B-splines (PEB-splines) are a generalization of the uniform B-splines constructed on integer knots. A PEB-spline of order $m \in \mathbb{N}$ is a C^{m-2} -differentiable, piecewise C^{m-1} -differentiable function supported on $[k, k+m]$ for some integer $k \in \mathbb{Z}$. The simplest non-trivial example is the *exponential* B-spline

$$\mathcal{E}_{m,\alpha}(x) := \prod_{s=1}^m {}^* e^{\alpha_s x} \chi_{[0,1)}(x), \quad \alpha \in \mathbb{R}^m,$$

where $\prod^*$ denotes the convolution product. The exact definition of PEB-splines is quite involved and we omit it here. It can be found in [123] as well as in Paper 1.

The main property that all PEB-splines share is the *Schoenberg-Whitney conditions*, also known as the interlacing property of PEB-splines [123].

Theorem 2.1. *Let g be a PEB-spline of order $m \in \mathbb{N}$ supported on $[0, m]$, and let $t_1 \leq \dots \leq t_D$ be sampling locations. We define the indices of repetition as*

$$d_i := \max \{\ell : t_i = \dots = t_{i-\ell}\} \leq m-1, \quad 1 \leq i \leq D.$$

Then the determinant of the collocation matrix

$$M \begin{pmatrix} t_1, \dots, t_D \\ g, \dots, T_{D-1}g \end{pmatrix} := \left(T_{\ell-1}g^{(d_i)}(t_i) \right)_{1 \leq i, \ell \leq D}$$

is nonnegative. It is positive if and only if

$$t_i \in \begin{cases} (i-1, i+m-1), & d_i < m-1 \\ [i-1, i+m-1), & d_i = m-1, \end{cases} \quad 1 \leq i \leq D. \quad (2.1)$$

Theorem 2.1 gives a characterization when the Hermite interpolation problem

$$f = \sum_{\ell=1}^D c_\ell T_{\ell-1}g, \quad L_{d_i}f(t_i) = \zeta_i, \quad 1 \leq i \leq D$$

has a unique solution and will play a key role in the results of Paper 1. Note that the condition (2.1) is a minimal requirement since g vanishes outside $(0, m)$. The stunning property of PEB-splines is the fact that the necessary condition is also sufficient for the invertibility of the collocation matrix. We remark that the Schoenberg-Whitney conditions are usually stated with a different set of differential operators. However, they can be replaced with the classical derivatives as stated in Theorem 2.1. The proof is included in Paper 1.

We turn back to the shift-invariant space $V^p(g)$. A necessary density condition for a pair (X, μ_X) to be a sampling pair for $V^p(g)$ is given by the weighted lower Beurling density defined as

$$D^-(X, \mu_X) := \liminf_{R \rightarrow \infty} \inf_{y \in \mathbb{R}} \frac{1}{2R} \sum_{x \in X \cap [y-R, y+R]} (1 + \mu_X(x)).$$

If the multiplicity is constant with $\mu_X \equiv S-1$, then the weighted Beurling density is a multiple of the classical Beurling density: $D^-(X, \mu_X) = S \cdot D^-(X)$.

For a large class of generators, a sampling pair has to satisfy $D^-(x, \mu_X) \geq 1$ [63]. One of the primary contributions of Paper 1 is the existence of sampling pairs (X, μ_X) with weighted Beurling density $D^-(X, \mu_X) < 1 + \varepsilon$ for all $\varepsilon > 0$.

It is evident that there cannot be a critical density with respect to the weighted lower Beurling density for this generator. To see this, let $\alpha \in (0, 1)$. Then the set $X = \alpha\mathbb{Z} \setminus [0, m]$ with any multiplicity function μ_X has density $D^-(X, \mu_X) \geq \alpha^{-1}$ but the function $g \in V^p(g)$ vanishes on X .

The two main results of Paper 1, Theorem 2.2 and 2.4, describe two different sets of sufficient conditions for a pair (X, μ_X) to be sampling pair for $V^p(g)$.

Theorem 2.2. *Let g be a PEB-spline of order $m \in \mathbb{N}$ supported on $[0, m]$. Assume $X \subseteq \mathbb{R}$ with multiplicity function μ_X satisfies the following:*

- (i) X is separated.

(ii) $S := \max_{x \in X} \mu_X(x) \leq m - 1$.

(iii) There exist integers $M, L > 0$ and $\varepsilon > 0$, such that for every $k \in \mathbb{Z}$, there exist points $x_1^k < x_2^k < \dots < x_{N(k)}^k$ in $X \cap [M + kL, M + kL + L)$ and nonnegative integers $\mu_1^k, \mu_2^k, \dots, \mu_{N(k)}^k$ with the following three properties:

$$\mu_j^k \leq \mu_X(x_j^k) \quad \text{for all } k \in \mathbb{Z}, 1 \leq j \leq N(k),$$

$$\sum_{j=1}^{N(k)} (1 + \mu_j^k) = L + m - 1 \quad \text{for all } k \in \mathbb{Z},$$

$$x_j^k \in X \cap \left([\mu_j^k - m + \varepsilon, -\varepsilon] + M + kL + 1 + \sum_{n=1}^{j-1} (1 + \mu_n^k) \right).$$

Then (X, μ_X) is a sampling set for $V^p(g)$ for all $p \in [1, \infty]$.

The proof is inspired by [9]. The idea is to partition $\mathbb{R}$ in intervals of length L and recover the restriction of f on $[M + kL, M + kL + L)$. The support condition of φ implies that only $L + m - 1$ shifts of g do not vanish on this interval, so by the Schoenberg-Whitney conditions this is a sufficient number of samples. We consider only a subsequence of the sampled data to be able to directly apply the Schoenberg-Whitney conditions, but one can replace this by considering rectangular collocation matrices with full range in the formulation of Theorem 2.1.

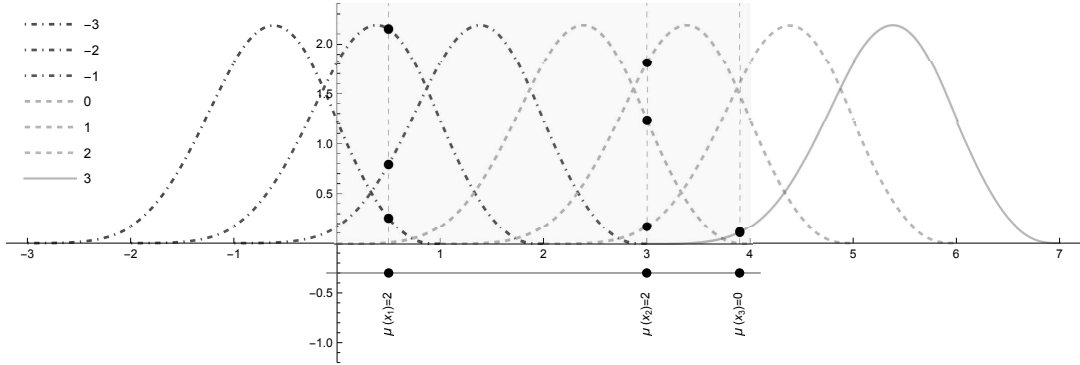


Figure 3: Nonvanishing shifts of $g(x) = \prod_{j=1}^4 e^x \chi_{[0,1)}(x)$. The sampling points are $x_1 = 0.5$, $x_2 = 3$, $x_3 = 3.9$, with multiplicities $\mu_X(x_1) = \mu_X(x_2) = 2$, $\mu_X(x_3) = 0$. The first sampling point lies in the support of the first three shifts of g (dot-dashed), the second point is in the support of the next three shifts of g (dashed), and the last point - in the support of the last shift of g (solid).

The shift-invariance of the space and the commutation of the derivatives with the time shifts allowed us to move the recovery of f on $[M + kL, M + kL + L)$ to a recovery of $f_k = T_{-M-kL}f$ on $[0, L)$ from the sampling locations $z_j^k = x_j^k - (M + kL)$,

$1 \leq j \leq N(k)$, $k \in \mathbb{Z}$. The ε -condition is slightly more stringent than what the Schoenberg-Whitney conditions require, but it is necessary to obtain the uniform bounds on the spectrum of the collocation matrices for $(f_k)_{k \in \mathbb{Z}}$, cf. Theorem 2.1, via a compactness argument. The conditions of Theorem 2.2 are optimal in the following sense: for any $\alpha, \beta \in \mathbb{Q}$, $\alpha < 1$ and $\beta < \frac{\alpha}{4}$, the set $X = \beta + \alpha(S+1)\mathbb{Z}$ with $\mu_X \equiv S$ satisfies the conditions of Theorem 2.2 and has weighted lower Beurling density $\mathcal{D}^-(x, \mu_X) = \alpha^{-1}$. Furthermore, if $\alpha \geq 1$, then $\alpha\mathbb{Z}$ is not a sampling set.

To formulate the second sampling theorem, we introduce a new density measure, the weighted maximum gap, which we found more suitable for PEB-spline generators. The *weighted maximum gap* $\mathbf{mg}(X, \mu_X)$ of a sampling set X with a multiplicity function $\mu_X : X \rightarrow \{0, \dots, m-1\}$ is defined as

$$\begin{aligned} \mathbf{mg}(X, \mu_X) &:= \max \left\{ \sup_{j \in \mathbb{Z}} \frac{x_{j+1} - x_j}{1 + \mu_X(x_j)}, \sup_{j \in \mathbb{Z}} \frac{x_{j+1} - x_j}{1 + \mu_X(x_{j+1})} \right\} \\ &= \sup_{j \in \mathbb{Z}} \frac{x_{j+1} - x_j}{1 + \min\{\mu_X(x_j), \mu_X(x_{j+1})\}}. \end{aligned}$$

The inverse of the weighted maximum gap defines a density comparable to the weighted lower Beurling density:

$$\mathbf{mg}(X, \mu_X)^{-1} \leq D^-(X, \mu_X).$$

A stepping stone to the second sampling result (Theorem 2.4) is the following sufficient condition for uniqueness pairs.

Theorem 2.3. *Let g be a PEB-spline of order $m \in \mathbb{N}$ and $X \subseteq \mathbb{R}$ be a separated set with a multiplicity function $\mu_X : X \rightarrow \{0, \dots, m-1\}$. If*

$$\mathbf{mg}(X, \mu_X) = \max \left\{ \sup_{j \in \mathbb{Z}} \frac{x_{j+1} - x_j}{1 + \mu_X(x_j)}, \sup_{j \in \mathbb{Z}} \frac{x_{j+1} - x_j}{1 + \mu_X(x_{j+1})} \right\} < 1,$$

then (X, μ) is a uniqueness pair for $V^p(g)$ for all $1 \leq p \leq \infty$.

The proof once again aims for a local recovery on an interval $[a, a+L]$ with integer ends $a, a+L \in \mathbb{Z}$ using the Schoenberg-Whitney conditions. In this case, though, instead of a partition, the proof creates a covering of overlapping intervals, of not necessarily the same length, which ensures that any function $f \in V^p(g)$ is recovered from the samples. As argued in the previous theorem, the restriction to $[a, a+L)$ involves $L+m-1$ shifts of g . However, a minor, but tangible difference here is the fact that the recovery intervals are *closed*, and due to the discontinuity of the last derivative $g^{(m-1)}$, in which case we only sample the right-sided derivatives, we have to take into account $L+m$ shifts. The maximum gap ensures that, provided the length L is sufficiently large, there are at least $L+m-1$ samples, and they satisfy the Schoenberg-Whitney conditions in Theorem 2.1. Using combinatorial arguments and the fact that the maximum gap is strictly less than 1, we could make "corrections" in case the last sample was a sample at a discontinuity point and more samples were required on the interval.

Theorem 2.4 (Maximum Gap Theorem). *Let g be a PEB-spline of order $m \geq 2$ and $X \subseteq \mathbb{R}$ be a separated set with multiplicity function $\mu_X : X \rightarrow \{0, \dots, m-1\}$. If the multiplicity function satisfies*

$$\text{dist}(\{x \in X : \mu_X(x) = m-1\}, \mathbb{Z}) > 0 \quad (2.2)$$

and the weighted maximum gap of (X, μ_X) satisfies

$$\mathbf{mg}(X, \mu_X) < 1, \quad (2.3)$$

then (X, μ_X) is a sampling set for $V^p(g)$, $1 \leq p \leq \infty$.

A necessary ingredient of the proof of Theorem 2.4 are *weak limits of integer translates* of the pair (X, μ_X) , see Paper 1 for the exact definition. A result of Gröchenig et al. [63, Thm. 2.1] states that a pair (X, μ_X) is a sampling pair if and only if any of its weak limits of integer translates (Y, μ_Y) is a uniqueness pair. Since the weighted maximum gap of a weak limit of translates (Y, μ_Y) is not larger than $\mathbf{mg}(X, \mu_X)$, the claim follows from Theorem 2.3. The uniform distance condition (2.2) is due to the discontinuity of the last derivative, and is needed for the adaptation of [63, Thm. 2.1] for piecewise continuous functions. If the multiplicity is bounded by $\mu_X \leq m-2$, then the distance condition (2.2) is void. In particular, for any multiplicity $\mu_X \equiv S \leq m-2$, any lattice $\alpha\mathbb{Z}$, $\alpha < S$, is a sampling set with

$$D^-(X, \mu_X) = \mathbf{mg}(X, \mu_X)^{-1} = \frac{S}{\alpha},$$

thus the result is near-optimal. The maximum gap condition (2.3) is critical since $\mathbb{Z}$ is not a sampling set for $V^p(g)$ for any $1 \leq p \leq \infty$.

In the unweighted case, Theorem 2.2 implies Theorem 2.4 [9]. However, when sampling (non-trivially) with derivatives, neither of the theorems above implies the other, and examples are included in Paper 1.

2.2 On the frame property of Hermite functions and analytic extensions of their frame sets

Despite the fact that the existence of Gabor frames is the cornerstone of Gabor analysis, there are so far a limited number of Gabor systems $\mathcal{G}(g, \Lambda)$ which are known to be frames. Even in the simplest case of $d = 1$ and rectangular lattices $a\mathbb{Z} \times b\mathbb{Z}$, there is only a handful of types of window functions $g \in L^2(\mathbb{R}^d)$ where we know whether $\mathcal{G}(g, a\mathbb{Z} \times b\mathbb{Z})$ is a Gabor frame, such as the Gaussian window [104, 125, 126], the hyperbolic secant [88, 90], rational functions [15, 99] and totally positive functions [60, 65, 75, 95].

This paper focuses on the Gabor systems of the Hermite function $\mathfrak{h}_n$ of order n , scaled to unit norm [53];

$$\mathfrak{h}_n(t) = (-1)^n \frac{2^{1/4-n/2}}{\sqrt{n!}(2\pi)^n} e^{\pi t^2} \frac{d^n}{dt^n} e^{-2\pi t^2}, \quad t \in \mathbb{R}, \quad n \in \mathbb{N}_0,$$

It is already known that the Gabor system $\mathcal{G}(\mathfrak{h}_n, a\mathbb{Z} \times b\mathbb{Z})$ is a frame if the lattice has density $(ab)^{-1} > n + 1$, $n \in \mathbb{N}$ [62]. Numerical results indicate that this is quite far from the full set of parameters (a, b) such that $\mathcal{G}(\mathfrak{h}_n, a\mathbb{Z} \times b\mathbb{Z})$ is a frame [56]. After observing that the Gabor system $\mathcal{G}(\mathfrak{h}_n, \frac{1}{\sqrt{n+1}}\mathbb{Z}^2)$ is not a frame for the first four Hermite functions $\mathfrak{h}_n$, $n \in \{0, 1, 2, 3\}$, [58, 101, 105], Lemvig [101] posed the question of whether $(ab)^{-1} > n + 1$ is the optimal necessary density condition.

Let $n \in \mathbb{N}$. Is there an $\varepsilon > 0$ such that $\mathcal{G}(g, a\mathbb{Z} \times b\mathbb{Z})$ is a Gabor frame for all $a, b > 0$ that satisfy $(ab)^{-1} \geq n + 1 - \varepsilon$?

For the first four Hermite functions the answer to Lemvig's question is negative due to the aforementioned Gabor systems on square lattices. Motivated by these examples, Paper Paper 2 is dedicated to the Gabor systems $\mathcal{G}(\mathfrak{h}_n, \frac{1}{\sqrt{n+1}}\mathbb{Z}^2)$ for $n \geq 4$. The main result is the peculiar fact that the pattern breaks at $n \geq 4$.

Theorem 2.5. *The Gabor system $\mathcal{G}(\mathfrak{h}_n, \frac{1}{\sqrt{n+1}}\mathbb{Z}^2)$ is a frame if and only if $n \geq 4$.*

The proof uses the *Janssen test* [140]. It is a *sufficient* criterion for a Gabor system on a lattice Λ to be a frame. Ideawise, the analogue was first shown for *finite* Gabor frames in [134], and it relies on the Janssen representation [58, 87] of the frame operator of $\mathcal{G}(g, \Lambda)$ as an expansion over the *adjoint lattice* Λ° .

Lemma 2.6 (Janssen test). *Let $\Lambda \subseteq \mathbb{R}^2$ be a lattice and $g \in L^2(\mathbb{R}^d)$ a normalized window function. If*

$$\sum_{\lambda^\circ \in \Lambda^\circ} |V_g g(\lambda^\circ)| < 2,$$

then $\mathcal{G}(g, \Lambda)$ constitutes a Gabor frame for $L^2(\mathbb{R})$.

To apply this to the Hermite functions, we introduce the Laguerre polynomials

$$\mathcal{L}_n(x) = \sum_{k=0}^n \binom{n}{k} \frac{(-x)^k}{k!}, \quad n \in \mathbb{N}_0.$$

The Laguerre polynomials are orthogonal on $\mathbb{R}_+$ with respect to the exponential weight $w(x) = e^{-x}$, and for all $n \in \mathbb{N}$, $\mathcal{L}_n(x)$ has n distinct roots in $\mathbb{R}_+$ [131]. For all $n \in \mathbb{N}$, the short-time Fourier transform $V_{\mathfrak{h}_n} \mathfrak{h}_n$ can be expressed as [53]

$$V_{\mathfrak{h}_n} \mathfrak{h}_n(x, \omega) = e^{-\pi i x \omega} \mathcal{L}_n \left(\pi(x^2 + \omega^2) \right) e^{-\frac{\pi}{2}(x^2 + \omega^2)}.$$

To prove Theorem 2.5, we split the series in Lemma 2.6 in a sum over small lattice elements, which have a large contribution to the value of the series, and the rest. We combine this with delicate estimates on Laguerre polynomials [97, 98, 131] to prove that $\mathcal{G}(\mathfrak{h}_n, \frac{1}{\sqrt{n+1}}\mathbb{Z}^2)$ passes the Janssen test for all $n \geq 4$.

2.3 Gabor frame bound optimizations

As noted in Subsection 1.3, the condition number $\kappa(\Lambda) = \frac{B(\Lambda)}{A(\Lambda)}$ of a Gabor frame $\mathcal{G}(g, \Lambda)$ is crucial for the convergence rate of the frame reconstruction algorithm [50, 143] and motivated us to study the following problem.

Let $g \in L^2(\mathbb{R})$ be a window function and Ξ a set of lattices Λ such that $\mathcal{G}(g, \Lambda)$ is a Gabor frame of a fixed density δ . Which lattice in Ξ has the best condition number?

The problem is rather difficult to answer since there are very few window functions g for which we know whether $\mathcal{G}(g, \Lambda)$ is a frame, even if Λ is rectangular, so we can only optimize over the known Gabor frames. Furthermore, in most cases, the frame property was proven without exact quantitative estimates. So far, the problem was tractable for a handful of windows g and rectangular lattices $\Lambda = a\mathbb{Z} \times b\mathbb{Z}$ with integer density $(ab)^{-1} = n$ [88].

The problem has only recently been solved for the Gaussian window on lattices of even density by Bétermin, Faulhuber, and Steinerberger [24], in which case the hexagonal lattice is the unique minimizer of the condition number among all lattices of fixed density. Montgomery [108] had shown that the upper frame bound is minimized by the hexagonal lattice, and the authors of [24] showed that the hexagonal lattice also maximizes the upper frame bound. This splitting approach led us to investigate three optimization problems:

$$\arg \max_{(ab)^{-1}=n} A(a\mathbb{Z} \times b\mathbb{Z}), \quad \arg \min_{(ab)^{-1}=n} B(a\mathbb{Z} \times b\mathbb{Z}), \quad \arg \min_{(ab)^{-1}=n} \kappa(a\mathbb{Z} \times b\mathbb{Z}).$$

We optimized the condition number and the frame bounds of well-known windows. The frame bounds were already computed in [88]. In most of the investigated cases, there exist unique optimal rectangular lattices, and this is proven by

1. studying the asymptotic behaviour of the parameter a as it tends to 0 and ∞ , and showing that the condition number and the bounds monotonically deteriorate, thus excluding the degenerate configurations as optimal.
2. showing that there exists a unique critical point of the object function (the condition number or the bounds), which has to be the optimal solution.

Note that in the setting with only one critical point, if the maximizer of the lower frame bound $A = A(a)$ does not coincide with the minimizer of the upper frame bound $B = B(a)$, then the minimizer of the condition number $\kappa = \kappa(a)$ is strictly between the frame bound optimizers: ¹

$$\kappa'(a) = 0 \quad \Longleftrightarrow \quad \frac{A'(a)}{A(a)} = \frac{B'(a)}{B(a)}.$$

¹With a slight abuse of notation, we use $A, A(a), A(a, b)$ (and analogously for B and κ) interchangeably, depending on the context and if clarification is necessary.

If A' and B' do not vanish at the same point, then a critical point of κ has to be a point where A' and B' have the same sign, i.e., the interval between the critical points of A and B .

We proceed with the main results of Paper 3 in the following three theorems.

Theorem 2.7. *Let $g(t) = \left(\frac{\pi}{2}\right)^{1/2} \text{sech}(\pi t)$ be the normalized hyperbolic secant and consider the rectangular lattice $a\mathbb{Z} \times b\mathbb{Z}$ of integer density $n \in \mathbb{N}$, i.e. $(ab)^{-1} = n$, with $n \geq 2$. By $A(a, b)$ and $B(a, b)$ we denote the sharp lower and upper frame bound of the Gabor system $\mathcal{G}(g, a\mathbb{Z} \times b\mathbb{Z})$, respectively. Then, we have*

$$A\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right) \geq A(a, b) \quad \text{and} \quad B\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right) \leq B(a, b),$$

with equality if and only if $a = b = \frac{1}{\sqrt{n}}$, i.e., if the lattice is the square lattice of density n .

Contrary to the hyperbolic secant, the one-sided and the two-sided exponential do not have a simultaneously optimal lattice.

Theorem 2.8. *Let $g(t) = \sqrt{2} e^{-t} \chi_{\mathbb{R}_+}(t)$ and consider the Gabor system $\mathcal{G}(g, a\mathbb{Z} \times b\mathbb{Z})$ with $(ab)^{-1} = n \in \mathbb{N}$. For fixed n , the frame condition number is minimized by the lattice $\frac{\text{arcsinh}(n)}{n} \mathbb{Z} \times \frac{1}{\text{arcsinh}(n)} \mathbb{Z}$, and this lattice is the unique minimizer. The lower and the upper frame bound have a unique optimal lattice $\frac{a_+}{n} \mathbb{Z} \times \frac{1}{a_+} \mathbb{Z}$ and $\frac{a_-}{n} \mathbb{Z} \times \frac{1}{a_-} \mathbb{Z}$ respectively. Furthermore,*

$$a_+ < \text{arcsinh}(n) < a_-.$$

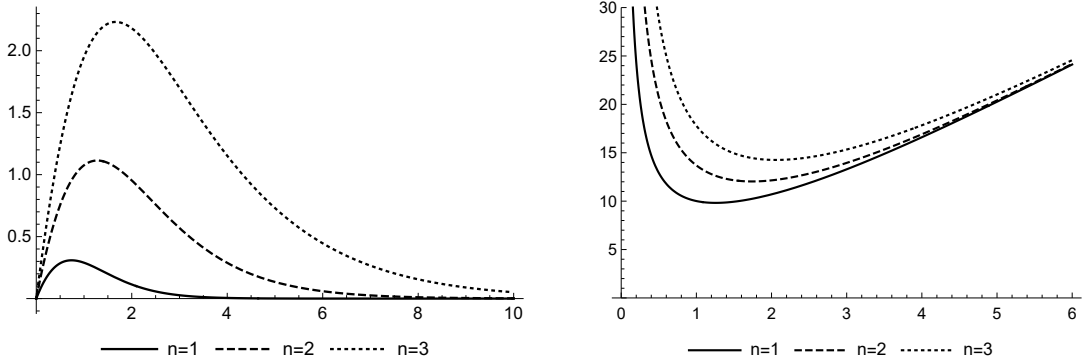


Figure 4: Lower (left plot) and upper (right plot) frame bounds of the Gabor frame of the one-sided exponential $\mathcal{G}(g, a\mathbb{Z} \times \frac{1}{an}\mathbb{Z})$ for density $n \in \{1, 2, 3\}$.

Theorem 2.9. *Let $g(t) = e^{-|t|}$ denote the two-sided exponential and consider the Gabor system $\mathcal{G}(g, a\mathbb{Z} \times b\mathbb{Z})$ with $(ab)^{-1} = n \in \mathbb{N}$, $n \geq 2$. For fixed n , the frame condition number, the lower and the upper frame bound have unique optimal lattices $\frac{a_0}{n} \mathbb{Z} \times \frac{1}{a_0} \mathbb{Z}$, $\frac{a_+}{n} \mathbb{Z} \times \frac{1}{a_+} \mathbb{Z}$, and $\frac{a_-}{n} \mathbb{Z} \times \frac{1}{a_-} \mathbb{Z}$, respectively. Furthermore,*

$$\text{arccosh}(n) < a^- < a_0 < a_+ < 4 \text{arccosh}(n).$$

The frame bounds in Theorems 2.7, 2.8, and 2.9 are analytic functions, and the proofs rely heavily on the first few derivatives of the frame bounds. We combined trigonometric identities, monotonicity arguments, and strict (log-) convexity properties of the frame bounds around the critical points with interval estimates on small scales to prove the uniqueness and the positioning of the optimizers.

We recall that the Gabor systems $\mathcal{G}(g, \Lambda)$ are invariant under metaplectic transformations $\mu(\mathcal{A}) \in \text{Mp}(2d, \mathbb{R})$ in the following manner [58]:

$$\begin{aligned} \mathcal{G}(g, \Lambda) \text{ is a frame with frame bounds } 0 < A \leq B < \infty \text{ if and only if} \\ \mathcal{G}(\mu(\mathcal{A})g, \mathcal{A}\Lambda) \text{ is a frame with the same bounds.} \end{aligned}$$

In particular, if we pick $\mathcal{A} = \mathcal{J}$ to be the standard skew-symmetric matrix,

$$\begin{aligned} \mathcal{G}(g, a\mathbb{Z} \times b\mathbb{Z}) \text{ is a frame with frame bounds } A, B \text{ if and only if} \\ \mathcal{G}(\hat{g}, b\mathbb{Z} \times a\mathbb{Z}) \text{ is a frame with the same bounds.} \end{aligned}$$

The hyperbolic secant is an eigenfunction of the Fourier transform, so a symmetry in the time-frequency coordinates is expected. In stark contrast, the Fourier transforms of the one-sided and the two-sided exponential are rational functions. In particular, their Fourier transforms have a significantly milder decay. This is starkly reflected in the optimal sampling lattices. The optimal lattices in Theorem 2.8 and Theorem 2.9 are significantly denser in the time direction than in the frequency direction.

Moreover, any of the Theorems 2.7, 2.8, 2.9 can be reformulated for dilated windows, e.g., for $g_\gamma(t) := \sqrt{2\gamma} e^{-\gamma t} \chi_{\mathbb{R}_+}(t)$, the optimal lattices of density $n \in \mathbb{N}$ for the condition number, the upper, and the lower frame bound are

$$\begin{aligned} \frac{\text{arcsinh}(n)}{\gamma n} \mathbb{Z} \times \frac{\gamma}{\text{arcsinh}(n)} \mathbb{Z}, \\ \frac{a^-(n)}{\gamma n} \mathbb{Z} \times \frac{\gamma}{a^-(n)} \mathbb{Z}, \\ \frac{a_+(n)}{\gamma n} \mathbb{Z} \times \frac{\gamma}{a_+(n)} \mathbb{Z}, \end{aligned}$$

respectively, since the dilation operator projects to $\mathcal{A} = \begin{pmatrix} \gamma^{-1} & 0 \\ 0 & \gamma \end{pmatrix} \in \text{Sp}(2, \mathbb{R})$.

2.4 A note on energy minimization in dimension 2

The hexagonal lattice Λ_2 given by

$$\Lambda_2 = \frac{\sqrt{2}}{\sqrt[4]{3}} \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} \end{pmatrix} \cdot \mathbb{Z}^2,$$

is the *unique minimizer* of the Gaussian energy with respect to the Gaussian window among the class of lattices of density 1 [108]. Thus Lemma 1.11 implies that the hexagonal lattice $\frac{1}{\sqrt{\delta}} \Lambda_2$ of density δ is universally optimal within the *the class of lattices* of density δ for all $\delta > 0$. The leading motivation for Paper 4 is the following popular speculation.

The hexagonal lattice is conjectured to be the universally optimal periodic point configuration in $\mathbb{R}^2$.

Paper 4 considers several new classes of periodic point configurations and confirms that the hexagonal lattice Λ_2 has energy strictly lower than any of the sets investigated in Paper 4. Recently, non-lattice periodic configurations have also been studied by Hardin and Tempas [77, 132].

We single out the honeycomb structure $\Gamma_\circ$, which is the natural non-lattice contender of the hexagonal lattice Λ_2 for the property of universal optimality. The honeycomb structure of density 1 can be written as the union of the hexagonal lattice of density $\frac{1}{2}$ and a translation of the hexagonal lattice of density $\frac{1}{2}$:

$$\Gamma_\circ = \frac{1}{\sqrt{2}}\Lambda_2 \cup \frac{1}{\sqrt{2}}(\Lambda_2 + x_0), \quad x_0 = \frac{1}{3}\sqrt{\frac{2}{\sqrt{3}}}\begin{pmatrix} \frac{3}{2} \\ \frac{\sqrt{3}}{2} \end{pmatrix}.$$

By rewriting the energy of $\Gamma_\circ$ and Λ_2 via the cubic theta functions introduced by Borwein and Borwein [22], we use their arithmetic-geometric mean properties and Poisson's summation formula to rewrite the difference of their energies $E_{\phi_\alpha}(\Lambda_2) - E_{\phi_\alpha}(\Gamma_\circ)$ as a series of strictly negative terms, and obtain the following.

Theorem 2.10. *For any fixed $\alpha > 0$ and fixed density $\delta > 0$ the hexagonal lattice $\frac{1}{\sqrt{\delta}}\Lambda_2$ has a smaller energy than the honeycomb structure $\frac{1}{\sqrt{\delta}}\Gamma_\circ$;*

$$E_{\phi_\alpha}(\frac{1}{\sqrt{\delta}}\Lambda_2) < E_{\phi_\alpha}(\frac{1}{\sqrt{\delta}}\Gamma_\circ).$$

We further show that among all unions

$$\frac{1}{\sqrt{2\delta}}\Lambda_2 \cup \frac{1}{\sqrt{2\delta}}(\Lambda_2 + x), \quad x \in \mathbb{R}^2, x \notin \frac{1}{\sqrt{2\delta}}\Lambda_2, \quad (2.4)$$

of 2 translations of the hexagonal lattice of density δ , the honeycomb has the minimal energy, thus by Theorem 2.10, none of the point configurations of the form (2.4) is universally optimal.

We summarize the rest of the considered sets in the following theorem.

Theorem 2.11. *In dimension 2, consider the set of lattices, the set of Cartesian tensor configurations and the set of unions of two shifted rectangular or hexagonal lattices. We denote these sets, respectively, by*

$$\begin{aligned} \Lambda &= \{\Lambda \subset \mathbb{R}^2 : \Lambda \text{ a lattice}\}, \\ \Gamma_\times &= \{\Gamma = \Gamma_1 \times \Gamma_2 : \Gamma_1, \Gamma_2 \text{ a periodic configuration in } \mathbb{R}\} \\ \Gamma_x &= \{\Gamma \subset \mathbb{R}^2 : \Gamma = \Lambda \cup (\Lambda + x), \Lambda \in \{(a\mathbb{Z} \times a^{-1}\mathbb{Z}), \Lambda_2\}, a > 0, x \in \mathbb{R}^2\}, \end{aligned}$$

For any fixed $\alpha > 0$ and for any fixed density $\delta > 0$, the hexagonal lattice $\frac{1}{\sqrt{\delta}}\Lambda_2$ is the unique minimizer among periodic configurations from the set $\Gamma_\times \cup \Gamma_x$;

$$E_{\phi_\alpha}(\frac{1}{\sqrt{\delta}}\Lambda_2) \leq E_{\phi_\alpha}(\Gamma), \quad \Gamma \in \Gamma_\times \cup \Gamma_x, \rho(\Gamma) = \delta.$$

In other words, the hexagonal lattice Λ_2 is universally optimal in $\Lambda \cup \Gamma_\times \cup \Gamma_x$.

The comparison within the set of lattices Λ was already established in [24]. Among the set of Cartesian products of periodic point configurations, we show that the optimal one is a square lattice by rewriting the energy of Cartesian products $\Gamma_1 \times \Gamma_2$ as the product of the respective energies. The solution follows from a result of Cohn and Kumar [31] which states that the optimal periodic point configurations in $\mathbb{R}$ are equispaced. Among all rectangular lattices of fixed density, the square lattice has the minimal energy, which is still larger than the energy of the hexagonal lattice of the same density.

We discussed the unions of two translations of the hexagonal lattice above. Similarly, we show that among all unions of *square lattices* of fixed density, any optimal solution is a lattice, thus covered by the comparison of Λ_2 with Λ in [24].

2.5 From completeness of discrete translates to phaseless sampling of the short-time Fourier transform

As discussed in Subsection 1.5, the phaseless sampling of the short-time Fourier transform on a lattice is not one-to-one on $L^2(\mathbb{R}^d)$ [68], and more apriori assumptions on the function class are necessary when sampling the phaseless short-time Fourier transform on lattices. The main objective of Paper 5 is to answer the following.

Can we recover compactly supported functions $f \in L^2(\mathbb{R}^d)$ from phaseless STFT samples $|V_g f(x)|$ for a suitable choice of a window $g \in L^2(\mathbb{R}^d)$ and a separated set $X \subseteq \mathbb{R}^{2d}$? Can X be a lattice?

The problem has already been studied in the one-dimensional setting when the window is the Gaussian window [136] via the connection to the Fock space and for $f \in L^4([-\sigma, \sigma])$ [69] with methods from sampling in shift-invariant spaces. We investigate the sampling of the phaseless short-time Fourier transform with respect to a large class of continuous windows $g \in C(K)$ which contains the Gaussian window and extend the result in [69] in several directions.

We solve the problem under the assumption that $\text{supp} f \subseteq K$ for a fixed compact set $K \subseteq \mathbb{R}^d$. The sets X we propose are Cartesian products $\Lambda \times \Gamma$, where Λ, Γ depend on K . In particular, X can be chosen to be a lattice. Under additional assumptions on f , we also provide an example of separated sets for Λ, Γ which are independent of K , thus X is independent of K . One of the main insights of Paper 5 is establishing a connection between the phase retrieval problem and the problem of *completeness of discrete translates*.

The completeness of translates problem is a well-studied problem in approximation theory and poses the following question:

Let $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ be a Banach space of functions on $\mathbb{R}^d$, $g : \mathbb{R}^d \rightarrow \mathbb{C}$ a function on $\mathbb{R}^d$ and $\Lambda \subseteq \mathbb{R}^d$ a non-trivial subset of $\mathbb{R}^d$. When is the system of translates

$$\mathfrak{T}(g, \Lambda) = \{T_{\lambda} g : \lambda \in \Lambda\}$$

complete in $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$?

There has been a vast interest in the completeness of such systems in various spaces, starting from [139], to more recent results [12, 23, 102, 103, 110, 111, 112]. The particular case we are interested in is completeness in the space of continuous functions $C(K)$ on compact sets $K \subseteq \mathbb{R}^d$, see, for instance, [3, 100, 114, 144].

To formulate the results, we use the notation $g_\omega = (T_\omega g)\bar{g}$, $\omega \in \mathbb{R}^d$. Additionally, $A \pm B$ denotes the set sum and set difference of $A, B \subseteq \mathbb{R}^d$ respectively, i.e.,

$$A + B = \{a + b : a \in A, b \in B\}, \quad A - B = \{a - b : a \in A, b \in B\}.$$

Theorem 2.12. *Let $\Lambda, \Gamma \subseteq \mathbb{R}^d$ and let $K \subseteq \mathbb{R}^d$ be a compact set. Let $g \in C(\mathbb{R}^d)$ be a window function such that for every $\omega \in K - K$ the system of translates*

$$\mathfrak{T}(g_\omega, \Lambda) = \{T_\lambda g_\omega : \lambda \in \Lambda\}$$

is complete in $(C(K), \|\cdot\|_\infty)$. If $\Gamma \subseteq \mathbb{R}^d$ is a uniqueness set for $PW_{K-K}^1(\mathbb{R}^d)$, then $\Lambda \times \Gamma$ is a uniqueness set for the STFT phase retrieval problem on $L^2(K)$. In particular, this holds whenever Γ is a lattice such that the difference set $K - K$ is contained in a fundamental domain of the reciprocal lattice Γ^* .

Note that the completeness condition is not trivial: even if g is particularly concentrated in phase space and its system of translates $\mathfrak{T}(g, \Lambda)$ is complete, this is in general not related to the completeness of $\mathfrak{T}(g_\omega, \Lambda)$ for all $\omega \in K - K$.

The proof of Theorem 2.12 has several steps:

1. For fixed $x \in \mathbb{R}^d$, the restriction $|V_g f(x, \cdot)|^2 = V_g f(x, \cdot) \cdot \overline{V_g f(x, \cdot)}$ is a band-limited function in $PW_{K-K}^1(\mathbb{R}^d)$. Thus, for each $x \in \mathbb{R}^d$, it is uniquely determined by the samples

$$|V_g f(x, \gamma)|^2, \quad \gamma \in \Gamma.$$

2. Equivalently, we can recover the Fourier transform of $|V_g f(x, \cdot)|^2$, which is given by

$$\mathcal{F}|V_g f(x, \cdot)|^2(\omega) = \int_K f_\omega(t) \overline{T_x g_\omega(t)} dt, \quad \omega \in \mathbb{R}^d.$$

This is precisely where the completeness property comes into play: it implies that we can recover f_ω , $\omega \in \mathbb{R}^d$, from the samples

$$\mathcal{F}|V_g f(\lambda, \cdot)|^2(\omega), \quad \lambda \in \Lambda, \omega \in \mathbb{R}^d.$$

3. If $f_\omega = h_\omega$ for some compactly supported $h \in L^2(K)$, then $\mathcal{F}f_\omega = \mathcal{F}h_\omega$ for all $\omega \in \mathbb{R}^d$. Therefore, $V_f f = V_h h$ on $\mathbb{R}^{2d}$, which implies that $f = \tau h$ for some $\tau \in \mathbb{T}$ [58].

The rest of Paper 5 is focused on finding windows g such that the functions g_ω are in a class of functions whose completeness in $C(K)$ is verifiable through complex analysis tools.

To state our class of functions, we recall that an entire function F on $\mathbb{C}^d$ has exponential type $\alpha \geq 0$ if for all $\varepsilon > 0$ exists a constant $C_\varepsilon > 0$ such that

$$|F(z)| \leq C_\varepsilon e^{(\alpha+\varepsilon)\|z\|_1}, \quad z \in \mathbb{C}^d.$$

We denote by $\mathcal{E}_\alpha(\mathbb{C}^d)$ the set of entire functions of exponential type α .

The simplest example of a window which will satisfy our requirements is the Gaussian window $g(t) = e^{-\pi|t|^2}$, in which case, g_ω are generalized Gaussian functions. More generally, let g belong to the class $\mathcal{P}_{\alpha,\beta}(\mathbb{C}^d)$, for some $0 \leq \alpha, \beta < \infty$, where

$$\mathcal{P}_{\alpha,\beta}(\mathbb{C}^d) := \left\{ z \mapsto p(z)e^{-z^T A z - b^T z} : \right. \\ \left. p \in \mathcal{E}_\alpha(\mathbb{C}^d), A \in \mathbb{C}^{d \times d}, b \in \mathbb{C}^d, \|A + A^T\|_1 \leq \beta \right\}.$$

We further require the functions $\varphi = g_\omega$ to *not* be mean-periodic. Mean-periodicity was studied, among others, by Delsarte, Schwartz, and Kahane [45, 46, 93, 124] and is a generalization of periodicity. A function $\varphi \in C(\mathbb{R}^d)$ is mean-periodic if there exists a complex Radon measure μ , $\mu \neq 0$, supported on a compact set $K \subseteq \mathbb{R}^d$ such that $\varphi * \mu = 0$. Equivalently, since the space of Radon measures $M(K)$ is the dual space of $C(K)$, φ is mean-periodic if the system of translates $\mathfrak{T}(\varphi, \mathbb{R}^d)$ is *not* complete in $C(K)$. The simplest example is a 1-periodic function, in which case the convolution $\varphi * (\delta_0 - \delta_1) = \varphi - T_1 \varphi$ vanishes. In light of Theorem 2.12, mean-periodic functions are not suitable as windows.

Finally, by duality, the family of translates $\mathfrak{T}(\varphi, \Lambda)$ is complete in $C(K)$ if and only if Λ is a uniqueness set for

$$\mathcal{O}_\varphi(K) := \left\{ F : \mathbb{C}^d \rightarrow \mathbb{C} : F(z) = \int_K T_z \varphi d\mu, \mu \in M(K) \right\}.$$

Indeed, a function $F \in \mathcal{O}_\varphi(K)$ which vanishes identically corresponds to some $\mu \in M(K)$ whose convolution $\varphi * \mu$ also vanishes identically. If $\varphi = g_\omega$ is not mean-periodic, then $\mu \equiv 0$. By duality, this implies the completeness of $\mathfrak{T}(g_\omega, \Lambda)$.

Theorem 2.13. *Let $K \subseteq \mathbb{R}^d$ be a compact set, $\varphi \in \mathcal{P}_{\alpha,\beta}(\mathbb{C}^d)$ not mean-periodic, and*

$$\sigma \geq \alpha + \frac{\beta}{2} \Delta(K), \quad \Delta(K) := \max_{t, t' \in K} \|t - t'\|_\infty.$$

If $\Lambda \subseteq \mathbb{R}^d$ is a uniqueness set for $\mathcal{E}_\sigma(\mathbb{C}^d)$, then $\mathfrak{T}(\varphi, \Lambda)$ is complete in $C(K)$.

To prove the statement, we showed that the functions $F \in \mathcal{O}_\varphi(K)$ are, up to a Gaussian factor independent of F , functions of exponential type σ . The simplest

case is the Gaussian function $\varphi(t) = e^{-|t|^2/2}$ and the hypercube $K = [-\sigma, \sigma]^d$, $\sigma > 0$. In this case, for any $F \in \mathcal{O}_\varphi(K)$,

$$\varphi(z)^{-1}F(z) = \varphi(z)^{-1}L_\mu(T_z\varphi) = \int_K \varphi(t)e^{z \cdot t} d\mu(t),$$

which is clearly in $\mathcal{E}_\sigma(\mathbb{C}^d)$.

Theorem 2.13 in combination with Theorem 2.12 implies the following phase retrieval result.

Theorem 2.14. *Let $K \subseteq \mathbb{R}^d$ be a compact set, $\varphi \in \mathcal{P}_{\alpha,\beta}(\mathbb{C}^d) \cap L^2(\mathbb{R}^d)$, $\varphi \neq 0$, and*

$$\sigma \geq 2\alpha + \beta\Delta(K).$$

If $\Lambda \subseteq \mathbb{R}^d$ is a uniqueness set for $\mathcal{E}_\sigma(\mathbb{C}^d)$ and if Γ is a uniqueness set for $PW_{K-K}^1(\mathbb{R}^d)$, then $\Lambda \times \Gamma$ is a uniqueness set for the STFT phase retrieval problem on $L^2(K)$. In particular, this holds for $K \subseteq [-\kappa, \kappa]^d$ for some $\kappa > 0$ and for $\sigma \geq 2\alpha + 2\beta\kappa$.

Most notably, Theorem 2.14 applies to the Hermite functions since they belong to $\mathcal{P}_{0,2\pi}(\mathbb{C}^d) \cap L^2(\mathbb{R}^d)$.

As noted several times already, the phaseless sampling of the short-time Fourier transform on lattices is not injective on $L^2(\mathbb{R}^d)$. In fact, it has been shown that if the window is the Gaussian window $g(t) = 2^{1/4}e^{-\pi t^2}$, the set of functions

$$\left\{ f \in L^2(\mathbb{R}^d) : \exists h \in L^2(\mathbb{R}^d) \text{ with } f \not\sim h, |V_g f(\lambda)| = |V_g h(\lambda)| \ \forall \lambda \in \Lambda \right\}$$

is dense in $\mathbb{R}^2$ for all lattices $\Lambda \subseteq \mathbb{R}^2$ [5,6]. Contrary to this, we show that if we omit the lattice structure, the phaseless sampling of the short-time Fourier transform is injective on a dense subset of $L^2(\mathbb{R}^d)$. We combine a modification of Theorem 2.12 and a result by Kahane on completeness of exponentials [94] to obtain a uniqueness result for a dense subset of $L^2(\mathbb{R}^d)$.

Theorem 2.15. *Let $g \in C(\mathbb{R}^d)$ be a real, multivariate Gaussian window function. Let $\Lambda \subseteq \mathbb{R} \setminus \{0\}$ be a sequence of distinct real numbers of density zero such that $\sum_{\lambda \in \Lambda} \frac{1}{|\lambda|} = \infty$. Further, let*

$$\mathcal{V} := \bigcup_{\kappa > 0} L^4[-\kappa, \kappa]^d.$$

Then $\mathcal{V}$ is dense in $L^2(\mathbb{R}^d)$ and there exists a separated set $\mathcal{K} \subseteq \mathbb{R}$ such that the following holds:

- (i) *every $f \in \mathcal{V}$ is determined up to a global phase by $|V_g f(\Lambda^d \times \mathcal{K}^d)|$;*
- (ii) *$\Lambda^d \times \mathcal{K}^d$ has density $\rho(\Lambda^d \times \mathcal{K}^d) = 0$.*

2.6 Benedicks-type uncertainty principle for metaplectic time-frequency representations

As discussed in Subsection 1.7, there are various ways of describing fundamental concentration obstructions for a pair $(f, \mathcal{F}f)$ or a joint time-frequency representation such as $V_g f$. Paper 6 investigates a Benedicks-type uncertainty principle for the metaplectic time-frequency representations

$$W_{\mathcal{A}}(f, g) = \mu(\mathcal{A})(f \otimes \bar{g}), \quad f, g \in L^2(\mathbb{R}^d), \quad \mathcal{A} \in \mathrm{Sp}(4d, \mathbb{R}).$$

We aim to obtain a systematic solution to the following support problem.

Let $\mathcal{A} \in \mathrm{Sp}(4d, \mathbb{R})$ be given. Are there non-zero functions $f, g \in L^2(\mathbb{R}^d)$ such that $W_{\mathcal{A}}(f, g)$ is supported on a set of finite measure? If so, is there a non-zero function f such that $W_{\mathcal{A}}(f, f)$ is supported on a set of finite measure?

There has been partial progress for several types of uncertainty principles. For instance, if the metaplectic operator is a composition of a partial Fourier transform and a dilation, i.e., $\mu(\mathcal{A}) = \mathcal{F}_2 \mathcal{D}_L$, under additional assumptions on $L \in \mathrm{GL}(2d, \mathbb{R})$, a Benedicks-type uncertainty principle holds [13]. Other authors have studied orthogonal symplectic matrices and Heisenberg's uncertainty principle [146, 147, 148]. However, so far it remained open: what would the exact deciding property of $\mathcal{A} \in \mathrm{Sp}(4d, \mathbb{R})$ be as to whether or not a particular uncertainty principle can hold for the metaplectic time-frequency representation $W_{\mathcal{A}}$? Note that this is not a trivial problem, even if we look only at the simplest generalization of the Wigner distribution, the τ -Wigner distribution

$$W_{\tau}(f, g)(x, \omega) := \int_{\mathbb{R}^d} f(x + \tau t) \overline{g(x - (1 - \tau)t)} e^{-2\pi i \omega t} dt, \quad (x, \omega) \in \mathbb{R}^{2d}.$$

The classical Wigner distribution is $W(f, g) = W_{1/2}(f, g)$, while the boundary cases $\tau = 0$ and $\tau = 1$ are the Rihaczek and the conjugate Rihaczek distribution respectively;

$$W_0(f, g)(x, \omega) = e^{2\pi i \omega \cdot x} f(x) \overline{\mathcal{F}g(\omega)}, \quad W_1(f, g)(x, \omega) = e^{2\pi i \omega \cdot x} \mathcal{F}f(\omega) \overline{g(x)}.$$

By Theorem 1.6, Benedicks's uncertainty principle admits a phase space formulation for some $\tau \in [0, 1]$, e.g., $\tau = \frac{1}{2}$. However, if $\tau = 0$, then there exist functions $f, g \in L^2(\mathbb{R}^d)$, $f \not\equiv 0$, $g \not\equiv 0$, such that

$$|W_0(f, g)| = |f| \otimes |\mathcal{F}g|$$

is even supported on a compact set. In contrast, if we only consider the quadratic Rihaczek distribution $W_0(f, f)$, then by Benedicks's theorem [11, 16]

$$|W_0(f, f)| = |f| \otimes |\mathcal{F}f|$$

is supported on a set of finite measure if and only if $f \equiv 0$.

To formulate the main result, we recall that $\mathrm{Sp}(2d, \mathbb{R})$ is a Lie group, and $\mathrm{Sp}(2d, \mathbb{R}) \cap \mathrm{O}(2d, \mathbb{R})$ is a maximal compact subgroup of $\mathrm{Sp}(2d, \mathbb{R})$. The subgroup is isomorphic to the unitary group $\mathrm{U}(d, \mathbb{C})$ via the isomorphism [44, 53]

$$\mathrm{U}(d, \mathbb{C}) \rightarrow \mathrm{Sp}(2d, \mathbb{R}) \cap \mathrm{O}(2d, \mathbb{R}), \quad U = A + iB \mapsto \mathcal{R}_{A+iB} := \begin{pmatrix} A & B \\ -B & A \end{pmatrix}.$$

Furthermore, $\mathrm{Sp}(2d, \mathbb{R})$ is a semisimple connected Lie group, so it admits an Iwasawa decomposition [96]. With a minor modification of the Iwasawa decomposition, we can decompose any symplectic matrix $\mathcal{A} \in \mathrm{Sp}(2d, \mathbb{R})$ as a product of a shearing matrix $\mathcal{V}_Q$, a block-diagonal matrix $\mathcal{D}_L$, and an orthogonal matrix $\mathcal{R}_U$;

$$\mathcal{A} = \mathcal{V}_Q \cdot \mathcal{D}_L \cdot \mathcal{R}_{A+iB}. \quad (2.5)$$

In the following, we refer to (2.5) as the pre-Iwasawa decomposition of $\mathcal{A} \in \mathrm{Sp}(2d, \mathbb{R})$. This factorization is beneficial since we know that the operators $\mu(\mathcal{V}_Q)$ and $\mu(\mathcal{D}_L)$ are linear "chirps" and dilations respectively (1.7). In particular, they have no substantial impact on the support:

$$\mathrm{supp} \, W_{\mathcal{A}}(f, g) = L \cdot \mathrm{supp} \, \mu(\mathcal{R}_U)(f \otimes \bar{g}) = L \cdot \mathrm{supp} \, W_{\mathcal{R}_U}(f, g). \quad (2.6)$$

Thus, we restrict the proof discussion to metaplectic time-frequency representations generated by orthogonal symplectic matrices. For generic symplectic matrices, the deciding factor is only the projection onto $\mathrm{O}(4d, \mathbb{R}) \cap \mathrm{Sp}(4d, \mathbb{R})$.

Another trivial case where we can exclude *any* type of an uncertainty principle are orthogonal symplectic matrices $\mathcal{R}_U \in \mathrm{O}(4d, \mathbb{R}) \cap \mathrm{Sp}(4d, \mathbb{R})$ with block-diagonal $U = \mathrm{diag}(U_1, U_2)$, $U_1, U_2 \in \mathrm{U}(d, \mathbb{C})$. In this case,

$$\mu(\mathcal{R}_U)(f \otimes \bar{g}) = \mu(\mathcal{R}_{U_1})f \otimes \overline{\mu(\mathcal{R}_{U_2})g}. \quad (2.7)$$

The main result of Paper 6 states that these two types of redundancies are the only cases where there is no uncertainty principle.

The first step in this direction is to understand a special case of symplectic matrices, namely, *free symplectic matrices*. A symplectic matrix $\mathcal{A} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ is called free if $\det B \neq 0$. Equivalently, these are the matrices which decompose as $\mathcal{A} = \mathcal{V}_Q \mathcal{D}_L \mathcal{J} \mathcal{V}_P$ [44]. As previously observed, the factors $\mathcal{V}_Q, \mathcal{D}_L$ are irrelevant for uncertainty principles and we only look into $\mathcal{J} \mathcal{V}_P$.

Theorem 2.16. *Let $\mathcal{A} = \mathcal{J} \mathcal{V}_P \in \mathrm{Sp}(4d, \mathbb{R})$ be a free symplectic matrix with*

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix},$$

such that the blocks on the side diagonal $P_{12} = P_{21}^t$ are not the zero matrix. Then the support of $W_{\mathcal{A}}(f, g)$ has finite measure for some functions $f, g \in L^2(\mathbb{R}^d)$ if and only if $f \equiv 0$ or $g \equiv 0$.

The proof relies on two observations:

$$\mathcal{J}\mathcal{V}_P = \mathcal{D}_{\text{diag}(W_1, W_2)} \mathcal{J}\mathcal{V} \begin{pmatrix} 0 & P_{12} \\ P_{21} & 0 \end{pmatrix} \mathcal{D}_{\text{diag}(W_1, W_2)} \mathcal{V}_{\text{diag}(P_{11}, P_{22})}$$

for some orthogonal matrices $W_1, W_2 \in \text{O}(d, \mathbb{R})$ and a diagonal matrix Γ of rank $1 \leq k \leq d$ with $W_1 \Gamma W_2^t = P_{12} = P_{21}^t$. With the same argument as in (2.6) and (2.7), we can assume that $W_1 = W_2 = I$ and $P_{11} = P_{22} = 0$.

The second part is rewriting $W_{\mathcal{J}\mathcal{V}_P}(f, g)$ as a dilated *partial* short-time Fourier transform

$$V_g^k \tilde{f}(x_1, x_2, \omega_1, \omega_2) = \int_{\mathbb{R}^k} \tilde{f}(t, x_2) \overline{\tilde{g}(t - x_1, -\omega_2)} e^{-2\pi i t \cdot \omega_1} dt$$

of some $\tilde{f}, \tilde{g}$ which are zero if and only if f, g are zero respectively. By restricting to subspaces of $\mathbb{R}^{2d}$ where $V_g^k \tilde{f}$ is the classical short-time Fourier transform of functions in $L^2(\mathbb{R}^k)$, we show that it cannot be supported on a set of finite measure unless $\tilde{f}$ or $\tilde{g}$ is the zero function, implying the desired statement.

The full characterization is a reduction to the special case in Theorem 2.16. By (2.7), if we choose $U_1 = U_2 = \tau I$ for some τI , we can reformulate as follows:

$$W_{\mathcal{R}_U}(f, g) = W_{\mathcal{R}_{\tau U}}(\mathcal{R}_{\tau I} f, \mathcal{R}_{\tau I} g).$$

Thus, there exist $f, g \in L^2(\mathbb{R}^d)$, $f, g \not\equiv 0$, such that $\text{supp } W_{\mathcal{R}_U}(f, g)$ has finite measure if and only if there exist $f_\tau, g_\tau \in L^2(\mathbb{R}^d)$, $f_\tau, g_\tau \not\equiv 0$ such that $\text{supp } W_{\mathcal{R}_{\tau U}}(f_\tau, g_\tau)$ has finite measure and constitutes the first step in the reduction.

Secondly, we show that for all $\mathcal{R}_U \in \text{O}(4d, \mathbb{R}) \cap \text{Sp}(2d, \mathbb{R})$ there exists a $\tau \in \mathbb{T}$ such that $\mathcal{R}_{\tau U}$ is free and we can apply Theorem 2.16. However, whether a τ exists such that

$$\mathcal{R}_{\tau U} = \mathcal{V}_Q \mathcal{D}_L \mathcal{J}\mathcal{V}_P$$

with a block-diagonal matrix P is not easy to verify and we present an equivalent characterization.

Theorem 2.17. *Let $\mathcal{A} \in \text{Sp}(4, \mathbb{R})$ be an arbitrary symplectic matrix with pre-Iwasawa decomposition*

$$\mathcal{A} = \mathcal{V}_Q \mathcal{D}_L \mathcal{R}_U.$$

Then the following statements hold.

- (i) *Assume $U^t U$ is not block-diagonal. Then a Benedicks-type uncertainty principle holds:*

*If $W_{\mathcal{A}}(f, g)$ is supported on a set of finite measure for some $f, g \in L^2(\mathbb{R}^d)$,
then $f \equiv 0$ or $g \equiv 0$.*

- (ii) *Assume $U^t U$ is block-diagonal. Then there exist $f, g \in L^2(\mathbb{R}^d)$, $f \not\equiv 0$, $g \not\equiv 0$, such that $W_{\mathcal{A}}(f, g)$ is supported on a compact set, in particular, on a set of finite measure.*

Reformulated as an algorithm, we obtain

Input: $\mathcal{A} \in \text{Sp}(4d, \mathbb{R})$.

1. Compute the pre-Iwasawa decomposition $\mathcal{A} = \mathcal{V}_Q \mathcal{D}_L \mathcal{R}_U$.
 2. Compute $U^t U$.
 3. If $U^t U$ is block-diagonal, then there exist $f, g \in L^2(\mathbb{R}^d)$, $f \not\equiv 0$, $g \not\equiv 0$, such that $W_{\mathcal{A}}(f, g)$ is supported on a compact set.
 4. Otherwise, whenever $W_{\mathcal{A}}(f, g)$ is supported on a set of finite measure, then $f \equiv 0$ or $g \equiv 0$.
-

The second part of Paper 6 focuses on *quadratic metaplectic time-frequency representations* $W_{\mathcal{A}}(f, f)$, $f \in L^2(\mathbb{R}^d)$. The Rihaczek distribution, also given as the τ -Wigner distribution with $\tau = 0$,

$$R(f, g)(x, \omega) = W_0(f, g)(x, \omega) = e^{2\pi i \omega \cdot x} f(x) \overline{\mathcal{F}g(\omega)}$$

is an intriguing example that motivated us to study the quadratic version $W_{\mathcal{A}}(f, f)$. Clearly, if the matrix $\mathcal{A}$ is covered by Theorem 2.17 (i), then $W_{\mathcal{A}}(f, f)$ cannot be supported on a set of finite measure unless f is the zero function. Otherwise, the problem is reduced to looking into representations of the type

$$\mu(\mathcal{A}_1)f \otimes \overline{\mu(\mathcal{A}_2)f} = \mu(\mathcal{A}_1 \mathcal{A}_2^{-1})\tilde{f} \otimes \tilde{f}, \quad \tilde{f} = \mu(\mathcal{A}_2)f.$$

The full characterization of the quadratic metaplectic time-frequency representations which fulfill a Benedicks-type uncertainty principle is summarized as follows.

Theorem 2.18. *Let $\mathcal{A} \in \text{Sp}(4d, \mathbb{R})$ be an arbitrary symplectic matrix with pre-Iwasawa decomposition*

$$\mathcal{A} = \mathcal{V}_Q \mathcal{D}_L \mathcal{R}_U.$$

Then the following statements hold.

- (i) *Assume $U^t U \neq \text{diag}(\Delta, \overline{\Delta})$ for all unitary matrices $\Delta \in \text{U}(d, \mathbb{C})$. Then a Benedicks-type uncertainty principle holds:*

If $W_{\mathcal{A}}(f, f)$ is supported on a set of finite measure for some $f \in L^2(\mathbb{R}^d)$, then $f \equiv 0$.

- (ii) *Assume $U^t U = \text{diag}(\Delta, \overline{\Delta})$ for some unitary matrix $\Delta \in \text{U}(d, \mathbb{C})$. Then there exists a $f \in L^2(\mathbb{R}^d)$, $f \not\equiv 0$, such that the support of $W_{\mathcal{A}}(f, f)$ is compact, in particular, has finite measure.*

The proof idea is comparable to the method in [85], with a more involved matrix decomposition and, similarly to Theorem 2.16, restrictions to subspaces of $\mathbb{R}^{2d}$.

References

- [1] ABDELMALEK, R., MNASRI, Z., AND BENZARTI, F. Audio signal reconstruction using phase retrieval: Implementation and evaluation. *Multimed. Tools Appl.* 81, 11 (2022), 15919–15946.
- [2] ADCOCK, B., GATARIC, M., AND HANSEN, A. C. Density theorems for nonuniform sampling of bandlimited functions using derivatives or bunched measurements. *J. Fourier Anal. Appl.* 23, 6 (2017), 1311–1347.
- [3] AKHIESER, N. I. *Theory of approximation*. Courier Corporation, 2013.
- [4] AKHIEZER, N. I. *The classical moment problem and some related questions in analysis*. Hafner Publishing Co., New York, 1965. Translated by N. Kemmer.
- [5] ALAIFARI, R., BARTOLUCCI, F., STEINERBERGER, S., AND WELLERSHOFF, M. On the connection between uniqueness from samples and stability in Gabor phase retrieval. *arXiv:2205.07013* (2022).
- [6] ALAIFARI, R., BARTOLUCCI, F., AND WELLERSHOFF, M. Uncovering the limits of uniqueness in sampled gabor phase retrieval: A dense set of counterexamples in $l^2(\mathbb{R})$. In *2023 International Conference on Sampling Theory and Applications (SampTA)* (2023), IEEE, pp. 1–5.
- [7] ALAIFARI, R., AND GROHS, P. Phase Retrieval In The General Setting Of Continuous Frames For Banach Spaces. *SIAM J. Math. Anal.* 49 (2016).
- [8] ALAIFARI, R., AND WELLERSHOFF, M. Phase Retrieval from Sampled Gabor Transform Magnitudes: Counterexamples. *J. Fourier Anal. Appl.* 28, 1 (2021), 9.
- [9] ALDROUBI, A., AND GRÖCHENIG, K. Beurling-Landau-type theorems for non-uniform sampling in shift invariant spline spaces. *J. Fourier Anal. Appl.* 6, 1 (2000), 93–103.
- [10] ALDROUBI, A., AND GRÖCHENIG, K. Nonuniform Sampling and Reconstruction in Shift-Invariant Spaces. *SIAM Rev.* 43, 4 (2001), 585–620.
- [11] AMREIN, W. O., AND BERTHIER, A. M. On support properties of L^p -functions and their Fourier transforms. *J. Functional Analysis* 24, 3 (1977), 258–267.
- [12] ATZMON, A., AND OLEVSKIĬ, A. Completeness of integer translates in function spaces on $\mathbf{R}$. *J. Approx. Theory* 87, 3 (1996), 291–327.
- [13] BAYER, D. *Bilinear Time-Frequency Distributions and Pseudodifferential Operators*. PhD thesis, University of Vienna, Vienna, Austria, 2010.

- [14] BECKNER, W. Inequalities in Fourier analysis. *Ann. of Math. (2)* 102, 1 (1975), 159–182.
- [15] BELOV, Y., KULIKOV, A., AND LYUBARSKII, Y. Gabor frames for rational functions. *Invent. Math.* 231, 2 (2023), 431–466.
- [16] BENEDICKS, M. On Fourier transforms of functions supported on sets of finite Lebesgue measure. *J. Math. Anal. Appl.* 106, 1 (1985), 180–183.
- [17] BÉNYI, A., AND OKOUDJOU, K. A. *Modulation spaces—with applications to pseudodifferential operators and nonlinear Schrödinger equations*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2020.
- [18] BERNSTEIN, S. Sur la définition et les propriétés des fonctions analytiques d’une variable réelle. *Math. Ann.* 75, 4 (1914), 449–468.
- [19] BERNSTEIN, S. Sur les fonctions absolument monotones. *Acta Math.* 52, 1 (1929), 1–66.
- [20] BONAMI, A., DEMANGE, B., AND JAMING, P. Hermite functions and uncertainty principles for the Fourier and the windowed Fourier transforms. *Rev. Mat. Iberoamericana* 19, 1 (2003), 23–55.
- [21] BORN, M., AND WOLF, E. *Principles of Optics: 60th Anniversary Edition*, 7 ed. Cambridge University Press, 2020.
- [22] BORWEIN, J. M., AND BORWEIN, P. B. A cubic counterpart of Jacobi’s identity and the AGM. *Trans. Amer. Math. Soc.* 323, 2 (1991), 691–701.
- [23] BRUNA, J., OLEVSKII, A., AND ULANOVSKII, A. Completeness in $L^1(\mathbb{R})$ of discrete translates. *Rev. Mat. Iberoam.* 22, 1 (2006), 1–16.
- [24] BÉTERMIN, L., FAULHUBER, M., AND STEINERBERGER, S. A variational principle for Gaussian lattice sums. *arXiv preprint 2110.06008* (2021).
- [25] CAHILL, J., CASAZZA, P. G., AND DAUBECHIES, I. Phase retrieval in infinite-dimensional Hilbert spaces. *Trans. Amer. Math. Soc. Ser. B* 3, 3 (2016), 63–76.
- [26] CHEN, Y., CHENG, C., SUN, Q., AND WANG, H. Phase retrieval of real-valued signals in a shift-invariant space. *Appl. Comput. Harmon. Anal.* 49, 1 (2020), 56–73.
- [27] CHENG, C., JIANG, J., AND SUN, Q. Phaseless sampling and reconstruction of real-valued signals in shift-invariant spaces. *J. Fourier Anal. Appl.* 25, 4 (2019), 1361–1394.

- [28] CHRIST, M., PINEAU, B., AND TAYLOR, M. A. Examples of hölder-stable phase retrieval. *arXiv:2205.00187* (2022).
- [29] CHRISTENSEN, O. *An Introduction to Frames and Riesz Bases*, 2 ed. Springer International Publishing, 2016.
- [30] CHRISTENSEN, O., DENG, B., AND HEIL, C. Density of Gabor frames. *Appl. Comput. Harmon. Anal.* 7, 3 (1999), 292–304.
- [31] COHN, H., AND KUMAR, A. Universally optimal distribution of points on spheres. *J. Amer. Math. Soc.* 20, 1 (2007), 99–148.
- [32] COHN, H., KUMAR, A., MILLER, S. D., RADCHENKO, D., AND VIAZOVSKA, M. Universal optimality of the E_8 and Leech lattices and interpolation formulas. *Ann. of Math. (2)* 196, 3 (2022), 983–1082.
- [33] COOK, C. E., AND BERNFELD, M. *Radar Signals: an Introduction to Theory and Applications*. Academic Press, New York, 1947.
- [34] CORBETT, J. V., AND HURST, C. A. Are wave functions uniquely determined by their position and momentum distributions? *J. Austral. Math. Soc. Ser. B* 20, 2 (1977/78), 182–201.
- [35] CORDERO, E., AND GIACCHI, G. Metaplectic Gabor frames and symplectic analysis of time-frequency spaces. *Appl. Comput. Harmon. Anal.* 68 (2024), Paper No. 101594, 31.
- [36] CORDERO, E., GIACCHI, G., AND RODINO, L. Wigner Analysis of Operators. Part II: Schrödinger equations. *arXiv:2208.00505* (2022).
- [37] CORDERO, E., GIACCHI, G., RODINO, L., AND VALENZANO, M. Wigner Analysis of Fourier Integral Operators with symbols in the Shubin classes. *arXiv:2402.02809* (2024).
- [38] CORDERO, E., AND RODINO, L. *Time-frequency analysis of operators*, vol. 75 of *De Gruyter Studies in Mathematics*. De Gruyter, Berlin, 2020.
- [39] CORDERO, E., AND RODINO, L. Wigner analysis of operators. Part I: Pseudodifferential operators and wave fronts. *Appl. Comput. Harmon. Anal.* 58 (2022), 85–123.
- [40] COWLING, M. G., AND PRICE, J. F. Bandwidth versus time concentration: the Heisenberg-Pauli-Weyl inequality. *SIAM J. Math. Anal.* 15, 1 (1984), 151–165.
- [41] DA SILVA, J. C., AND MENZEL, A. Elementary signals in ptychography. *Opt. Express* 23, 26 (2015), 33812–33821.

- [42] DAUBECHIES, I., LANDAU, H. J., AND LANDAU, Z. Gabor time-frequency lattices and the Wexler-Raz identity. *J. Fourier Anal. Appl.* 1, 4 (1995), 437–478.
- [43] DE GOSSON, M. *Symplectic Geometry and Quantum Mechanics*. Birkhäuser Basel, 2006.
- [44] DE GOSSON, M. A. *Symplectic Methods in Harmonic Analysis and in Mathematical Physics*. Springer Basel, 2011.
- [45] DELSARTE, J. Les fonctions moyenne-périodiques. *J. Math. Pures et Appl.* 14 (1935), 403–453.
- [46] DELSARTE, J. Theorie des Fonctions Moyenne-Periodiques de Deux Variables. *Ann. of Math* 72 (1960), 121–178.
- [47] DEMANGE, B. Uncertainty principles for the ambiguity function. *J. London Math. Soc. (2)* 72, 3 (2005), 717–730.
- [48] DEMANGE, B. Uncertainty principles associated to non-degenerate quadratic forms. *Mém. Soc. Math. Fr. (N.S.)*, 119 (2009), 98.
- [49] DONOHO, D. L., AND STARK, P. B. Uncertainty principles and signal recovery. *SIAM J. Appl. Math.* 49, 3 (1989), 906–931.
- [50] DUFFIN, R. J., AND SCHAEFFER, A. C. A class of nonharmonic Fourier series. *Trans. Amer. Math. Soc.* 72 (1952), 341–366.
- [51] FANNJIANG, A., AND STROHMER, T. The numerics of phase retrieval. *Acta Numer.* 29 (2020), 125–228.
- [52] FAULHUBER, M. A short note on the frame set of odd functions. *Bull. Aust. Math. Soc.* 98, 3 (2018), 481–493.
- [53] FOLLAND, G. B. *Harmonic analysis in phase space*, vol. 122 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1989.
- [54] FOLLAND, G. B., AND SITARAM, A. The uncertainty principle: a mathematical survey. *J. Fourier Anal. Appl.* 3, 3 (1997), 207–238.
- [55] FORSTNER, A., KRAHMER, F., MELNYK, O., AND SISSOUNO, N. Well-conditioned ptychographic imaging via lost subspace completion. *Inverse Problems* 36, 10 (2020), 105009, 26.
- [56] GHOSH, R., AND SELVAN, A. On Gabor frames generated by B-splines, totally positive functions, and hermite functions. *arXiv:2304.11322* (2023).

- [57] GRÖCHENIG, K. Reconstruction algorithms in irregular sampling. *Math. Comp.* 59, 199 (1992), 181–194.
- [58] GRÖCHENIG, K. *Foundations of time-frequency analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Inc., Boston, MA, 2001.
- [59] GRÖCHENIG, K. Phase-Retrieval in Shift-Invariant Spaces with Gaussian Generator. *J. Fourier Anal. Appl.* 26, 3 (Jun 2020), 52.
- [60] GRÖCHENIG, K. Totally positive functions and Gabor frames over rational lattices. *Adv. Math.* 427 (2023), Paper No. 109113, 12.
- [61] GRÖCHENIG, K., AND KOPPENSTEINER, S. Gabor frames: characterizations and coarse structure. In *New trends in applied harmonic analysis. Vol. 2—harmonic analysis, geometric measure theory, and applications*, Appl. Numer. Harmon. Anal. Birkhäuser/Springer, Cham, [2019] ©2019, pp. 93–120.
- [62] GRÖCHENIG, K., AND LYUBARSKII, Y. Gabor frames with Hermite functions. *C. R. Math. Acad. Sci. Paris* 344, 3 (2007), 157–162.
- [63] GRÖCHENIG, K., ROMERO, J. L., AND STÖCKLER, J. Sharp results on sampling with derivatives in shift-invariant spaces and multi-window Gabor frames. *Constr. Approx.* 51, 1 (2020), 1–25.
- [64] GRÖCHENIG, K., AND SCHWAB, H. Fast local reconstruction methods for nonuniform sampling in shift-invariant spaces. *SIAM J. Matrix Anal. Appl.* 24, 4 (2003), 899–913.
- [65] GRÖCHENIG, K., AND STÖCKLER, J. Gabor frames and totally positive functions. *Duke Math. J.* 162, 6 (2013), 1003–1031.
- [66] GRÖCHENIG, K., AND ZIMMERMANN, G. Hardy’s theorem and the short-time Fourier transform of Schwartz functions. *J. London Math. Soc. (2)* 63, 1 (2001), 205–214.
- [67] GROHS, P., KOPPENSTEINER, S., AND RATHMAIR, M. Phase Retrieval: Uniqueness and Stability. *SIAM Rev.* 62, 2 (2020), 301–350.
- [68] GROHS, P., AND LIEHR, L. On foundational discretization barriers in STFT phase retrieval. *J. Fourier Anal. Appl.* 28, 2 (2022), Paper No. 39, 21.
- [69] GROHS, P., AND LIEHR, L. Injectivity of Gabor phase retrieval from lattice measurements. *Appl. Comput. Harmon. Anal.* 62 (2023), 173–193.
- [70] GROHS, P., AND LIEHR, L. Non-uniqueness theory in sampled STFT phase retrieval. *SIAM J. Math. Anal.* 55, 5 (2023), 4695–4726.

- [71] GROHS, P., LIEHR, L., AND RATHMAIR, M. Multi-window STFT phase retrieval: lattice uniqueness. *arXiv:2207.10620* (2022).
- [72] GROHS, P., LIEHR, L., AND RATHMAIR, M. Phase retrieval in Fock space and perturbation of Liouville sets. *arXiv:2308.00385* (2023).
- [73] GROHS, P., AND RATHMAIR, M. Stable Gabor phase retrieval and spectral clustering. *Comm. Pure Appl. Math.* 72, 5 (2019), 981–1043.
- [74] GRÖCHENIG, K., ORTEGA-CERDÀ, J., AND ROMERO, J. L. Deformation of Gabor systems. *Adv. Math.* 277 (2015), 388–425.
- [75] GRÖCHENIG, K., ROMERO, J. L., AND STÖCKLER, J. Sampling theorems for shift-invariant spaces, Gabor frames, and totally positive functions. *Invent. Math.* 211, 3 (2017), 1119–1148.
- [76] GUANLEI, X., XIAOTONG, W., AND XIAOGANG, X. The logarithmic, Heisenberg’s and short-time uncertainty principles associated with fractional Fourier transform. *Signal Processing* 89, 3 (2009), 339–343.
- [77] HARDIN, D., AND TENPAS, N. Bounds for Periodic Energy and the Optimality of Two Periodic Point Configurations in the Plane. *arXiv:2307.15822* (2023).
- [78] HARDY, G. H. A Theorem Concerning Fourier Transforms. *J. London Math. Soc.* 8, 3 (1933), 227–231.
- [79] HAUSDORFF, F. Summationsmethoden und Momentfolgen. I. *Math. Z.* 9, 1-2 (1921), 74–109.
- [80] HAUSDORFF, F. Summationsmethoden und Momentfolgen. II. *Math. Z.* 9, 3-4 (1921), 280–299.
- [81] ISMAGILOV, R. S. On the Pauli problem. *Funktsional. Anal. i Prilozhen.* 30, 2 (1996), 82–84.
- [82] IWEN, M., PERLMUTTER, M., SISOONO, N., AND VISWANATHAN, A. Phase retrieval for $L^2([-\pi, \pi])$ via the provably accurate and noise robust numerical inversion of spectrogram measurements. *J. Fourier Anal. Appl.* 29, 1 (2023), Paper No. 8, 45.
- [83] JAMING, P. Principe d’incertitude qualitatif et reconstruction de phase pour la transformée de Wigner. *C. R. Acad. Sci. Paris Sér. I Math.* 327, 3 (1998), 249–254.
- [84] JAMING, P. Phase retrieval techniques for radar ambiguity problems. *J. Fourier Anal. Appl.* 5, 4 (Jul 1999), 309–329.

- [85] JAMING, P. A simple observation on the uncertainty principle for the fractional Fourier transform. *J. Fourier Anal. Appl.* 28, 3 (2022), Paper No. 51, 8.
- [86] JAMING, P., AND RATHMAIR, M. Gabor phase retrieval via semidefinite programming. *arXiv:2310.11214* (2023).
- [87] JANSSEN, A. J. E. M. Duality and biorthogonality for Weyl-Heisenberg frames. *J. Fourier Anal. Appl.* 1, 4 (1995), 403–436.
- [88] JANSSEN, A. J. E. M. Some Weyl-Heisenberg frame bound calculations. *Indag. Math.* 7, 2 (1996), 165–183.
- [89] JANSSEN, A. J. E. M. Proof of a conjecture on the supports of Wigner distributions. *J. Fourier Anal. Appl.* 4, 6 (1998), 723–726.
- [90] JANSSEN, A. J. E. M., AND STROHMER, T. Hyperbolic secants yield Gabor frames. *Appl. Comput. Harmon. Anal.* 12, 2 (2002), 259–267.
- [91] JIA, R.-Q., AND MICCHELLI, C. A. Using the Refinement Equations for the Construction of Pre-Wavelets II: Powers of Two. In *Curves and Surfaces*. Elsevier, 1991, pp. 209–246.
- [92] JIANG, Y., CHEN, Z., HAN, Y., DEB, P., GAO, H., XIE, S., PUROHIT, P., TATE, M. W., PARK, J., GRUNER, S. M., ELSEER, V., AND MULLER, D. A. Electron ptychography of 2d materials to deep sub-ångström resolution. *Nature* 559, 7714 (2018), 343–349.
- [93] KAHANE, J. *Lectures on On Mean Periodic Functions*. Tata Institute of Fundamental Research, 1959.
- [94] KAHANE, J.-P. Sur la totalité des suites d’exponentielles imaginaires. *Ann. Inst. Fourier* 8 (1958), 273–275.
- [95] KLOOS, T., AND STÖCKLER, J. Zak transforms and Gabor frames of totally positive functions and exponential B-splines. *J. Approx. Theory* 184 (2014), 209–237.
- [96] KNAPP, A. W. *Lie groups beyond an introduction*, vol. 140 of *Progress in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, 1996.
- [97] KRASIYOV, I. Inequalities for Laguerre polynomials. *East J. Approx.* 11, 3 (2005), 257–268.
- [98] KRASIYOV, I. Inequalities for orthonormal Laguerre polynomials. *J. Approx. Theory* 144, 1 (2007), 1–26.

- [99] KULIKOV, A. Openness of the frame set on the hyperbolas. *arXiv:2404.00585* (2024).
- [100] LANDAU, H. J. On the completeness of a set of translates. *J. Approximation Theory* 5 (1972), 438–440.
- [101] LEMVIG, J. On some Hermite series identities and their applications to Gabor analysis. *Monatshefte für Mathematik* 182, 4 (2017), 889–912.
- [102] LEV, N. Completeness of uniformly discrete translates in $l^p(\mathbb{R})$. *arXiv:2401.15588* (2024).
- [103] LEV, N., AND OLEVSKII, A. Wiener’s ‘closure of translates’ problem and Piatetski-Shapiro’s uniqueness phenomenon. *Ann. of Math. (2)* 174, 1 (2011), 519–541.
- [104] LYUBARSKII, Y. Frames in the Bargmann space of entire functions. In *Entire and Subharmonic Functions*. American Mathematical Society, Providence, RI, 1992, p. 167–180.
- [105] LYUBARSKII, Y., AND NES, P. G. Gabor frames with rational density. *Applied and Computational Harmonic Analysis* 34, 3 (2013), 488–494.
- [106] MALLAT, S. G. Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(\mathbb{R})$. *Trans. Am. Math. Soc.* 315, 1 (1989), 69.
- [107] MELNYK, O. On connections between Amplitude Flow and Error Reduction for phase retrieval and ptychography. *Sampl. Theory Signal Process. Data Anal.* 20, 2 (2022), 16.
- [108] MONTGOMERY, H. L. Minimal theta functions. *Glasgow Math. J.* 30, 1 (1988), 75–85.
- [109] NAZAROV, F. L. Local estimates for exponential polynomials and their applications to inequalities of the uncertainty principle type. *Algebra i Analiz* 5, 4 (1993), 3–66.
- [110] OLEVSKII, A., AND ULANOVSKII, A. Almost integer translates. Do nice generators exist? *J. Fourier Anal. Appl.* 10, 1 (2004), 93–104.
- [111] OLEVSKII, A., AND ULANOVSKII, A. Discrete Translates in Function Spaces. *Anal. Math.* 44, 2 (2018), 251–261.
- [112] OLEVSKII, A., AND ULANOVSKII, A. Discrete translates in $L^p(\mathbb{R})$. *Bull. Lond. Math. Soc.* 50, 4 (2018), 561–568.
- [113] PAULI, W. *Die allgemeinen Prinzipien der Wellenmechanik*. Springer-Verlag Berlin Heidelberg, 1990.

- [114] PINKUS, A. Density in Approximation Theory. *Surv. Approx. Theory* 1 (2005), 1–45.
- [115] RAMANATHAN, J., AND STEGER, T. Incompleteness of sparse coherent states. *Appl. Comput. Harmon. Anal.* 2, 2 (1995), 148–153.
- [116] RAWN, M. D. A stable nonuniform sampling expansion involving derivatives. *IEEE Trans. Inform. Theory* 35, 6 (1989), 1223–1227.
- [117] RAZAFINJATOVO, H. N. Iterative reconstructions in irregular sampling with derivatives. *J. Fourier Anal. Appl.* 1, 3 (1995), 281–295.
- [118] RICAUD, B., AND TORRÉSANI, B. A survey of uncertainty principles and some signal processing applications. *Adv. Comput. Math.* 40, 3 (2014), 629–650.
- [119] ROMERO, J. Sign retrieval in shift-invariant spaces with totally positive generator. *J. Fourier Anal. Appl.* 27, 2 (Mar 2021), 27.
- [120] RON, A. Introduction to shift-invariant spaces. Linear independence. In *Multivariate Approximation and Applications*. Cambridge University Press, 2001, pp. 112–151.
- [121] RON, A., AND SHEN, Z. Weyl-Heisenberg frames and Riesz bases in $L_2(\mathbf{R}^d)$. *Duke Math. J.* 89, 2 (1997), 237–282.
- [122] SCHEMPP, W. Radar ambiguity functions, the Heisenberg group, and holomorphic theta series. *Proc. Amer. Math. Soc.* 92, 1 (1984), 103–110.
- [123] SCHUMAKER, L. *Spline Functions: Basic Theory*, third ed. Cambridge University Press, 2007.
- [124] SCHWARTZ, L. Theorie Generale des Fonctions Moyenne-Periodiques. *Ann. of Math.* 48, 4 (1947), 857–929.
- [125] SEIP, K. Density theorems for sampling and interpolation in the Bargmann-Fock space I. *J. Reine Angew. Math.* 429, 429 (1992), 91–106.
- [126] SEIP, K., AND WALLSTÉN, R. Density theorems for sampling and interpolation in the Bargmann-Fock space II. *J. Reine Angew. Math.* 429 (1992), 107–114.
- [127] SELVAN, A. A., AND GHOSH, R. Sampling with derivatives in periodic shift-invariant spaces. *Numer. Funct. Anal. Optim.* 43, 13 (2022), 1591–1615.
- [128] SHAH, F. A., NISAR, K. S., LONE, W. Z., AND TANTARY, A. Y. Uncertainty principles for the quadratic-phase Fourier transforms. *Math. Methods Appl. Sci.* 44, 13 (2021), 10416–10431.

- [129] SHI, J., LIU, X., AND ZHANG, N. On uncertainty principle for signal concentrations with fractional fourier transform. *Signal Processing* 92, 12 (2012), 2830–2836.
- [130] STEIN, E. M., AND SHAKARCHI, R. *Fourier analysis*, vol. 1 of *Princeton Lectures in Analysis*. Princeton University Press, Princeton, NJ, 2003. An introduction.
- [131] SZEGÖ, G. *Orthogonal polynomials*, fourth ed., vol. Vol. XXIII of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 1975.
- [132] TENPAS, N. A Family of Optimal Configurations for Periodic Energy. *arXiv:2311.05594* (2023).
- [133] TER MORSCHÉ, H. G., AND OONINCX, P. J. On the Integral Representations for Metaplectic Operators. *J. Fourier Anal. Appl.* 8 (2002), 245–258.
- [134] TSCHURTSCHENTHALER, T. *The Gabor Frame Operator (its Structure and Numerical Consequences)*. Master’s thesis, University of Vienna, Vienna, Austria, 2000.
- [135] WEIL, A. Sur certains groupes d’opérateurs unitaires. *Acta Math.*, 111 (1964), 143–211.
- [136] WELLERSHOFF, M. Sampling at twice the Nyquist rate in two frequency bins guarantees uniqueness in Gabor phase retrieval. *J. Fourier Anal. Appl.* 29, 1 (2023), Paper No. 7, 9.
- [137] WEXLER, J., AND RAZ, S. Discrete Gabor expansions. *Signal Process.* 21, 3 (1990), 207–220.
- [138] WIDDER, D. V. *The Laplace Transform*. Princeton University Press, 1941.
- [139] WIENER, N. Tauberian Theorems. *Ann. of Math.* 33, 1 (1932), 1–100.
- [140] WIESMEYR, C. *Construction of frames by discretization of phase space*. PhD thesis, University of Vienna, Vienna, Austria, 2013.
- [141] WILCZOK, E. New uncertainty principles for the continuous Gabor transform and the continuous wavelet transform. *Doc. Math.* 5 (2000), 201–226.
- [142] WOODWARD, P. M. *Probability and information theory, with applications to radar*, second ed. Pergamon Press, Oxford-Edinburgh-New York-Paris-Frankfurt, 1964.
- [143] YOUNG, R. *An Introduction to Non-Harmonic Fourier Series*, revised ed. Academic Press, 2001.

- [144] ZALIK, R. A. On approximations by shifts and a theorem of Wiener. *Trans. Amer. Math. Soc.* 243 (1978).
- [145] ZHANG, F., CHEN, B., MORRISON, G. R., VILA-COMAMALA, J., GUIZAR-SICAIROS, M., AND ROBINSON, I. K. Phase retrieval by coherent modulation imaging. *Nat. Commun.* 7, 1 (2016), 13367.
- [146] ZHANG, Z. Uncertainty principle of complex-valued functions in specific free metaplectic transformation domains. *J. Fourier Anal. Appl.* 27, 4 (2021), Paper No. 68, 32.
- [147] ZHANG, Z., ZHU, Z., LI, D., AND HE, Y. Free metaplectic Wigner distribution: definition and Heisenberg’s uncertainty principles. *IEEE Trans. Inform. Theory* 69, 10 (2023), 6787–6810.
- [148] ZHANG, Z.-C. Uncertainty principle for linear canonical transform using matrix decomposition of absolute spread matrix. *Digit. Signal Process.* 89 (2019), 145–154.

3 Publications

3.1 Sampling theorems with derivatives in shift-invariant spaces generated by periodic exponential B-splines

Bibliographic Information

Gröchenig, K., Shafkulovska, I., Sampling theorems with derivatives in shift-invariant spaces generated by periodic exponential B-splines, *submitted* (2023)

Submission date

29 November 2023

Contribution

Karlheinz Gröchenig and Irina Shafkulovska contributed equally to the idea, concept, development, and writing of the paper.

E-Print

arXiv:2311.08352

SAMPLING THEOREMS WITH DERIVATIVES IN SHIFT-INVARIANT SPACES GENERATED BY PERIODIC EXPONENTIAL B-SPLINES

KARLHEINZ GRÖCHENIG * AND IRINA SHAFKULOVSKA *

ABSTRACT. We derive sufficient conditions for sampling with derivatives in shift-invariant spaces generated by a periodic exponential B-spline. The sufficient conditions are expressed with a new notion of measuring the gap between consecutive samples. These conditions are near optimal, and, in particular, they imply the existence of sampling sets with lower Beurling density arbitrarily close to the necessary density.

1. INTRODUCTION

By a sampling problem, we understand the question of whether and how a function in a given space can be recovered from discrete information, which is usually given by point evaluations (samples). The given space is an a priori signal model, and for a century this model has been the Paley-Wiener space of bandlimited functions. The significance of the Paley-Wiener space and the associated Shannon-Whittaker-Kotelnikov sampling theorem cannot be overestimated; it is at the origin of the information theory by Shannon, it is the technical basis for analog/digital conversion in signal processing, and it informs a whole branch of complex analysis. From the point of view of mathematics, the decisive results are the necessary density conditions of Landau [26] and the sufficient conditions for sampling by Beurling [10]. Their work remains the point of reference for the exploration of sampling problems in analysis and signal processing. See [4, 8, 35] for some representative expositions of modern sampling theory.

In this paper, we study sampling in a shift-invariant space generated by a periodic exponential B-spline (PEB-spline). Given a PEB-spline, or more generally a generator φ subject to some mild decay conditions, the shift-invariant space generated by φ in $L^p(\mathbb{R})$ is defined to be

$$V^p(\varphi) := \left\{ f \in L^p(\mathbb{R}) : f(x) = \sum_{\ell \in \mathbb{Z}} c_\ell \varphi(x - \ell), (c_\ell)_{\ell \in \mathbb{Z}} \in \ell^p(\mathbb{Z}) \right\} \subseteq L^p(\mathbb{R}).$$

*FACULTY OF MATHEMATICS, UNIVERSITY OF VIENNA, OSKAR-MORGENSTERN-PLATZ 1, 1090 VIENNA, AUSTRIA

E-mail addresses: `karlheinz.groechenig@univie.ac.at`, `irina.shafkulovska@univie.ac.at`.

2010 *Mathematics Subject Classification.* 42C15, 41A15, 94A20, 42C40.

Key words and phrases. Shift-invariant spaces, splines, Chebyshev B-splines, collocation matrix, Schoenberg-Whitney condition, nonuniform sampling, Gabor frames.

K. Gröchenig was supported by the Austrian Science Fund (FWF) project P31887-N32. I. Shafkulovska was funded by the Austrian Science Fund (FWF) project P33217 and by the Austrian Academy of Sciences through an internship at the Acoustics Research Institute.

If $\varphi(x) = \frac{\sin \pi x}{\pi x}$, then $V^2(\varphi) = \{f \in L^2(\mathbb{R}) : \text{supp } \hat{f} \subseteq [-1/2, 1/2]\}$ is the Paley-Wiener space $PW^2(\mathbb{R})$ of band-limited functions.

Shift-invariant spaces form a natural class of *signal models* in sampling theory, as they generalize the Paley-Wiener space and are useful for practical and numerical reasons. The generator φ can be considered a parameter of the signal model that can be tuned to the specifics of the data acquisition. For instance, if φ is compactly supported, then every point evaluation can be computed as a finite sum, and a sample $f(x)$ affects only the coefficients c_k in a neighbourhood of x . Additional local properties of $f \in V^p(\varphi)$, such as differentiability and analyticity, can be implemented by imposing them on the generator φ . By contrast, the Paley-Wiener space consists of entire functions, and every sample of f has a long-range effect on the values of f . Shift-invariant spaces are central in approximation theory [13, 22, 25, 33] and are increasingly used in signal processing.

For the mathematical formulation of the sampling problem, we assume that the signal space $V^p(\varphi)$ is known and that $f \in V^p(\varphi)$ is sampled on a set X together with the first derivatives $f^{(s)}(x)$, $s = 0, \dots, \mu(x)$, for some multiplicity function μ . Whereas the standard sampling problem only takes samples on a discrete subset of $\mathbb{R}$, the inclusion of derivatives has become important in several applications. First derivatives model trends, and higher derivatives indicate convexity properties. In higher dimensional spaces, i.e., for images or physical fields, one speaks of “gradient-augmented measurements” [2], in event-based sampling “information is transmitted only when a significant change in the signal occurs, justifying the acquisition of a new sample” [29], in other words, the value of the first derivative determines the position of the next sample.

For the Paley-Wiener space, sampling with derivatives was studied early on [16, 30, 31], the problem was taken up again recently with an emphasis on shift-invariant spaces in [1, 15, 17, 36].

The main objective is to recover a function from its samples and derivatives on a set $X = (x_j)_{j \in \mathbb{Z}}$. We indicate the number of derivatives at every point with a *multiplicity function* $\mu_X : X \rightarrow \{0, \dots, S\}$. The goal is to derive a sampling inequality of type

$$(1.1) \quad A_p \|f\|_p^p \leq \sum_{x \in X} \sum_{s=0}^{\mu_X(x)} |f^{(s)}(x)|^p \leq B_p \|f\|_p^p \quad \text{for all } f \in V^p(\varphi).$$

Note that the number of derivatives depends on the sampling point x . If (1.1) is satisfied, we call (X, μ_X) a sampling set (or a set of stable sampling) for $V^p(\varphi)$. Once a sampling inequality is available, one can use standard reconstruction algorithms from frame theory to reconstruct f from (X, μ_X) .

Our final choice is a class of *generators* for the shift-invariant space. We suggest the use of PEB-splines as suitable and practical generators. These have compact support, they are well understood in approximation theory [12, 34] and in numerics. In sampling theory, they lead to $\mathcal{O}(N)$ algorithms, when N is the number of available data [18]. Most importantly and in contrast to other classes of generators, for PEB-spline generators one can prove optimal results. This is precisely the objective of this paper.

Before proving anything at all, one should understand what is possible. How much information is necessary to recover a function from its samples? This question is answered by a necessary density condition. For shift-invariant spaces the relevant quantity is the weighted Beurling density defined as

$$D^-(X, \mu_X) := \liminf_{R \rightarrow \infty} \inf_{y \in \mathbb{R}} \frac{1}{2R} \sum_{x \in X \cap [y-R, y+R]} (1 + \mu_X(x)).$$

If $\mu \equiv 0$, then $D^-(X, \mu_X) = D^-(X)$ is just the usual lower Beurling density. The weighted Beurling density counts the average number of data per unit interval. The necessary density condition, first proved by Landau [27] for bandlimited functions in full generality (with arbitrary spectra and in higher dimensions), turns out to be a universal information-theoretic bound for sampling. In the context of sampling with derivatives in shift-invariant spaces it can be expressed as follows [21, Prop. 3.7.].

Theorem 1.1. *Assume that μ_X is bounded by M and $\varphi, \varphi', \dots, \varphi^{(M)}$ satisfies a mild decay condition, e.g., $|\varphi^s(x)| \leq C(1 + |x|)^{-1-\varepsilon}$ for some $\varepsilon > 0$.*

If (X, μ) is a sampling set for $V^p(\varphi)$, then

$$(1.2) \quad D^-(X, \mu_X) \geq 1.$$

The problem of sufficient conditions that guarantee the reconstruction of $f \in V^p(\varphi)$ from its samples (X, μ_X) is more interesting and also more difficult. Under minimal conditions on the smoothness and decay of φ one can always show with elementary tools that any sufficiently dense set is sampling [3]. For some types of generators, the density condition (1.2) is (almost) a characterization. If φ is either the cardinal sine $\frac{\sin \pi x}{\pi x}$, whence $V^2(\varphi) = PW$, or the Gaussian or some related function, then $D^-(X, \mu) > 1$ is a sufficient condition for sampling [10, 21]. For other generators, (1.2) may not provide a characterization, in particular, if φ has compact support, then the maximal gap between sampling points is limited by the size of $\text{supp } \varphi$, whereas for given Beurling density large gaps are permitted. Therefore one often uses different notions to measure the density of a set.

Contributions. For PEB-splines as generators, we will derive a set of conditions that allow the construction of sampling sets arbitrarily close to the necessary condition.

Our first theorem is somewhat technical because the sufficient conditions are expressed in terms of local Schoenberg-Whitney conditions that are well-known in spline theory. These conditions come close to a characterization of sampling sets with derivatives but are not complete. We postpone the precise formulation to Section 4, and state the most important consequence.

Theorem 1.2. *Let φ be a PEB-spline of order m with support in $[0, m]$. For every $\varepsilon > 0$ and given multiplicity sequence $\mu = (\mu_j)_{j \in \mathbb{Z}}$ with values in $\{0, \dots, m-1\}$, there exists a set $X = (x_j)_{j \in \mathbb{Z}}$, such that (X, μ_X) is sampling for $V^p(\varphi)$, $\mu_X(x_j) = \mu_j$ for all $j \in \mathbb{Z}$, and $D^-(X, \mu_X) < 1 + \varepsilon$.*

Our second result uses the maximum gap between consecutive samples, also called the mesh width or the covering density. The maximum gap of a set $X = (x_j)_{j \in \mathbb{Z}} \subseteq \mathbb{R}$ consisting

of consecutive samples is defined as

$$\mathbf{mg}(X) = \sup_{j \in \mathbb{Z}} (x_{j+1} - x_j),$$

and is used frequently to formulate conditions for sampling, e.g., in [3]. Our new contribution is a weighted maximum gap that takes into consideration the number of derivatives at each point. It is defined as

$$\mathbf{mg}(X, \mu_X) = \max \left\{ \sup_{j \in \mathbb{Z}} \frac{x_{j+1} - x_j}{1 + \mu_X(x_j)}, \sup_{j \in \mathbb{Z}} \frac{x_{j+1} - x_j}{1 + \mu_X(x_{j+1})} \right\}$$

The rationale for this definition is that several derivatives at a point x , say $f^{(s)}(x)$, $s = 0, \dots, \mu_X(x)$, amount to $\mu_X(x) + 1$ data, so the next data point should be at distance $< \mu_X(x) + 1$ from x , in other words, $\frac{x_{j+1} - x_j}{\mu_X(x) + 1} < 1$. The precise formulation is as follows.

Theorem 1.3 (Maximum Gap Theorem). *Let φ be a PEB-spline of order m and let $X \subseteq \mathbb{R}$ be a separated set with multiplicity function $\mu_X : X \rightarrow \{0, \dots, m - 2\}$. If the weighted maximum gap of (X, μ_X) satisfies*

$$(1.3) \quad \mathbf{mg}(X, \mu_X) < 1,$$

then (X, μ_X) is a sampling set for $V^p(\varphi)$.

In contrast to the more technical theorem, the maximum gap condition is easy to understand and to verify.

As a special case, we mention sampling with the same number of derivatives at every point.

Corollary 1.4. *Let φ be a PEB-spline of order $m \geq 2$ and let $X \subseteq \mathbb{R}$ be a separated set. Assume that μ_X is constant, $\mu_X(x) = s < m - 1$ for all $x \in X$. If*

$$\mathbf{mg}(X) < s + 1,$$

then (X, μ_X) is a sampling set for $V^p(\varphi)$, $1 \leq p \leq \infty$.

In particular, a lattice $X = \alpha\mathbb{Z}$ is a sampling set for $V^p(\varphi)$, $1 \leq p \leq \infty$, if and only if $\alpha < s + 1$.

When using $\mathbf{mg}(X, \mu_X)$ as the relevant density and PEB-splines as generators, then the weighted gap condition (1.3) is best possible in the following sense. One can show that for $a > 1$ and given multiplicity sequence $(\mu_j)_{j \in \mathbb{Z}}$ there exist sets $X = (x_j)$, such that $\mathbf{mg}(X, \mu) = a$, but (X, μ) is not a sampling set.

It is easy to see that $D^-(X, \mu_X) \geq \mathbf{mg}(X, \mu_X)^{-1}$. This explains why the maximum gap condition is often problematic. So far there is only a handful of generators [4, 14, 20, 21] for which the condition $\mathbf{mg}(X) < 1$ or $\mathbf{mg}(X, \mu_X) < 1$ has been proved to be sufficient for sampling. In a surprisingly large number of publications, a sufficient condition $\mathbf{mg}(X) \leq \delta_0$ has been derived for unknown or small δ_0 . Since the comparison to the necessary condition is missing, more research is required to understand the relevance of such results.

In the last section, we exploit a connection between sampling in shift-invariant spaces and the theory of Gabor frames and derive some new results about Gabor frames with PEB windows.

Methods. Following [4] we partition $\mathbb{R}$ into intervals of finite length and then study the local problem. Since a PEB-spline of order $m \in \mathbb{N}$ has support in the interval $[0, m]$, the restriction of a function f in $V^p(\varphi)$ to an interval $[a, b]$, $a, b \in \mathbb{Z}$, sees only the finite sum

$$f(x) = \sum_{\ell=a-m+1}^{b-1} c_\ell \varphi(x - \ell) \quad \text{for } x \in [a, b].$$

This is then a finite-dimensional problem, and the intervening matrix is the *collocation matrix* associated to the PEB-spline φ . Its invertibility is well understood and can be expressed with the help of the Schoenberg-Whitney conditions on the sampling points in $[a, b]$. The characterization of invertible collocation matrices is one of the cornerstones of spline theory [34], and our proofs pay homage to this fundamental result. We develop two approaches centred around the collocation matrix. In one approach we force uniform bounds on the collocation matrix and then glue together the local reconstructions. In an alternative approach, we use Beurling's technique of weak limits to arrive at a sampling theorem.

The paper is organized as follows. Section 2 introduces shift-invariant spaces and sampling sets and prepares some technical tools for sampling with derivatives. Section 3 is dedicated to the basic theory of ECC-systems and B-splines with a special focus on collocation matrices for sampling with derivatives. The first main result and the existence of sampling sets near the optimal density are proven in Section 4. Section 5 is devoted to Beurling's technique adapted to sampling in shift-invariant spaces. With this at hand, we prove the main sampling with the maximum gap condition in Section 6. We conclude with Section 7 with a discussion of the two methods and Section 8, where we derive the implications of the two sampling theorems for Gabor systems. A technical result is postponed to the appendix.

2. SHIFT-INVARIANT SPACES AND SAMPLING

We begin with properties of shift-invariant spaces and various formulations of the sampling problem.

2.1. (Vector-valued) shift-invariant spaces. Let us consider a shift-invariant space $V^p(\varphi)$. Since each element $f \in V^p(\varphi)$ is represented as $f = \sum_{\ell \in \mathbb{Z}} c_\ell T_\ell \varphi$ for some sequence $(c_\ell)_{\ell \in \mathbb{Z}} \in \ell^p(\mathbb{Z})$, the desirable generators are those which come with the norm equivalence

$$(2.1) \quad A_p \|c\|_{\ell^p} \leq \left\| \sum_{\ell \in \mathbb{Z}} c_\ell T_\ell \varphi \right\|_{L^p} \leq B_p \|c\|_{\ell^p}, \quad c \in \ell^p(\mathbb{Z}),$$

where the constants $A_p, B_p > 0$ depend only on p . If this holds, we say that φ has *p-stable integer translates (shifts)* and denote (2.1) as $\|f\|_{L^p} \asymp \|c\|_{\ell^p}$. If norm equivalence holds for

all $p \in [1, \infty]$, we omit the reference to p and speak of *stable integer translates* and a *stable generator*.

The following theorem summarizes some well-known conditions for stability.

Theorem 2.1. *Let $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ be a bounded, compactly supported function. Then the following statements are equivalent:*

- (i) *The generator φ has stable translates.*
- (ii) *The generator φ has stable ∞ -translates.*
- (iii) *The translates $\{T_\ell \varphi : \ell \in \mathbb{Z}\}$ are ℓ^∞ -independent, i.e.,*

$$\sum_{\ell \in \mathbb{Z}} c_\ell T_\ell \varphi \neq 0 \quad \text{for all } c \in \ell^\infty(\mathbb{Z}) \setminus \{0\}.$$

- (iv) *The Fourier transform $\hat{\varphi}$ does not have a real 1-periodic zero, i.e.,*

$$0 < \sum_{\ell \in \mathbb{Z}} |\hat{\varphi}(\omega + \ell)|^2 \quad \text{for all } \omega \in \mathbb{R}.$$

We refer to Ron's survey article [33]. The equivalences follow from the results in [23, Thm. 3.5] and [33, Thm. 29]. For further references, the reader can consult [28, Eq. (9)], [23, Thm. 3.3]. For sampling in shift-invariant spaces the stability condition is explained and used in [4, 5] and [20, Sec. 2].

For the special case of a bounded, compactly supported generator φ above with $\text{supp}(\varphi) \subseteq [-L, L]$, the important lower inequality is almost trivial:

$$\|f\|_{L^p}^p \leq \int_{\mathbb{R}} \sum_{\ell \in \mathbb{Z}} |c_\ell T_\ell \varphi(t)|^p dt \leq \sum_{k \in \mathbb{Z}} \int_K^{K+1} \sum_{\ell=K-2L+1}^L |c_\ell|^p \|\varphi\|_\infty^p dt = 4L \|\varphi\|_\infty^p \|c\|_p^p,$$

if $p < \infty$, with the analogous estimate for $p = \infty$.

Following [21, Sec. 2], we further consider vector-valued shift-invariant spaces

$$V^p(\Phi) := \left\{ \sum_{\ell \in \mathbb{Z}} c_\ell T_\ell \Phi : c \in \ell^p(\mathbb{Z}) \right\} \subseteq (L^p(\mathbb{R}))^{S+1}$$

generated by a vector $\Phi = (\Phi^0, \dots, \Phi^S)^t \in (L^1(\mathbb{R}))^{S+1}$ with norm

$$\|F\|_p = \left(\sum_{s=0}^S \|F^s\|_{L^p}^p \right)^{1/p},$$

with the usual modification for $p = \infty$. We will assume that Φ has stable integer shifts, i.e.,

$$(2.2) \quad \left\| \sum_{\ell \in \mathbb{Z}} c_\ell T_\ell \Phi \right\|_p \asymp \|c\|_p \quad \text{for all } c \in \ell^p(\mathbb{Z}).$$

2.2. Sampling sets. To accommodate derivatives and multiplicities in sampling, we consider tuples of sets $\vec{X} = (X^0, \dots, X^S)$ with $X^s \subseteq \mathbb{R}$, $0 \leq s \leq S$. Assume that the generator has stable integer shifts and that the shift-invariant space $V^p(\Phi)$ contains tuples of functions $F = (F_0, \dots, F^S)$ which are pointwise well-defined. In the following, $V^p(\Phi^s)$ contains only continuous or piecewise continuous functions with well-defined one-sided limits at the finitely many discontinuities. We say that $\vec{X}$ is a *sampling set* for $V^p(\Phi)$, $p \in [1, \infty]$, if

$$(2.3) \quad \sum_{s=0}^S \sum_{x \in X^s} |F^s(x)|^p \asymp \|F\|_p^p \quad \text{for all } F \in V^p(\Phi),$$

where the inequality constants depend only on p and Φ (with the usual adaptation for $p = \infty$). A weaker notion is that of a uniqueness set. We call $\vec{X}$ a *uniqueness set* if the linear operator $F \mapsto (F^0|_{X^0}, \dots, F^S|_{X^S})$ is injective on $V^p(\Phi)$, that is,

$$F^s|_{X^s} \equiv 0 \quad \text{for all } s \in \{0, \dots, S\} \quad \Leftrightarrow \quad F \equiv 0.$$

The case $S = 0$ is the standard case of sampling sets. As usual, the upper bound for sampling is easy to obtain in a general setting, whereas, for the lower bound, the specific features of the generator have to be exploited.

Theorem 2.2. *Let φ be a compactly supported, piecewise continuous function with finitely many jump discontinuities and assume that φ has stable integer translates. Then*

- (i) *All functions in $V^p(\varphi)$ are piecewise continuous. Furthermore, the set of discontinuities is separated, consisting only of jump discontinuities.*
- (ii) *The sampling operator $f \rightarrow f|_X$ is bounded on $V^p(\varphi)$. Here the samples at discontinuities are to be understood as samples of the right-sided limits.*

The analogous statement holds with left-sided limits.

Proof. (i) A series $\sum_{\ell \in \mathbb{Z}} c_\ell T_\ell \varphi$ is locally finite, so it inherits the properties of the generator.

Concerning (ii), we refer the reader to [5, Thm. 3.1(iii)] for the proof ¹. $\square$

To relate sampling in vector-valued spaces to sampling with derivatives with a multiplicity function $\mu_X : X \rightarrow \{0, \dots, S\}$, we set $\Phi := (\varphi, \varphi^{(1)}, \dots, \varphi^{(S)})$ and

$$X^s := \{x \in X : \mu_X(x) \geq s\}, \quad 1 \leq s \leq S.$$

In this case, the inequality (2.3) for vector-valued functions is the same as the sampling inequality

$$\sum_{x \in X} \sum_{s=0}^{\mu_X(x)} |f^{(s)}(x)|^p \asymp \|f\|_p^p \quad \text{for all } f \in V^p(\varphi).$$

¹Note that the proof does not require the continuity of φ .

3. SOME SPLINE THEORY

In this section, we introduce ECC-systems and Chebyshev B-splines and cite and prove statements about collocation matrices arising in the sampling problem. We follow the terminology in [34].

3.1. ECC-systems and Chebyshev B-splines. We call a system of functions $u_\ell : [a, b] \rightarrow \mathbb{C}$, $1 \leq \ell \leq m$, a *Chebyshev system*, if the collocation matrix with entries

$$(u_\ell(t_i))_{1 \leq i, \ell \leq m}$$

has a strictly positive determinant for all $a \leq t_1 < \dots < t_m \leq b$. It follows from the definition that a Chebyshev system is linearly independent. The span of $(u_1, \dots, u_m)$ is called a Chebyshev space.

The definition of a Chebyshev system implies immediately the following property of the zero set. One of the main properties of these spaces is the number of solutions of $f(x) = 0$.

Lemma 3.1. *If $(u_1, \dots, u_m)$ is a Chebyshev system, then either $(c_\ell)_{1 \leq \ell \leq m} = 0$ or*

$$\#\left\{t \in [a, b] : \sum_{\ell=1}^m c_\ell u_\ell(t) = 0\right\} \leq m - 1.$$

Moving further towards splines, we choose positive weight functions $w_s \in C^{m-s}[a, b]$ on an interval $[a, b] \subseteq \mathbb{R}$ and define

$$\begin{aligned} u_1(x) &= w_1(x), \\ u_2(x) &= w_1(x) \int_a^x w_2(x_2) dx_2, \\ &\vdots \\ u_m(x) &= w_1(x) \int_a^x w_2(x_2) \int_a^{x_2} w_3(x_3) \cdots \int_a^{x_{m-1}} w_m(x_m) dx_m \dots dx_2. \end{aligned}$$

These are functions in $\mathbb{C}^{m-1}[a, b]$ and form a so-called *extended complete Chebyshev (ECC) system* on $[a, b]$. This means that for all $1 \leq D \leq m$ and all $t_1 \leq \dots \leq t_D \in [a, b]$ the *collocation matrix of Hermite interpolation*, i.e., the matrix with entries

$$(u_\ell^{(d_i)}(t_i))_{1 \leq i, \ell \leq D},$$

where

$$d_i := \max \{\ell : t_i = \dots = t_{i-\ell}\}, \quad 1 \leq i \leq D,$$

has a positive determinant. This is the matrix that arises in the Hermite interpolation problem

$$f = \sum_{\ell=1}^D c_\ell u_\ell, \quad f^{(d_i)}(t_i) = \xi_i, \quad 1 \leq i \leq D.$$

The associated ECC-space is given by $\mathcal{U}_m := \text{span}_{\mathbb{C}}(u_1, \dots, u_m)$. As the simplest example (and the motivation), we consider $w_1(x) = 1$ and $w_s = s - 1$, $2 \leq s \leq m$. In this case, we obtain the polynomials $u_s = (x - a)^{s-1}$, $1 \leq s \leq m$, i.e., the ECC-space $\mathcal{U}_m = \mathcal{P}_m$ is the space of polynomials of degree at most $m - 1$. One of the particularly useful properties of polynomial spaces is the self-containment with respect to the derivatives, i.e.,

$$p^{(s)} \in \mathcal{P}_{m-s}, \quad 0 \leq s \leq m - 1, \quad p \in \mathcal{P}_m.$$

To preserve this property for ECC-systems, one replaces the usual derivatives with related differential operators. We define $D_0 f = f$ and

$$D_s f := \frac{d}{dx} \frac{f}{w_s}, \quad L_s := D_s D_{s-1} \dots D_0, \quad 0 \leq s \leq m.$$

By Leibniz's rule, L_s is a differential operator of order s with variable coefficients.

Just as with the polynomials and the well-known B-splines, we can now consider Chebyshev splines. For an ECC-system $(u_1, \dots, u_m)$ on $[a, b]$ and knots $a = y_1 < \dots < y_{D+m} = b$, a *spline* with these knots is a function $B \in C^{m-2}(\mathbb{R})$ satisfying

$$B|_{(y_\ell, y_{\ell+1})} \in \mathcal{U}_m, \quad 1 \leq \ell \leq D + m - 1.$$

This condition implies that B is piecewise C^{m-1} and at the knots $B^{(m-1)}$ has jump discontinuities and only one-sided derivatives exist. The integer m is the order of the spline. In this case we define $L_{m-1}B$ at the knots y_i to be the right-sided limit

$$L_{m-1}B(y_i) := \lim_{x \searrow y_i} L_{m-1}B(x).$$

A major result in spline theory is the existence of Chebyshev B-splines (CB-splines) associated to knots $a \leq y_1 < \dots < y_{m+D} \leq b$. Precisely, the CB-splines are the splines $(B_m^\ell)_{1 \leq \ell \leq D}$ with controlled support

$$\text{supp}(B_m^\ell) = [y_\ell, y_{\ell+m}], \quad 1 \leq \ell \leq D.$$

The CB-splines B_m^ℓ are unique up to normalization. The traditional construction is based on divided differences and can be found in full detail in Schumaker's book [34, Sec. 9.4].

We conclude by considering matrices associated to CB-splines. The definition of ECC-systems calls for the collocation matrix with the standard derivatives. We adapt this to the new differential operators L_s . Given $t_1 \leq t_2 \leq \dots \leq t_D$, let

$$d_i := \max \{ \ell : t_i = \dots = t_{i-\ell} \} \leq m - 1, \quad 1 \leq i \leq D.$$

Let $(B_m^1, \dots, B_m^D)$ be CB-splines associated with the knots $y_1 < y_2 < \dots < y_{m+D}$. We consider the collocation matrix

$$(3.1) \quad (L_{d_i} B_m^\ell(t_i))_{1 \leq i, \ell \leq D}.$$

This is the matrix tied to the Hermite interpolation problem

$$f = \sum_{\ell=1}^D c_\ell B_m^\ell, \quad L_{d_i} f(t_i) = \zeta_i, \quad 1 \leq i \leq D.$$

We can determine exactly when this matrix is invertible [34, Thm. 9.33], cf. [34, Thm. 4.67].

Theorem 3.2 (Interlacing property, Schoenberg-Whitney conditions). *The determinant of the collocation matrix in (3.1) is nonnegative. It is positive if and only if*

$$t_i \in \begin{cases} (y_i, y_{i+m}), & \text{if } d_i < m-1 \\ [y_i, y_{i+m}), & \text{if } d_i = m-1, \end{cases} \quad 1 \leq i \leq D.$$

The theorem says that it suffices for t_i to be in the support of B_m^i . The invertibility of the collocation matrix is independent of the normalization of the CB-splines, as the normalization corresponds to multiplying the collocation matrix with an invertible diagonal matrix.

Example 3.3. As a small example, we consider CB-splines of order 5. Let the knots be $0, 1, \dots, 10$. Then the splines $B_5^0, \dots, B_5^5$ are supported on

$$(0, 5), \quad (1, 6), \quad (2, 7), \quad (3, 8), \quad (4, 9), \quad \text{and} \quad (5, 10),$$

respectively. If we choose $t_1 = t_2 = 0.5$, $t_3 = 4$, $t_4 = t_5 = t_6 = 7.7$, then the corresponding multiplicities are

$$d = (0 \quad 1 \quad 0 \quad 0 \quad 1 \quad 2).$$

The matrix is not invertible because $t_2 = 0.5$ is not in $(1, 6)$. Visually, the issue is clear:

$$\begin{pmatrix} B_5^0(0.5) & B_5^1(0.5) & B_5^2(0.5) & B_5^3(0.5) & B_5^4(0.5) & B_5^5(0.5) \\ L_1 B_5^0(0.5) & L_1 B_5^1(0.5) & L_1 B_5^2(0.5) & L_1 B_5^3(0.5) & L_1 B_5^4(0.5) & L_1 B_5^5(0.5) \\ B_5^0(4) & B_5^1(4) & B_5^2(4) & B_5^3(4) & B_5^4(4) & B_5^5(4) \\ B_5^0(7.7) & B_5^1(7.7) & B_5^2(7.7) & B_5^3(7.7) & B_5^4(7.7) & B_5^5(7.7) \\ L_1 B_5^0(7.7) & L_1 B_5^1(7.7) & L_1 B_5^2(7.7) & L_1 B_5^3(7.7) & L_1 B_5^4(7.7) & L_1 B_5^5(7.7) \\ L_2 B_5^0(7.7) & L_2 B_5^1(7.7) & L_2 B_5^2(7.7) & L_2 B_5^3(7.7) & L_2 B_5^4(7.7) & L_2 B_5^5(7.7) \end{pmatrix}$$

$$= \begin{pmatrix} B_5^0(0.5) & 0 & 0 & 0 & 0 & 0 \\ L_1 B_5^0(0.5) & 0 & 0 & 0 & 0 & 0 \\ B_5^0(4) & B_5^1(4) & B_5^2(4) & B_5^3(4) & 0 & 0 \\ 0 & 0 & 0 & B_5^3(7.7) & B_5^4(7.7) & B_5^5(7.7) \\ 0 & 0 & 0 & L_1 B_5^3(7.7) & L_1 B_5^4(7.7) & L_1 B_5^5(7.7) \\ 0 & 0 & 0 & L_2 B_5^3(7.7) & L_2 B_5^4(7.7) & L_2 B_5^5(7.7) \end{pmatrix}$$

$$= \begin{pmatrix} * & 0 & 0 & 0 & 0 & 0 \\ * & 0 & 0 & 0 & 0 & 0 \\ * & * & * & * & 0 & 0 \\ 0 & 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * & * \end{pmatrix}$$

If instead we sample at $t_1 = t_2 = 1.5 \in (1, 5)$, $t_3 = 4$, $t_4 = t_5 = t_6 = 7.7$, all points are positioned correctly: 4 is in $(2, 7)$ and 7.7 lies in $(3, 8)$, $(4, 9)$ and $(5, 10)$. In this case, the collocation matrix is invertible.

3.2. Periodic exponential B-splines. We have now gathered all the necessary background on CB-splines and continue with the special case of CB-splines which are shifts of a single CB-spline. For this, we require the following structure:

(S1) the knots are the integers $y_i = i - 1$, $i \in \mathbb{Z}$,²

(S2) the weights w_s , $1 \leq s \leq m$ are given by

$$w_s(x) = e^{\gamma_s x} r_s(x), \quad \gamma_s \in \mathbb{R}, \quad T_\ell r_s = r_s > 0 \quad \text{for all } \ell \in \mathbb{Z}.$$

CB-splines in this form are called *periodic exponential B-splines* (PEB-splines) introduced in [7, 32]. If $r_s \equiv 1$ for all $1 \leq s \leq m$, then we refer to them as *exponential B-splines* (EB-splines). The benefit of (S1) and (S2) is the fact that the CB-spline associated with the knots $y_\ell, \dots, y_{\ell+m}$ is (up to scaling) uniquely determined by any other of the B-splines, e.g., by $\varphi := B_m^1$. It is clear that any shift $T_\ell \varphi$ is a Chebyshev spline with

$$(3.2) \quad \text{supp}(T_\ell \varphi) = [0 + \ell, m + \ell], \quad \ell \in \mathbb{Z}.$$

By the uniqueness, $(T_\ell \varphi)_{\ell \in \mathbb{Z}}$ differs up to positive factors $(\theta_\ell)_{\ell \in \mathbb{Z}}$ from the classically constructed CB-splines with divided differences $(B_m^\ell)_{\ell \in \mathbb{Z}}$ [34, Sec. 9.4]. The invertibility of the collocation matrix in Theorem 3.2 is now tied to the integers. We take a closer look at the associated differential operators. For ease of reading, we define for all $1 \leq s \leq m$ the functions

$$q_s(x) := \frac{1}{r_s(x)}, \quad \text{so that} \quad w_s^{-1}(x) = e^{-\gamma_s x} q_s(x).$$

Since both q_s and w_s^{-1} are strictly positive functions, they are in the same differentiability class. Moreover, q_s is periodic. A simple calculation shows

$$\begin{aligned} D_s f(x) &= \frac{d}{dx} [e^{-\gamma_s x} q_s(x) \cdot f(x)] \\ (3.3) \quad &= -\gamma_s e^{-\gamma_s x} q_s(x) \cdot f(x) + e^{-\gamma_s x} q_s'(x) \cdot f(x) + e^{-\gamma_s x} q_s(x) \cdot f'(x) \\ &= \frac{1}{w_s(x)} \left(\left(-\gamma_s + \frac{q_s'(x)}{q_s(x)} \right) \cdot \text{id} + \frac{d}{dx} \right) f(x) \end{aligned}$$

²The choice $y_i = i - 1$ instead of $y_i = i$ is for the support condition (3.2) and consistency of Theorem 3.2 with later applications.

and

$$\begin{aligned} D_s T_\ell f(x) &= \frac{d}{dx} \left[e^{-\gamma_s x} q_s(x) f(x - \ell) \right] \\ &= \frac{d}{dx} \left[e^{-\gamma_s \ell} \cdot e^{-\gamma_s (x - \ell)} q_s(x - \ell) f(x - \ell) \right] \\ &= e^{-\gamma_s \ell} T_\ell D_s f(x), \quad \ell \in \mathbb{Z}. \end{aligned}$$

We obtain the commutator rule

$$L_s T_\ell f = e^{-\ell \sum_{k=1}^s \gamma_k} T_\ell L_s f.$$

The differential operators L_s are natural in the context of ECC-systems, but in practice, information usually comes as samples of standard derivatives. Therefore, in the following statements we derive the relation between the standard derivatives and the differential operators associated with an ECC-system.

Lemma 3.4. *Let $r_s \in C^{m-s}(\mathbb{R})$, $1 \leq s \leq m$, be 1-periodic, strictly positive functions and $\gamma_1, \dots, \gamma_m \in \mathbb{R}$. Let $L_1, \dots, L_m$ be the differential operators associated with the ECC-system induced by*

$$w_s(x) = e^{\gamma_s x} r_s(x), \quad x \in \mathbb{R}.$$

Set for $0 \leq s \leq m$

$$v_s(x) = \prod_{k=1}^{s-1} \frac{1}{w_k(x)}, \quad x \in \mathbb{R}.$$

For each $0 \leq s \leq m$ there exist periodic functions $p_{s,k} \in C^{m-1-s}(\mathbb{R})$ satisfying

$$L_s f(x) = v_s(x) \cdot \left(f^{(s)}(x) + \sum_{k=0}^{s-1} p_{s,k}(x) \cdot f^{(k)}(x) \right).$$

In a matrix notation, this can be written as

$$(L_0, L_1, L_2, \dots, L_s)^t = \mathcal{V}_s \cdot \mathcal{P}_s \cdot \left(\text{id}, \frac{d}{dx}, \frac{d^2}{dx^2}, \dots, \frac{d^s}{dx^s} \right)^t,$$

where $\mathcal{V}_s : \mathbb{R} \rightarrow \mathbb{R}^{(s+1) \times (s+1)}$ is the diagonal matrix

$$\mathcal{V}_s(x) = \text{diag}(v_0(x), \dots, v_s(x))$$

and $\mathcal{P}_s : \mathbb{R} \rightarrow \mathbb{R}^{(s+1) \times (s+1)}$ is the periodic lower triangular matrix

$$\mathcal{P}_s(x) = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ p_{1,0}(x) & 1 & 0 & \cdots & 0 \\ p_{2,0}(x) & p_{2,1}(x) & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{s,0}(x) & p_{s,1}(x) & \cdots & p_{s,s-1}(x) & 1 \end{pmatrix}, \quad x \in \mathbb{R}.$$

In particular, $\mathcal{V}_s(x)$ and $\mathcal{P}_s(x)$ are invertible for all $x \in \mathbb{R}$ and are continuous on $\mathbb{R}$.

Proof. We prove this formula by induction. Clearly, $f^{(0)} = L_0 f$. For the induction step we use (3.3) and $w_k^{-1}(x) = e^{-\gamma_k x} q_k(x)$.

$$\begin{aligned}
L_{s+1}f &= D_{s+1}L_s f \stackrel{(3.3)}{=} \frac{1}{w_{s+1}} \left(\left(-\gamma_{s+1} + \frac{q'_{s+1}}{q_{s+1}} \right) \cdot L_s f + (L_s f)' \right) \\
&\stackrel{\text{Induction step}}{=} \frac{1}{w_{s+1}} \left\{ \left(-\gamma_{s+1} + \frac{q'_{s+1}}{q_{s+1}} \right) \cdot v_s \cdot \sum_{k=0}^s p_{s,k} \cdot f^{(k)} \right. \\
&\quad \left. + v_s \cdot \left(\sum_{k=0}^s p_{s,k} \cdot f^{(k)} \right)' + v'_s \cdot \sum_{k=0}^s p_{s,k} \cdot f^{(k)} \right\} \\
&= \frac{1}{w_{s+1}} \left(-\gamma_{s+1} + \frac{q'_{s+1}}{q_{s+1}} \right) \cdot v_s \cdot \sum_{k=0}^s p_{s,k} \cdot f^{(k)} \\
&\quad + \frac{v_s}{w_{s+1}} \cdot \sum_{k=0}^s p'_{s,k} \cdot f^{(k)} + \frac{v_s}{w_{s+1}} \cdot \sum_{k=0}^s p_{s,k} \cdot f^{(k+1)} \\
&\quad + \frac{v_s}{w_{s+1}} \cdot \sum_{n=1}^s \left(-\gamma_n + \frac{q'_n}{q_n} \right) \cdot \sum_{k=0}^s p_{s,k} \cdot f^{(k)},
\end{aligned}$$

where we used

$$v'_s = \sum_{n=1}^s w_n \cdot v_s \cdot (w_n^{-1})' = \sum_{n=1}^s w_n \cdot v_s \cdot w_n^{-1} \left(-\gamma_n + \frac{q'_n}{q_n} \right) = v_s \cdot \sum_{n=1}^s \left(-\gamma_n + \frac{q'_n}{q_n} \right).$$

for the last equality. We recall that $v_{s+1} = \frac{v_s}{w_{s+1}}$ and $p_{s,s} = 1$ by the induction hypothesis. We set $p_{s,-1} := 0$ to obtain

$$\begin{aligned}
L_{s+1}f &= v_{s+1} \cdot \left(f^{(s+1)} + \sum_{k=0}^s \left(\left(-\gamma_{s+1} + \frac{q'_{s+1}}{q_{s+1}} \right) \cdot p_{s,k} + p'_{s,k} + p_{s,k-1} \right) \cdot f^{(k)} \right) \\
&\quad + v_{s+1} \sum_{n=1}^s \left(-\gamma_n + \frac{q'_n}{q_n} \right) \cdot \sum_{k=0}^s p_{s,k} \cdot f^{(k)} \\
&= v_{s+1} \cdot f^{(s+1)} + v_{s+1} \cdot \underbrace{\sum_{k=0}^s \left(p'_{s,k} + p_{s,k-1} + p_{s,k} \sum_{n=1}^{s+1} \left(-\gamma_n + \frac{q'_n}{q_n} \right) \right)}_{=: p_{s+1,k}} \cdot f^{(k)},
\end{aligned}$$

concluding the proof. $\square$

Lemma 3.5. *Let the knots $y_i = i - 1, i = 1, \dots, D$ and the weights $w_s, 1 \leq s \leq m$ be given and let $B_m^1, \dots, B_m^D = T_{D-1} B_m^1$ be the associated PEB-splines. Then the determinant of*

the collocation matrix

$$M \begin{pmatrix} t_1, \dots, t_D \\ B_m^1, \dots, B_m^D \end{pmatrix} := \left((B_m^\ell)^{(d_i)}(t_i) \right)_{1 \leq i, \ell \leq D} = \left(T_{\ell-1}(B_m^1)^{(d_i)}(t_i) \right)_{1 \leq i, \ell \leq D}$$

is nonnegative. It is positive if and only if

$$t_i \in \begin{cases} (i-1, i+m-1), & d_i < m-1 \\ [i-1, i+m-1), & d_i = m-1, \end{cases} \quad 1 \leq i \leq D.$$

Proof. We first address the second equality in the definition of the matrix M . As already noted in (3.2), the PEB-splines for equispaced knots are shifts of a single PEB-spline, namely, $B_m^\ell = T_{\ell-1}B_m^1$. Since differentiation and translation operators commute, we have $(B_m^\ell)^{(d_i)}(t_i) = T_{\ell-1}(B_m^1)^{(d_i)}(t_i)$.

Let as before d_i count the previous occurrences of the point t_i , i.e.,

$$d_i = \max \{ \ell : t_i = \dots = t_{i-\ell} \} \leq m-1, \quad 1 \leq i \leq D.$$

Then $t = (t_1, t_2, t_3, \dots, t_D)$ can be rewritten as

$$t = \left(\underbrace{t_1, \dots, t_1}_{d_1+1\text{-times}}, \underbrace{t_{d_1+2}, \dots, t_{d_1+2}}_{d_2+1\text{-times}}, \underbrace{t_{d_1+d_2+3}, \dots, t_{d_1+d_2+3}}_{d_3+1\text{-times}}, \dots, \underbrace{t_D, \dots, t_D}_{d_N+1\text{-times}} \right).$$

We define the block-diagonal matrices

$$\mathbf{P}(t) := \begin{pmatrix} \mathcal{P}_{d_1}(t_1) & 0 & 0 & \cdots & 0 \\ 0 & \mathcal{P}_{d_2}(t_{d_1+2}) & 0 & \cdots & 0 \\ 0 & 0 & \mathcal{P}_{d_3}(t_{d_1+d_2+3}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \mathcal{P}_{d_N}(t_D) \end{pmatrix} \in \mathbb{R}^{D \times D},$$

$$\mathbf{V}(t) := \begin{pmatrix} \mathcal{V}_{d_1}(t_1) & 0 & 0 & \cdots & 0 \\ 0 & \mathcal{V}_{d_2}(t_{d_1+2}) & 0 & \cdots & 0 \\ 0 & 0 & \mathcal{V}_{d_3}(t_{d_1+d_2+3}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \mathcal{V}_{d_N}(t_D) \end{pmatrix} \in \mathbb{R}^{D \times D}.$$

Since each $\mathcal{P}_{d_i}$ has determinant 1, the block matrix $\mathbf{P}(t)$ also has determinant 1. By construction, the matrix $\mathbf{V}(t)$ is a diagonal matrix with strictly positive diagonal entries. By Lemma 3.4, the collocation matrix with the associated differential operators and the collocation matrix with the standard derivatives are related by

$$(L_{d_i} B_m^\ell(t_i))_{1 \leq i, \ell \leq D} = \mathbf{V}(t) \cdot \mathbf{P}(t) \cdot M \begin{pmatrix} t_1, \dots, t_D \\ B_m^1, \dots, B_m^D \end{pmatrix}.$$

The claim follows from the multiplicativity of the determinant and Theorem 3.2 with $y_i = i - 1$. $\square$

We finally relate the collocation matrix to sampling with multiplicities.

Corollary 3.6. *Let φ be a PEB-spline of order m supported on the interval $[0, m]$. Further let $x = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$ be a vector of pairwise distinct sampling points, $x_1 < \dots < x_N$, with multiplicities $\mu = (\mu_1, \dots, \mu_N) \in \{0, \dots, m-1\}^N$. Set $D = \sum_{j=1}^N (1 + \mu_j)$. The Hermite interpolation problem*

$$(3.4) \quad f = \sum_{\ell=\ell_0}^{\ell_0+D-1} c_\ell T_\ell \varphi, \quad f^{(s)}(x_j) = \xi_{j,s}, \quad 0 \leq s \leq \mu_j, \quad 1 \leq j \leq N$$

has a unique solution if and only if

$$(3.5) \quad \begin{aligned} x_j &\in (\mu_j + \ell_0, m + \ell_0) + \sum_{n=1}^{j-1} (1 + \mu_n), & \text{if } \mu_j < m-1, \\ x_j &\in [\mu_j + \ell_0, m + \ell_0) + \sum_{n=1}^{j-1} (1 + \mu_n), & \text{if } \mu_j = m-1. \end{aligned}$$

In this case, the solution depends linearly on the interpolation data.

If the points satisfy (3.5), we say that they satisfy the Schoenberg-Whitney conditions.

Proof. To apply the interlacing property to the interpolation problem (3.4), we set

$$t = \left(\underbrace{x_1, \dots, x_1}_{\mu_1+1\text{-times}}, \underbrace{x_2, \dots, x_2}_{\mu_2+1\text{-times}}, \dots, \underbrace{x_N, \dots, x_N}_{\mu_N+1\text{-times}} \right) \in \mathbb{R}^D,$$

to be precise,

$$t_i = x_j \quad \text{if} \quad \sum_{n=1}^{j-1} (1 + \mu_n) < i \leq \sum_{n=1}^j (1 + \mu_n).$$

Then the vector of multiplicities is given by

$$d = \left(0, 1, \dots, \mu_1, 0, 1, \dots, \mu_2, \dots, 0, 1, \dots, \mu_N \right) \in \{0, \dots, m-1\}^D.$$

Accordingly, we set $\xi \in \mathbb{C}^D$ to be the list of the samples $\xi = (\xi_i)_{i=1, \dots, D}$,

$$\xi = \left(\xi_{1,0}, \xi_{1,1}, \xi_{1,2}, \dots, \xi_{1,\mu_1}, \xi_{2,0}, \xi_{2,1}, \xi_{2,2}, \dots, \xi_{2,\mu_2}, \dots, \xi_{N,0}, \xi_{N,1}, \xi_{N,2}, \dots, \xi_{N,\mu_N} \right).$$

With this notation, the interpolation problem (3.4) can be recast as

$$f = \sum_{\ell=\ell_0}^{\ell_0+D-1} c_\ell T_\ell \varphi, \quad f^{(d_i)}(t_i) = \xi_i, \quad 1 \leq i \leq D.$$

The corresponding collocation matrix is

$$M \begin{pmatrix} t_1, \dots, t_D \\ T_{\ell_0}\varphi, \dots, T_{\ell_0+D-1}\varphi \end{pmatrix} = \left(T_{\ell-1}\varphi^{(d_i)}(t_i - \ell_0) \right)_{1 \leq i, \ell \leq D}.$$

For further reference, we denote it by

$$(3.6) \quad \mathbf{M}(x, \mu, \ell_0) := M \begin{pmatrix} t_1, \dots, t_D \\ T_{\ell_0}\varphi, \dots, T_{\ell_0+D-1}\varphi \end{pmatrix}.$$

By Lemma 3.5, the collocation matrix is invertible, and thus a unique solution of (3.4) exists, if and only if $t_i - \ell_0 \in (i - 1, i + m - 1)$ for all $1 \leq i \leq D$ (with the modification $t_i - \ell_0 \in [i - 1, i + m - 1)$ if $d_i = m - 1$). Expressed with the sampling points x_j , the collocation matrix is invertible if and only if

$$x_j \in \bigcap_{s=0}^{\mu_j} \left((s, s + m) + \ell_0 + \sum_{n=1}^{j-1} (1 + \mu_n) \right) = (\mu_j + \ell_0, m + \ell_0) + \sum_{n=1}^{j-1} (1 + \mu_n).$$

for all $1 \leq j \leq N$. The second condition in (3.5) is the adaptation for $\mu_j = m - 1$. $\square$

3.3. Stability of PEB-Splines. To use PEB-splines as generators of shift-invariant spaces, we need to verify the stability of PEB-splines and their derivatives.

Lemma 3.7. *Let φ be a PEB-spline of order m . Then*

- (i) *The generator φ has stable integer shifts.*
- (ii) *For all $0 \leq s \leq m - 1$, the tuple $(\varphi, \dots, \varphi^{(s)})$ has stable integer shifts.*

Proof. (i) Without loss of generality, we can assume that φ is supported on $[0, m]$. By Theorem 2.1, it suffices to show the ℓ^∞ -independence of $(T_\ell \varphi)_{\ell \in \mathbb{Z}}$. To that end, let $c \in \ell^\infty(\mathbb{Z})$, $c \neq 0$, and assume that $c_{\ell_0} \neq 0$. Since $\text{supp}(T_\ell \varphi) = [\ell, \ell + m]$, the restriction of f to the interval $(\ell_0 + m - 1, \ell_0 + m)$ is given by

$$f = \sum_{\ell=\ell_0}^{\ell_0+m-1} c_\ell T_\ell \varphi.$$

The restriction of f to $(\ell_0 + m - 1, \ell_0 + m)$ clearly belongs to the ECC-system. By Lemma 3.1, it has at most $m - 1$ zeros. In particular, it does not vanish on this interval.

(ii) follows from (i), because the inequality

$$A_p \|c\|_{\ell^p}^p \leq \left\| \sum_{\ell \in \mathbb{Z}} c_\ell T_\ell \varphi \right\|_{L^p}^p \leq \sum_{s=0}^S \left\| \sum_{\ell \in \mathbb{Z}} c_\ell T_\ell \varphi^{(s)} \right\|_{L^p}^p.$$

implies that the generator $\Phi = (\varphi, \varphi', \dots, \varphi^{(S)})$ has stable integer shifts. The upper inequality in (2.2) are obvious. $\square$

3.4. EB-splines. Our initial motivation to consider PEB-splines was the particular sub-class of exponential B-splines. These have already been used in time-frequency analysis in [6, 24]. Whereas general PEB-splines are usually defined recursively via divided differences, EB-splines also have a direct definition by their Fourier transform as follows from [11]. Let $m \in \mathbb{N}$ and let $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{R}^m$. An EB-spline $\mathcal{E}_{m,\alpha} : \mathbb{R} \rightarrow \mathbb{R}$ of order m for the m -tuple α is a function of the form

$$\mathcal{E}_{m,\alpha}(x) := \prod_{s=1}^m {}^* e^{\alpha_s x} \chi_{[0,1)}(x),$$

where $\prod {}^*$ denotes the convolution product. Its Fourier transform is given by

$$(3.7) \quad \widehat{\mathcal{E}}_{m,\alpha}(\omega) = \prod_{s=1}^m \frac{e^{\alpha_s - 2\pi i \omega} - 1}{\alpha_s - 2\pi i \omega}.$$

If $\alpha_1 = \dots = \alpha_m = 0$, then this is a classical B-spline. To establish the connection with the original definition of PEB-splines, we set $\gamma_1 := \alpha_1$ and $\gamma_s = \alpha_s - \alpha_{s-1}$ for $2 \leq s \leq m$. Then the weight functions $w_s(x) = e^{\gamma_s x}$ induce the EB-spline $B_m = \mathcal{E}_{m,\alpha}$. For an EB-spline, the stability follows directly from the factorization (3.7) and Theorem 2.1(iv).

4. A SAMPLING THEOREM VIA UNIFORM BOUNDS ON THE COLLOCATION MATRICES

In this section, we prove a first sampling theorem with derivatives in a shift-invariant space generated by a PEB spline. The proof technique is inspired by [4], and its main point is to establish uniform bounds on a family of collocation matrices. The use of derivatives requires more spline theory and leads to more technicalities.

Our standing assumption is that the generator φ is a periodic exponential B-spline of order m , as described in Subsection 3.2. Without loss of generality, we assume that φ is supported on $[0, m]$. We always assume that X is separated, i.e., $\inf_{x,y \in X, x \neq y} |x-y| = \delta > 0$, and δ is called the separation constant of X .

As in [4], we partition $\mathbb{R}$ into suitable intervals and analyze the sampling problem locally. Given integers $M \in \mathbb{Z}$ and $L \in \mathbb{N}$, we partition $\mathbb{R}$ in intervals

$$\mathbb{R} = \bigcup_{k \in \mathbb{Z}} [M + kL, M + kL + L) = \bigcup_{k \in \mathbb{Z}} I_{M,L}(k).$$

Remark. Let $m \in \mathbb{N}$ and let $\ell_0 \leq x_1 < \dots < x_N \leq \ell_0 + L$ be sampling locations with multiplicities $\mu_1, \dots, \mu_N \in \{0, 1, \dots, m-1\}$ such that

$$\sum_{j=1}^N (1 + \mu_j) = L + m - 1$$

and $\mu_N \leq m - 2$ if $x_N = \ell_0 + L$. Then for all $(c_\ell)_{\ell \in \mathbb{Z}} \in \ell^2(\mathbb{Z})$ we have

$$(4.1) \quad \left(\sum_{\ell \in \mathbb{Z}} c_\ell T_\ell \varphi \right)^{(s_j)}(x_j) = \left(\sum_{\ell=\ell_0-m+1}^{\ell_0+L-1} c_\ell T_\ell \varphi \right)^{(s_j)}(x_j), \quad 0 \leq s_j \leq \mu_j, \quad 1 \leq j \leq N.$$

Since φ vanishes outside of $[0, m]$, (4.1) is clear for all $x_1, \dots, x_{N-1}$. The additional assumption that $\mu_N \leq m - 2$ if $x_N = \ell_0 + L$ implies that the measurements at the last sampling point do not include a point of discontinuity of $T_{\ell_0+L}\varphi^{(m-1)}$. Thus for all $0 \leq s \leq \mu_N$ and all $\ell \geq \ell_0 + L$, $T_\ell\varphi^{(s)}(\ell_0 + L) = 0$. These observations imply (4.1).

To avoid unnecessary banalities, we first settle the easy case of PEB-splines of order 1.

Lemma 4.1. *Let φ be a PEB-spline of order 1 supported on $[0, 1)$. Assume that $X \subseteq \mathbb{R}$ is separated. Then the following are equivalent:*

- (i) *For every $k \in \mathbb{Z}$, there exists a point $x^k \in X \cap [k, k + 1)$.*
- (ii) *(X, μ) is a sampling set for $V^p(\varphi)$ for all $p \in [1, \infty]$.*
- (iii) *(X, μ) is a uniqueness set for $V^p(\varphi)$ for all $p \in [1, \infty]$.*

Proof. The implication (ii) $\Rightarrow$ (iii) is trivial. A PEB-spline of the first order is the cut-off of the continuous weight function, namely $\varphi = w_1 \cdot \chi_{[0,1)}$. Thus, if there exists a $k \in \mathbb{Z}$ with $X \cap [k, k + 1) = \emptyset$, then $T_k\varphi|_X = 0$, although this function is clearly not the zero function. This shows (iii) $\Rightarrow$ (i).

Now assume that (i) holds. The upper sampling bound is satisfied by Theorem 2.2. Since w does not vanish, φ has no zeros on $[0, 1)$ and vanishes outside this interval. Therefore, $\varphi(x^k - \ell) \neq 0$ holds if and only if $k = \ell$. Since w is continuous and is strictly positive on $\mathbb{R}$, it attains a minimum $a > 0$ on the compact interval $[0, 1]$. For all $p \in [1, \infty)$ we obtain

$$\sum_{x \in X} |f(x)|^p \geq \sum_{k \in \mathbb{Z}} |f(x^k)|^p = \sum_{k \in \mathbb{Z}} |c_k \varphi(x^k - k)|^p \geq a^p \sum_{k \in \mathbb{Z}} |c_k|^p,$$

with an analogous estimate for $p = \infty$. This shows (i) $\Rightarrow$ (ii) and completes the proof. $\square$

From now on, we assume without loss of generality that $m \geq 2$. Our first theorem looks a bit technical, but is definitely useful.

Theorem 4.2. *Let φ be a PEB-spline of order $m \in \mathbb{N}$ supported on $[0, m]$. Assume $X \subseteq \mathbb{R}$ with multiplicity function μ_X satisfies the following properties:*

- (i) *X is separated.*
- (ii) *$S := \max_{x \in X} \mu_X(x) \leq m - 1$.*
- (iii) *There exist integers $M, L > 0$ and $\varepsilon > 0$, such that for every $k \in \mathbb{Z}$, there exist points $x_1^k < x_2^k < \dots < x_{N(k)}^k$ in $X \cap I_{M,L}(k)$ and nonnegative integers $\mu_1^k, \mu_2^k, \dots, \mu_{N(k)}^k$ with the following three properties:*

$$(4.2) \quad \mu_j^k \leq \mu_X(x_j^k) \quad \text{for all } k \in \mathbb{Z}, 1 \leq j \leq N(k),$$

$$(4.3) \quad \sum_{j=1}^{N(k)} (1 + \mu_j^k) = L + m - 1 \quad \text{for all } k \in \mathbb{Z},$$

$$(4.4) \quad x_j^k \in X \cap \left([\mu_j^k - m + \varepsilon, -\varepsilon] + M + kL + 1 + \sum_{n=1}^{j-1} (1 + \mu_n^k) \right).$$

Then (X, μ_X) is a sampling set for $V^p(\varphi)$ for all $p \in [1, \infty]$.

Remark. The conditions (4.2) - (4.4) are compatible with (4.1) and in some sense are necessary. Since the restriction of f to an interval $I_{M,L}(k)$ belongs to an $(L + m - 1)$ -dimensional space, we need at least $L + m - 1$ samples, whence (4.3). If more information is available, we either remove points or the highest derivatives, which is (ii) and (4.3). As we intend to use the invertibility of the collocation matrix (Corollary 3.6), the retained information must be chosen diligently and involves a Schoenberg-Whitney condition of type (4.4), which says that the support of each shift must contain a sampling point (see Figure 1).

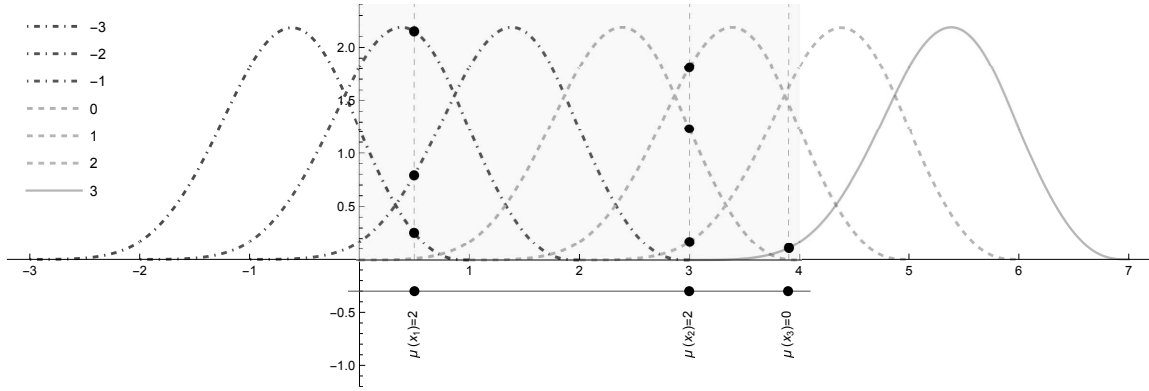


Figure 1. Nonvanishing shifts of $\varphi(x) = \prod_{j=1}^* 4 e^{x\chi_{[0,1)}(x)}$. The sampling points are $x_1 = 0.5$, $x_2 = 3$, $x_3 = 3.9$, with multiplicities $\mu_X(x_1) = \mu_X(x_2) = 2$, $\mu_X(x_3) = 0$. The first sampling point lies in the support of the first three shifts of φ (dot-dashed), the second point is in the support of the next three shifts of φ (dashed), and the last point - in the support of the last shift of φ (solid).

Proof. Fix $p \in [1, \infty)$ and let

$$f = \sum_{\ell \in \mathbb{Z}} c_\ell T_\ell \varphi \in V^p(\varphi).$$

Since $\|f\|_p \asymp \|c\|_p$ for $f \in V^p(\varphi)$ by Lemma 3.7, it suffices to show the inequality

$$\sum_{x \in X} \sum_{s=0}^{\mu_X(x)} |f^{(s)}(x)|^p \asymp \|c\|_{\ell^p}^p.$$

For given $M, L \in \mathbb{N}$ every interval $I_{M,L}(k) = [M + kL, M + (k + 1)L)$ will be investigated separately. The aim is to use Corollary 3.6.

Step 1: The upper bound. Since $\varphi^{(s)}$, $0 \leq s \leq m - 1$, are piecewise continuous, compactly supported functions with finitely many jump discontinuities, Theorem 2.2 (ii) applies and yields the upper bound for the sampling inequality.

Step 2: A class of collocation matrices with uniform bounds. Let $\mathcal{M}$ be the set

$$\mathcal{M} := \left\{ \mu \in \{0, \dots, S\}^N : N \in \mathbb{N}, \sum_{j=1}^N (1 + \mu_j) = L + m - 1 \right\}.$$

If $\mu \in \mathcal{M} \cap \mathbb{Z}^N$, then $N \leq L + m - 1$. Therefore,

$$\mathcal{M} \subseteq \bigcup_{N=1}^{L+m-1} \{0, \dots, S\}^N,$$

in particular, the set $\mathcal{M}$ is finite. Let $\Omega_\mu \subseteq \mathbb{R}^N$ for $\mu \in \mathcal{M} \cap \mathbb{Z}^N$ be the set of vectors $z = (z_1, z_2, \dots, z_N)^t \in \mathbb{R}^N$ with

$$\min_{i \neq j} |z_i - z_j| \geq \delta$$

and

$$(4.5) \quad \begin{aligned} z_j &\in [0, L] \cap \left([\mu_j - m + \varepsilon, -\varepsilon] + 1 + \sum_{n=1}^{j-1} (1 + \mu_n) \right) \\ &= \left[\max \left\{ 0, \mu_j - m + 1 + \varepsilon + \sum_{n=1}^{j-1} (1 + \mu_n) \right\}, \min \left\{ L, 1 - \varepsilon + \sum_{n=1}^{j-1} (1 + \mu_n) \right\} \right]. \end{aligned}$$

In other words, Ω_μ contains all point configurations in $[0, L]$ that satisfy Condition (4.4). The set is bounded and closed, hence compact in $\mathbb{R}^N$. For each $z \in \Omega_\mu$ we consider the collocation matrix

$$\mathbf{M}(z, \mu, -m + 1) \in \mathbb{R}^{(L+m-1) \times (L+m-1)}$$

as defined in the proof of Corollary 3.6 in (3.6) with parameters $D = L + m - 1$ and $\ell_0 = -m + 1$. By Corollary 3.6 and the positions of the points (cf. (4.5)), the matrix is invertible. In addition, $z \mapsto \mathbf{M}(z, \mu, -m + 1)$ is continuous on the compact set Ω_μ , so the matrices are bounded from above and below uniformly in $\omega \in \Omega_\mu$. To see this, a trivial upper bound is a multiple of the maximum of the entries of $\mathbf{M}(z, \mu, -m + 1)$. For the lower bound, we note that $z \mapsto \mathbf{M}(z, \mu, -m + 1)^{-1}$ is continuous by Cramer's rule, thus we can again obtain a uniform upper bound $0 < a_p(\mu)^{-1}$ for $\mathbf{M}(z, \mu, -m + 1)^{-1}$. Then $a_p(\mu)$ is a strictly positive uniform lower bound of $\mathbf{M}(z, \mu, -m + 1)$, i.e., for all $c \in \mathbb{R}^{L+m-1}$

$$\|\mathbf{M}(z, \mu, -m + 1) \cdot c\|_p \geq a_p(\mu) \|c\|_p.$$

Since $\mathcal{M}$ is finite, we can set

$$A_p := \min\{a_p(\mu) : \mu \in \mathcal{M}\} > 0.$$

Step 3: The lower estimate. Let $x^k = (x_1^k, \dots, x_{N(k)}^k)$ be the points in $X \cap I_{M,L}(k)$ that satisfy Condition (iii), i.e., they satisfy (4.2), (4.3), and (4.4). We set $\mu^k = (\mu_1^k, \dots, \mu_{N(k)}^k) \in \mathcal{M}$ and $z_j^k = x_j^k - (M + kL)$. Then $z^k \in \Omega_{\mu^k}$. Let $\mathbf{F}(x^k)$ be the complete list of evaluations

$$\mathbf{F}(x^k) := \left(f(x_1^k), \dots, f(\mu_1^k)(x_1^k), \dots, f(x_{N(k)}^k), \dots, f(\mu_{N(k)}^k)(x_{N(k)}^k) \right) \in \mathbb{C}^{L+m-1}.$$

For $p \in [1, \infty)$, we estimate from below the sum

$$\sum_{x \in X} \sum_{s=0}^{\mu_X(x)} |f^{(s)}(x)|^p \geq \sum_{k \in \mathbb{Z}} \sum_{j=1}^{N(k)} \sum_{s=0}^{\mu_j^k} |f^{(s)}(x_j^k)|^p = \sum_{k \in \mathbb{Z}} \|\mathbf{F}(x^k)\|_p^p.$$

We now consider the restriction of f to the interval $I_{M,L}(k)$ and shift it to $[0, L)$. The resulting function is

$$T_{-(M+kL)}(f|_{I_{M,L}(k)}) = T_{-(M+kL)} \sum_{\ell=M+kL-m+1}^{M+kL+L-1} c_\ell T_\ell \varphi = \sum_{\ell=-m+1}^{L-1} c_{\ell+M+kL} T_\ell \varphi.$$

We denote by $\mathbf{c}_k = (c_{M+kL-m+1}, \dots, c_{M+kL+L-1})^t$ the extracted coefficients which are active on $I_{M,L}(k)$. By Corollary 3.6, specifically (3.6), the measurements $\mathbf{F}(x^k)$ are given by

$$\mathbf{F}(x^k) = \mathbf{M}(z^k, \mu^k, -m+1) \cdot \mathbf{c}_k.$$

From this follows

$$\sum_{x \in X} \sum_{s=0}^{\mu_X(x)} |f^{(s)}(x)|^p \geq \sum_{k \in \mathbb{Z}} \|\mathbf{M}(z^k, \mu^k, -m+1) \cdot \mathbf{c}_k\|_p^p \geq A_p^p \sum_{k \in \mathbb{Z}} \|\mathbf{c}_k\|_p^p \geq A_p^p \|c\|_p^p.$$

In the last inequality, we have used the fact that every $\ell \in \mathbb{Z}$ is contained in some interval $M+kL-m+1 \leq \ell \leq M+kL+L-1$. The estimate for $p = \infty$ is analogous, and we are done. $\square$

Remark. The assumptions in Theorem 4.2 are not easy to visualize. We will therefore investigate a different set of conditions in Section 6. But Theorem 4.2 is a powerful result, and the imposed conditions are necessary and optimal in several regards.

(i) The restriction of $f \in V^p(\varphi)$ to the interval $I_{M,L}(k)$ is given by (4.1) and is a linear combination of $L+m-1$ linearly independent shifts $T_\ell \varphi$, this is Condition (4.3)).

(ii) To recover the restriction of f on the interval $I_{M,L}(k)$ from $L+m-1$ samples, the associated collocation matrix must be invertible. This is implied by Condition (4.4). The margin $\varepsilon > 0$ leads to uniform estimates in k .

(iii) Condition (4.3) implies that (X, μ_X) contains a subset $X' \subseteq X$ with multiplicity function $\mu_{X'}$ satisfying $\mu_{X'}(x) \leq \mu_X(x)$, $x \in X'$, such that $D^-(X', \mu_{X'}) = 1 + \frac{m-1}{L}$, cf. [21, Prop. 3.7.]. Thus every sampling set subject to the assumptions of Theorem 4.2 contains a sampling set with a weighted density close to the necessary density condition.

To illustrate the strength of Theorem 4.2, we draw two consequences. First, we check standard sampling sets without multiplicities. It will be convenient to label X as a bi-infinite strictly increasing sequence $X = (x_j)_{j \in \mathbb{Z}}$, $x_j < x_{j+1}$. In this notation the set X is δ -separated if $\inf_{j \in \mathbb{Z}} (x_{j+1} - x_j) \geq \delta > 0$, and its maximum gap is defined as

$$\mathbf{mg}(X) := \sup_{j \in \mathbb{Z}} (x_{j+1} - x_j).$$

The following result extends [4, Thm. 4] from B-splines to general PEB-splines.

Corollary 4.3. *Let φ be a PEB-spline of order $m \in \mathbb{N}$. If the sampling set $X \subseteq \mathbb{R}$ is separated and the maximum gap satisfies $\mathbf{mg}(X) < 1$, then X is a sampling set for $V^p(\varphi)$.*

If $\mu_X \equiv 0$, then the assumptions in Theorem 4.2 are identical with those in [4, Thm. 3]. Further, the proof of [4, Thm. 4] shows that a set with $\mathbf{mg}(X) < 1$ satisfies the conditions in [4, Thm. 4], hence, the conditions in Theorem 4.2. We will give an alternative proof in Section 6.

To construct sampling sets with multiplicities, one can start with a sampling set without derivatives and remove certain sampling points while adding multiplicities at neighbouring sampling points (see Figure 2).

Finally, we show that there exist sets with a given multiplicity function that satisfies the sufficient conditions of Theorem 4.2 with a density arbitrarily close to the necessary condition.

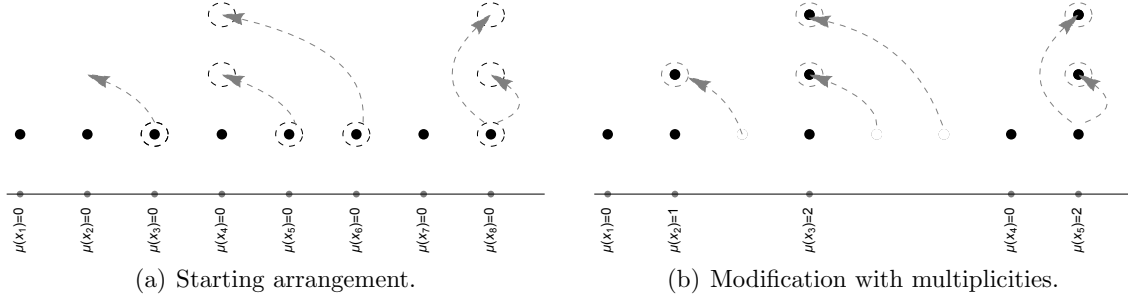


Figure 2. Local modification of a sampling set that preserves the conditions of Thm. 4.2.

Theorem 4.4. *For every $\nu > 0$ and given multiplicity sequence $\mu = (\mu_j)_{j \in \mathbb{Z}}$ with values in $\{0, \dots, m-1\}$, there exists a sequence $X = (x_j)_{j \in \mathbb{Z}}$, such that (X, μ_X) is sampling for $V^p(\varphi)$, $\mu_X(x_j) = \mu_j$ for all $j \in \mathbb{Z}$, and $D^-(X, \mu_X) < 1 + \nu$.*

Proof. The sampling set X will be a suitable subset of a sufficiently dense lattice $\alpha\mathbb{Z}$. In line with Theorem 4.2 we use a suitable partition and argue locally. Since X will be a subset of $\alpha\mathbb{Z}$, it is separated, and Condition (ii) in Theorem 4.2 is satisfied by assumption. The main part is to construct points x and submultiplicities $\tilde{\mu}$ to satisfy the third condition.

Step 1. Global parameters. Given $\nu > 0$, choose $L \in \mathbb{N}$ such that $\frac{2m-2}{L} < \nu$. We set $\varepsilon = \alpha = \frac{1}{(m+1)}$. We will work with the interval partitioning $I_{L,0}(k) = [kL, (k+1)L)$. The key observation here is the fact that $\#([\ell + \varepsilon, \ell + 1 - \varepsilon] \cap \alpha\mathbb{Z}) = m$, and for a fixed $\ell \in \mathbb{Z}$

$$[\ell + \varepsilon, \ell + 1 - \varepsilon] \cap \left([\mu_j^k - m + \varepsilon, -\varepsilon] + kL + 1 + \sum_{n=1}^{j-1} (1 + \mu_n^k) \right)$$

is non-empty for at most m consecutive indices $j, j+1, \dots, j+m-1$. Thus, in $[\ell + \varepsilon, \ell + 1 - \varepsilon]$ we will need at most m sampling points in $\alpha\mathbb{Z}$, which are available by choice of α .

Step 2. Local construction. Let $N \in \mathbb{N}$ be the minimal natural number with $\sum_{j=1}^N (1 + \mu_j) \geq L + m - 1$. We set

$$\tilde{\mu}_j := \mu_j, \quad 1 \leq j \leq N-1, \quad \tilde{\mu}_N := (L + m - 1) - \sum_{j=1}^N (1 + \mu_j) - 1.$$

The minimality of N implies that $0 \leq \tilde{\mu}_N \leq \mu_N$, and the definition of $\tilde{\mu}_N$ implies that

$$\sum_{j=1}^N (1 + \tilde{\mu}_j) = L + m - 1.$$

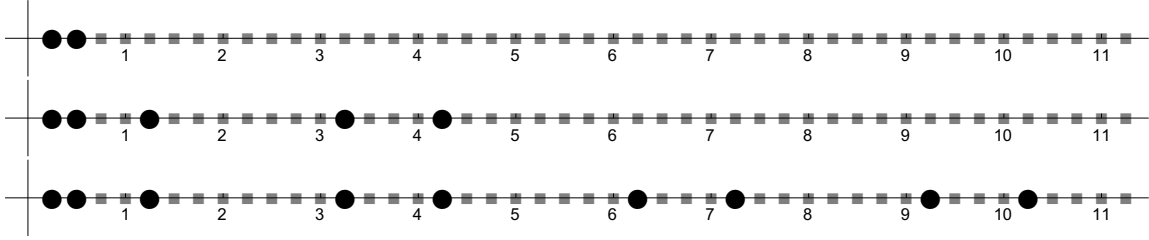


Figure 3. Choice of points (black circles) for $m = 3$, $\mu = (1 + (-1)^j)_{j \in \mathbb{Z}}$ and $\nu = \frac{1}{2}$ from the lattice $\frac{1}{4}\mathbb{Z}$ (gray squares). The interval length is $L = 13$, and the number of points is $N = 9$. The sampling locations are αk , $k \in \{1, 2, 5, 13, 17, 25, 29, 37, 41\}$. The plots represent the choice of the first j points, $j \in \{2, 5, 9\}$.

We now choose points $x_j \in [0, L)$ inductively (see Figure 3) and start with $x_1 = \alpha = \varepsilon$, which clearly satisfies (4.4). Assume now we have picked $x_1, \dots, x_{j-1}$. We choose x_j to be *the minimal element in $\alpha\mathbb{Z}$ that satisfies the Schoenberg-Whitney conditions (4.4)*. Precisely, let

$$\ell = \max \left\{ 0, \tilde{\mu}_j^k - m + 1 + \sum_{n=1}^{j-1} (1 + \tilde{\mu}_n^k) \right\},$$

and define

$$x_j = \min \{ \alpha k : k \in \mathbb{Z}, \alpha k \in [\ell + \varepsilon, \ell + 1 - \varepsilon] \}.$$

As observed in Step 1, we can always choose such x_j . The condition $x_j \in [\ell + \varepsilon, \ell + 1 - \varepsilon]$ is clearly stronger than the Schoenberg-Whitney conditions (4.4). The chosen points are in $[0, L)$ because

$$0 \leq \ell \leq \tilde{\mu}_N - m + 1 + \sum_{n=1}^{N-1} (1 + \tilde{\mu}_n) = L - 1 < L.$$

Step 3. Partitioning μ and $\mathbb{R}$. We now partition the sequence μ into blocks $M_k := \{\mu_{N_{k-1}+1}, \dots, \mu_{N_k}\}$. We set $N_{-1} = 0$ and construct N_k inductively. Given N_{k-1} , we choose $N_k \in \mathbb{Z}$ minimal with

$$\sum_{n=N_{k-1}+1}^{N_k} (1 + \mu_j) \geq L + m - 1.$$

We carry out the local step for each block and obtain points $x_{N_{k-1}+j} \in [kL, (k+1)L)$ with submultiplicities $\tilde{\mu}_{N_{k-1}+j}$ satisfy all conditions of Theorem 4.2. Consequently, the set $\{x_m : m \in \mathbb{Z}\} = \bigcup_{k \in \mathbb{Z}} \{x_{N_{k-1}+j} : j = 1, \dots, N_k\}$ with multiplicities $(\mu_m)_{m \in \mathbb{Z}}$ form a sampling set in $V^p(\varphi)$.³

Step 4. Density. Since we have

$$\sum_{n=N_{k-1}+1}^{N_k} (1 + \mu_n) = L + m - 1 + (\mu_{N_k} - \tilde{\mu}_{N_k}) \leq L + 2m - 2$$

samples in every interval $[kL, (k+1)L)$, the weighted Beurling density is bounded by $D^-(X, \mu) \leq \frac{L+2m-2}{L} = 1 + \frac{2m-2}{L} < 1 + \nu$. $\square$

Remark. Note that a controlled perturbation of X by setting $\tilde{x}_j = x_j + \gamma_j$, $j \in \mathbb{Z}$ for $|\gamma_j| \leq \tau\epsilon$ for a fixed $\tau \in (0, 1)$ still satisfies the conditions of Theorem 4.2. Additionally, there exist sampling sets with multiplicities (X, μ_X) which are sampling for $V^p(\varphi)$, $D^-(X, \mu_X) < 1 + \nu$, but $D^-(X) < 1$, in particular, X is not sampling for $V^p(\varphi)$. A simple example is a sequence that alternates between s_1 and s_2 , $s_1 + s_2 > 0$. If $L \in \mathbb{N}$ satisfies $L + m - 1 = T(s_1 + s_2 + 2)$ for sufficiently large $T \in \mathbb{N}$, then $\tilde{\mu}_j = \mu_j$ for all $j \in \mathbb{Z}$, there are $2T$ sampling locations in each interval $[kL, (k+1)L)$ and the resulting set X is L -periodic. Its lower Beurling density is given by

$$\begin{aligned} D^-(X) &= \lim_{K \rightarrow \infty} \frac{\#(X \cap [-kL, kL])}{2kL} = \frac{2T}{L} = \frac{2T}{T(s_1 + s_2 + 2) - (m - 1)} \\ &= \left(\frac{s_1 + s_2}{2} + 1 - \frac{m - 1}{2T} \right)^{-1} < 1, \end{aligned}$$

whenever $T > \frac{m-1}{s_1+s_2}$. This argument can be generalized to more complicated sequences μ .

5. WEAK LIMITS

A particular case to consider is sampling with a constant multiplicity S for some integer $0 \leq S \leq m - 1$. Since every point comes with $S + 1$ data, one expects that a suitably modified gap condition might be sufficient for sampling with derivatives. Based on Corollary 4.3, it is natural to conjecture that

$$(5.1) \quad \frac{\text{mg}(X)}{1 + S} < 1$$

is sufficient for sampling. This is indeed correct, but it does not follow from Theorem 4.2. One can show that there are sets that satisfy (5.1) but violate the conditions (4.2), (4.3), and (4.4). These examples will be addressed in Section 7.

To study maximal gap conditions and achieve an optimal result, we will use the technique of weak limits. These go back to Beurling [10] and have been introduced for the investigation of shift-invariant spaces in [20, 21].

³Note that only at the interval ends N_k we had to reduce the multiplicities to apply Theorem 4.2.

We recall Beurling's notion of a weak limit of a sequence of sets. A sequence $(X_n)_{n \in \mathbb{N}}$ of subsets of $\mathbb{R}$ *converges weakly* to a set $Y \subseteq \mathbb{R}$, denoted by $X_n \xrightarrow{w} Y$, if for every open bounded interval (a, b) and every $\varepsilon > 0$, there exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$,

$$X_n \cap (a, b) \subseteq Y + (-\varepsilon, \varepsilon) \quad \text{and} \quad Y \cap (a, b) \subseteq X_n + (-\varepsilon, \varepsilon).$$

Given a tuple of sets $\vec{X} = (X^0, \dots, X^S)$, a tuple $\vec{Y} = (Y^0, \dots, Y^S)$ is called a *weak limit of integer translates* of $\vec{X}$ if there exists a sequence $(k_n)_{n \in \mathbb{N}} \subseteq \mathbb{Z}$ such that $X^s - k_n \xrightarrow{w} Y^s$ for all $0 \leq s \leq S$. An important part of the definition is the fact that the sequence of translates is the same for all Y^s . The set of weak limits of integer translates of X is denoted by $W_{\mathbb{Z}}(\vec{X})$. Weak limits enter sampling theory through the following characterization of sampling sets [21, Thm. 2.1].

Theorem 5.1. *Assume that $\Phi = (\Phi^0, \dots, \Phi^S)$ consists of continuous functions with a minimal decay property $|\Phi^s(x)| \leq C(1 + |x|)^{-1-\varepsilon}$ and Φ has stable integer shifts. Let $\vec{X} = (X^0, \dots, X^S)$ be a tuple of separated sets. Then the following are equivalent.*

- (i) $\vec{X}$ is a sampling set for $V^p(\Phi)$ for some $p \in [1, \infty]$.
- (ii) $\vec{X}$ is a sampling set for $V^p(\Phi)$ for all $p \in [1, \infty]$.
- (iii) Every weak limit $\vec{Y} \in W_{\mathbb{Z}}(\vec{X})$ is a sampling set for $V^\infty(\Phi)$.
- (iv) Every weak limit $\vec{Y} \in W_{\mathbb{Z}}(\vec{X})$ is a uniqueness set for $V^\infty(\Phi)$.

When sampling on (X, μ_X) , this theorem is applied to the generator $\Phi = (\varphi, \varphi', \dots, \varphi^{(S)})$ and the sets

$$X^s := \{x \in X : \mu_X(x) \geq s\}, \quad 0 \leq s \leq \max \{\mu_X(x) : x \in X\}.$$

In this case, $\vec{X} - k_n$ can be identified with $(X, \mu_X) - k_n := (X - k_n, \mu_{X-k_n})$, where

$$\mu_{X-k_n}(x - k_n) = \mu_X(x), \quad x \in X.$$

For separated sets, we can use an alternative characterization of weak convergence (cf. [19, Lem. 4.4.], [21, Prop. 3.2.]).

Proposition 5.2. *Let (X, μ_X) be a separated set with multiplicity function $\mu_X : X \rightarrow \{0, \dots, S\}$, (Y, μ_Y) be a set with multiplicity function μ_Y , and $(k_n)_{n \in \mathbb{N}} \subseteq \mathbb{Z}$ a sequence of integers. Then $X^s - k_n \xrightarrow{w} Y^s$ as $n \rightarrow \infty$ for all $0 \leq s \leq S$, if and only if*

$$\sum_{x \in X} (1 + \mu_X(x)) \delta_{x-k_n} \longrightarrow \sum_{y \in Y} (1 + \mu_Y(y)) \delta_y, \quad \text{as } n \longrightarrow \infty,$$

in the (C_c^*, C_c) -topology (where C_c denotes the class of continuous functions with compact support).

For splines, an additional difficulty arises that is not formally covered by Theorem 5.1. The last derivative $\varphi^{(m-1)}$ is only piecewise continuous with finitely many jump discontinuities. In this case, we have to impose an additional condition on the m -th sampling set. The modified version of Theorem 5.1 goes as follows:

Corollary 5.3. *Let $\varphi \in C^{S-1}(\mathbb{R})$ be a compactly supported, piecewise C^S -function with finitely many jump discontinuities $\mathcal{J} \subseteq \mathbb{R}$. Further, assume that φ has stable integer translates. Let X be a separated set with multiplicity function $\mu_X : X \rightarrow \{0, \dots, S\}$ satisfying*

$$(5.2) \quad \text{dist}(\{x \in X : \mu_X(x) = S\}, \mathcal{J} + \mathbb{Z}) > 0.$$

Then the following statements are equivalent:

- (i) (X, μ_X) is a sampling set for $V^p(\varphi)$ for some $p \in [1, \infty]$.
- (ii) (X, μ_X) is a sampling set for $V^p(\varphi)$ for all $p \in [1, \infty]$.
- (iii) Every weak limit $(Y, \mu_Y) \in W_{\mathbb{Z}}(X, \mu_X)$ is a sampling set for $V^\infty(\varphi)$.
- (iv) Every weak limit $(Y, \mu_Y) \in W_{\mathbb{Z}}(X, \mu_X)$ is a uniqueness set for $V^\infty(\varphi)$.

Remark. The equivalence of (i) and (ii) holds without condition (5.2). For PEB-splines of order m , (5.2) is only needed if $X^{m-1} = \{x \in X : \mu(x) = m-1\}$ is non-empty, i.e., when $(m-1)$ -th derivatives are part of the data. If $X^{m-1} = \emptyset$, then (5.2) is void and not required.

The proof of Corollary 5.3 is a small generalization of [20, Thm. 3.1]. For the sake of completeness, we sketch the modifications in the Appendix.

When we label a separated set $X \subseteq \mathbb{R}$ as a strictly increasing sequence, we can define the weighted maximum gap as

$$\begin{aligned} \mathbf{mg}(X, \mu_X) &:= \max \left\{ \sup_{j \in \mathbb{Z}} \frac{x_{j+1} - x_j}{1 + \mu_X(x_j)}, \sup_{j \in \mathbb{Z}} \frac{x_{j+1} - x_j}{1 + \mu_X(x_{j+1})} \right\} \\ &= \sup_{j \in \mathbb{Z}} \frac{x_{j+1} - x_j}{1 + \min\{\mu_X(x_j), \mu_X(x_{j+1})\}}. \end{aligned}$$

We will show that $\mathbf{mg}(X, \mu_X) < 1$ is sufficient for a set to be a sampling set.

To apply Corollary 5.3, we first study the weighted maximum gap of a weak limit $(Y, \mu_Y) \in W_{\mathbb{Z}}(X, \mu_X)$.

Lemma 5.4. *Let (X, μ_X) be a separated set with a multiplicity function μ_X and (Y, μ_Y) be a weak limit of (X, μ_X) . Then*

$$\mathbf{mg}(Y, \mu_Y) \leq \mathbf{mg}(X, \mu_X).$$

Proof. Let (Y, μ_Y) be a weak limit of (X, μ_X) . Since X is separated, we can use Proposition 5.2 to describe the weak convergence by means of measures

$$(5.3) \quad \sum_{x \in X} (1 + \mu_X(x)) \delta_{x-k_n} \longrightarrow \sum_{y \in Y} (1 + \mu_Y(y)) \delta_y, \quad \text{as } n \longrightarrow \infty.$$

For arbitrary $\varepsilon > 0$, we choose a pair $y_j < y_{j+1}$ of consecutive points in Y such that

$$\frac{y_{j+1} - y_j}{1 + \min\{\mu_Y(y_j), \mu_Y(y_{j+1})\}} > \mathbf{mg}(Y, \mu_Y) - \varepsilon.$$

Let $\delta > 0$ be the separation constant of X and $f_0, f_1 \in C_c(\mathbb{R})$ be two auxiliary functions such that $0 \leq f_0, f_1 \leq 1$,

$$\begin{aligned} \text{supp}(f_0) &\subseteq \left(-\frac{\delta}{2}, \frac{\delta}{2}\right), & f_0(x) &= 1 \text{ for all } x \in \left(-\frac{\delta}{4}, \frac{\delta}{4}\right), \\ \text{supp}(f_1) &\subseteq (y_j, y_{j+1}), & f_1(x) &= 1 \text{ for all } x \in \left(y_j + \frac{\delta}{2}, y_{j+1} - \frac{\delta}{2}\right). \end{aligned}$$

Then by (5.3)

$$\begin{aligned} \sum_{x \in X} (1 + \mu_X(x)) f_0(x - k_n - y_j) &= \\ = \sum_{x \in X} (1 + \mu_X(x)) \delta_{x - k_n}(T_{y_j} f_0) &\rightarrow \sum_{y \in Y} (1 + \mu_Y(y)) \delta_y(T_{y_j} f_0) = 1 + \mu_Y(y_j). \end{aligned}$$

Since X is δ -separated, $f_0 \geq 0$ and $\text{supp}(f_0) \subseteq \left(-\frac{\delta}{2}, \frac{\delta}{2}\right)$, for every $n \in \mathbb{N}$ there is at most one $x \in X$, say x_{n_j} , such that

$$|x_{n_j} - k_n - y_j| < \frac{\delta}{2}.$$

So $x_{n_j} - k_n \rightarrow y_j$ as n tends to infinity and consequently $\mu_X(x_{n_j}) = \mu_Y(y_j)$ for all $n \in \mathbb{N}$. Likewise, there is a sequence $\tilde{x}_{n_j} - k_n - y_{j+1}$ such that

$$(5.4) \quad |\tilde{x}_{n_j} - k_n - y_{j+1}| < \frac{\delta}{2}$$

and $\mu_X(\tilde{x}_{n_j}) = \mu_Y(y_{j+1})$ for all $j \in \mathbb{N}$ sufficiently large.

Using the weak*-convergence for f_1 , we obtain

$$\sum_{x \in X} (1 + \mu_X(x)) f_1(x - k_n) \longrightarrow \sum_{y \in Y} (1 + \mu_Y(y)) f_1(y) = 0,$$

because $\text{supp}(f_1) \cap Y = \emptyset$. As $f_1 \geq 0$, this implies that for all sufficiently large $n \in \mathbb{N}$, $\text{supp}(f_1) \cap (X - k_n) \subseteq [y_j, y_j + \frac{\delta}{2}) \cap [y_{j+1} - \frac{\delta}{2}, y_{j+1})$.

By (5.4) and the δ -separability of the sets, these must be the points $x_{n_j} - k_n$ and $\tilde{x}_{n_j} - k_n$, so that necessarily $\tilde{x}_{n_j} = x_{n_{j+1}}$. Consequently, since $\mu_{X - k_n} = \mu_X(\cdot + k_n)$,

$$\begin{aligned} \mathbf{mg}(X, \mu_X) &\geq \frac{x_{n_{j+1}} - x_{n_j}}{1 + \min\{\mu_X(x_{n_j}), \mu_X(x_{n_{j+1}})\}} \\ &\geq \frac{(x_{n_{j+1}} - k_n) - (x_{n_j} - k_n)}{1 + \min\{\mu_{X - k_n}(x_{n_j} - k_n), \mu_{X - k_n}(x_{n_{j+1}} - k_n)\}} \\ &\geq \frac{y_{j+1} - y_j}{1 + \min\{\mu_Y(y_j), \mu_Y(y_{j+1})\}} \\ &> \mathbf{mg}(Y, \mu_Y) - \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, the claim follows. $\square$

6. SAMPLING VIA MAXIMAL GAP AND WEAK LIMITS

In this section, we will prove two sampling theorems that use gap conditions for the sampling set. These conditions are easier to check than the technical conditions of Theorem 4.2.

We first relate the weighted maximum gap to the Schoenberg-Whitney conditions (3.5). The following statement does not make reference to splines and only involves arrangements of points on the line.

Lemma 6.1 (Combinatorial lemma). *Let $m \geq 2$ and let (X, μ_X) be a separated set with multiplicity function μ_X . Assume that $\mathbf{mg}(X, \mu_X) < 1$. Then for all $x \in X$ there exist $N, L \in \mathbb{N}$, $N, L \geq 2$, depending on x , and points*

$$x_1^* < x_2^* < \cdots < x_N^* \text{ in } X \cap [\lfloor x \rfloor, \lfloor x \rfloor + L]$$

with submultiplicities $\mu_j \in \mathbb{N}$,

$$\mu_j \leq \mu_X(x_j^*), \quad 1 \leq j \leq N,$$

satisfying the following properties:

(i) dimension count:

$$(6.1) \quad \sum_{j=1}^N (1 + \mu_j) = L + m - 1,$$

(ii) Schoenberg-Whitney conditions:

$$(6.2) \quad x_j^* \in (\mu_j - m + 1, 1) + \lfloor x \rfloor + \sum_{n=1}^{j-1} (1 + \mu_n), \quad 1 \leq j \leq N$$

with the usual adaptation if $\mu_j = m - 1$,

(iii) $X \cap [\lfloor x \rfloor + L - 1, \lfloor x \rfloor + L] \neq \emptyset$.

(*) If $x_N^* = \lfloor x \rfloor + L \in \mathbb{Z}$, then, in addition, we have $\mu_N \leq m - 2$.

Although the number of points N and the length of the interval L depend on x , they may be bounded uniformly by

$$N \leq 1 + \lceil (m + 1)(1 - \mathbf{mg}(X, \mu_X))^{-1} \rceil = N_0, \quad \text{and } L \leq m(N_0 + m + 2).$$

Proof. The proof is split into several steps. We make a first selection of points $x_1^*, \dots, x_K^*$ that satisfy all conditions. In the generic case, we are done, but in some exceptional cases, we need to adapt the intermediate parameters and add more points.

Reduction. Since the problem is invariant under integer shifts and reindexing, we may assume without loss of generality that $\lfloor x \rfloor = 0$ and $x_1 = \min\{x \in X : x \geq 0\}$. We need to find $N, L \in \mathbb{N}$, points $x^* \in [0, L]^N$ and submultiplicities $\mu \in \{0, \dots, m - 1\}^N$ with the stated properties.

Step 1: (Preliminary) choice of N and L . The condition $\mathbf{mg}(X, \mu_X) < 1$ leads to a bound for consecutive samples $x_{j+1} - x_j \leq \mathbf{mg}(X, \mu_X)(1 + \mu_X(x_j))$ for all $j \in \mathbb{Z}$. By telescoping we obtain ⁴

$$(6.3) \quad x_n - x_1 = \sum_{j=1}^{n-1} (x_{j+1} - x_j) \leq \mathbf{mg}(X, \mu_X) \sum_{j=1}^{n-1} (1 + \mu_X(x_j)) < \sum_{j=1}^{n-1} (1 + \mu_X(x_j)).$$

Thus the number of data $\sum_{j=1}^{n-1} (1 + \mu_X(x_j))$ always exceeds the length of the sampling interval $x_n - x_1$ by fixed factor $\mathbf{mg}(X, \mu_X)^{-1} > 1$. This leaves some room for manipulation.

In particular, for n large enough, $(n-1)(1 - \mathbf{mg}(X, \mu_X)) > m+1$, we obtain

$$\begin{aligned} \lceil x_n \rceil + m - 1 &\leq x_n - x_1 + 2 + m - 1 \\ &\leq \mathbf{mg}(X, \mu_X) \sum_{j=1}^{n-1} (1 + \mu_X(x_j)) + m + 1 \\ &\leq \mathbf{mg}(X, \mu_X) \sum_{j=1}^{n-1} (1 + \mu_X(x_j)) + (1 - \mathbf{mg}(X, \mu_X))(n-1) \\ &\leq \sum_{j=1}^{n-1} (1 + \mu_X(x_j)). \end{aligned}$$

Now we define K as the smallest integer with

$$(6.4) \quad \lceil x_K \rceil + m - 1 \leq \sum_{j=1}^K (1 + \mu_X(x_j)).$$

Clearly $K \geq 1$ and by the preceding argument $K \leq (m+1)/(1 - \mathbf{mg}(X, \mu_X))$. This choice of K implies the reverse inequality

$$(6.5) \quad \lceil x_j \rceil + m - 1 > \sum_{k=1}^j (1 + \mu_X(x_k)) \quad \text{for } j = 1, \dots, K-1.$$

The inequalities (6.4) and (6.5) contain the key idea of the proof.

As the preliminary choice for the interval length we set

$$L_0 = \lceil x_K \rceil.$$

The final choice of the number of points N will be from $\{2, 3, K, K+1\}$ and $L \leq \lceil x_K \rceil + m \leq (K+1)m + m$, whence we obtain their boundedness independent of the initial choice of $x \in X$.

Step 2: Choice of the initial x^* and their submultiplicities μ when $K > 1$. Assume first that $K > 1$. The exceptional case $K = 1$ will be treated later. We choose

$$x_j^* := x_j, \quad \mu_j := \mu_X(x_j), \quad 1 \leq j \leq K-1,$$

⁴Here we use $\frac{x_{j+1} - x_j}{1 + \mu_X(x_j)} \leq \mathbf{mg}(X, \mu_X)$.

and first show the Schoenberg-Whitney conditions (6.2) for these points. Since $\lfloor x_1 \rfloor = 0$, for $j = 1$ these are the inequalities $x_1 < 1$ and $x_1 \geq 0 \geq \mu_1 - m + 1$ with equality if and only if $\mu_1 = m - 1$.

For all $j = 2, \dots, K - 1$, (6.3) yields the upper inequality

$$(6.6) \quad \begin{aligned} x_j^* &= x_j = x_j - x_1 + x_1 \\ &\leq \mathbf{mg}(X, \mu_X) \sum_{n=1}^{j-1} (1 + \mu_X(x_n)) + 1 < 1 + \sum_{n=1}^{j-1} (1 + \mu_n). \end{aligned}$$

The lower bound is contained in (6.5) via

$$x_j > \lceil x_j \rceil - 1 > -m + \sum_{n=1}^j (1 + \mu_n) = 1 + \mu_j - m + \sum_{n=1}^{j-1} (1 + \mu_n).$$

Step 3. For $j = K$ we set $x_K^* = x_K$ and

$$\mu_K := \lceil x_K \rceil + m - 2 - \sum_{j=1}^{K-1} (1 + \mu_j).$$

This choice ensures condition (6.1);

$$(6.7) \quad \sum_{j=1}^K (1 + \mu_j) = \lceil x_K \rceil + m - 1 = L_0 + m - 1.$$

To check the range of μ_K , we use (6.4) and obtain

$$\begin{aligned} 1 + \mu_K &= \lceil x_K \rceil + m - 1 - \sum_{j=1}^{K-1} (1 + \mu_j) \\ &\leq \sum_{j=1}^K (1 + \mu_j) - \sum_{j=1}^{K-1} (1 + \mu_j) = 1 + \mu_X(x_K), \end{aligned}$$

and by (6.5) for $j = K - 1$

$$\begin{aligned} 1 + \mu_K &= \lceil x_K \rceil + m - 1 - \sum_{j=1}^{K-1} (1 + \mu_j) \\ &\geq \lceil x_{K-1} \rceil + m - 1 - \sum_{j=1}^{K-1} (1 + \mu_j) > 0. \end{aligned}$$

Since $\mu_K \in \mathbb{Z}$, we obtain $0 \leq \mu_K \leq \mu_X(x_K^*)$. To show that x_K^* satisfies the Schoenberg-Whitney condition (6.2), we recall that for $j \leq K - 1$, $\mu_j = \mu_X(x_j)$, thus the estimate in (6.6) holds for x_K^* as well. The lower bound holds by choice of μ_K :

$$x_K^* = x_K > \lceil x_K \rceil - 1 = \mu_K - m + 1 + \sum_{n=1}^{K-1} (1 + \mu_n).$$

Step 4: Generic Case $K > 1, x_K > 1$. If $x_K > 1$, and additionally $x_{K-1} \notin \mathbb{Z}$ or $\mu_{K-1} \leq m-2$, then we set $N := K$ and $L := L_0 = \lceil x_K \rceil$ so that $L-1 \leq x_N^* = x_K \leq L$. By construction the points $x_1^* < \dots < x_N^*$ with submultiplicities $\mu_1, \dots, \mu_N$ satisfy (i) - (iii) and we are done.

It remains to consider the exceptional cases: (i) $x_K \geq 1$ with $x_K \in \mathbb{Z}$ and $\mu_K = m-1$, (ii) $x_K < 1$, and (iii) $K = 1$.

Step 5. Case $x_K^* = x_K \geq 1, x_K \in \mathbb{Z}, \mu_K = m-1$. In this case, we add the next sampling point x_{K+1} and set

$$x_{K+1}^* := x_{K+1} \quad \text{and} \quad \mu_{K+1} := \lceil x_{K+1} \rceil - x_K - 1.$$

Since $\lceil x_{K+1} \rceil - x_K > 0$ is an integer, we have $\mu_{K+1} \geq 0$. We set $N = K+1$ and $L = \lceil x_{K+1} \rceil$ and verify items (i) — (iii) after adding the new point x_{K+1}^* . We substitute (6.7) and the definition of μ_K and obtain

$$\sum_{j=1}^{K+1} (1 + \mu_j) = \lceil x_K \rceil + m - 1 + 1 + \mu_{K+1} = \lceil x_{K+1} \rceil + m - 1 = L + m - 1.$$

Next,

$$\mu_{K+1} = \lceil x_{K+1} \rceil - \lceil x_K \rceil - 1 < x_{K+1} - x_K < \mu_X(x_{K+1}) + 1. \quad ^5$$

This inequality of integers implies $\mu_{K+1} \leq \mu_X(x_{K+1})$. If also $x_{K+1} \in \mathbb{Z}$, then

$$\mu_{K+1} = x_{K+1} - x_K - 1 < \mu_X(x_{K+1})$$

and thus $\mu_{K+1} \leq m-2$, which takes care of the extra condition (*).

To verify the Schoenberg-Whitney conditions (6.2), we use $\sum_{n=1}^K (1 + \mu_n) = \lceil x_K \rceil + m - 1$ from (6.7) and find the upper estimate from

$$\begin{aligned} x_{K+1}^* &= x_{K+1} - x_K + \lceil x_K \rceil \\ &< 1 + \mu_X(x_K) + \sum_{n=1}^K (1 + \mu_n) - m + 1 \leq 1 + \sum_{n=1}^K (1 + \mu_n), \end{aligned}$$

whereas the lower estimate follows from

$$\begin{aligned} \mu_{K+1} - m + 1 + \sum_{n=1}^K (1 + \mu_n) &= \lceil x_{K+1} \rceil - \lceil x_K \rceil - 1 - m + 1 + \lceil x_K \rceil + m - 1 \\ &= \lceil x_{K+1} \rceil - 1 < x_{K+1}. \end{aligned}$$

The last condition (iii) is satisfied by definition because $x_{K+1}^* \in [L, L+1)$.

Step 6: Case $x_K < 1, K > 1$. Let $\tilde{K}$ be the minimal index with $x_{\tilde{K}} > 1$. Due to its minimality and the maximum gap condition, the points $x_{\tilde{K}-1}, x_{\tilde{K}}$ satisfy

$$(6.8) \quad x_K \leq x_{\tilde{K}-1} \leq 1 \quad \text{and} \quad 1 < x_{\tilde{K}} < x_{\tilde{K}-1} + m \leq 1 + m.$$

⁵Here we use $\frac{x_{j+1} - x_j}{1 + \mu_X(x_{j+1})} \leq \mathbf{mg}(X, \mu_X) < 1$.

We set $N := K + 1$, $L := \lceil x_{\tilde{K}} \rceil$, and

$$x_{K+1}^* := x_{\tilde{K}}, \quad \mu_{K+1} := \lceil x_{\tilde{K}} \rceil - 2 < x_{\tilde{K}} - 1.$$

This choice guarantees that $L = \lceil x_{\tilde{K}} \rceil \geq 2$, $\mu_{K+1} \geq 0$, and $x_{K+1}^* \in [L - 1, L] \cap X$. We have

$$\mu_{K+1} < x_{\tilde{K}} - 1 \leq x_{\tilde{K}} - x_{\tilde{K}-1} < 1 + \mu_X(x_{\tilde{K}})^6,$$

and again $\mu_{K+1} \leq \mu_X(x_{K+1}^*)$.

For the Schoenberg-Whitney condition (6.2) for x_{K+1} , we observe that (6.7) says that $\sum_{j=1}^K (1 + \mu_j) = \lceil x_K \rceil + m - 1 = m$, so that (6.8) says that

$$x_{K+1}^* = x_{\tilde{K}} < 1 + m = 1 + \sum_{j=1}^K (1 + \mu_j),$$

and

$$\mu_{K+1} - m + 1 + \sum_{j=1}^K (1 + \mu_j) = \lceil x_{\tilde{K}} \rceil - 2 - m + 1 + m = \lceil x_{K+1}^* \rceil - 1 < x_{K+1}^*.$$

Step 7: Case $K = 1$. In this case, (6.5) is void and cannot be used for the Schoenberg-Whitney conditions for $j > 1$. We need to modify the selection of points. We distinguish three subcases.

If $x_1 > 0$, then the inequalities $\mu_X(x_1) - m + 1 \leq 0 < x_1 < 1$ express the Schoenberg-Whitney condition for x_1 . We proceed as in Step 6 and choose $x_2^* = x_{\tilde{K}}$ with $x_{\tilde{K}-1} \leq 1 < x_{\tilde{K}}$ and $\mu_2 = \lceil x_{\tilde{K}} \rceil - 2$. We set $N = 2$ and $L = \lceil x_{\tilde{K}} \rceil \geq 2$. Conditions (i) — (iii) are verified readily as before.

If $x_1 = 0$, then we must have $\mu_X(x_1) = m - 1$. If $x_2 \leq 1$, then we set

$$\begin{aligned} x_2^* &:= x_2, & \mu_2 &:= 0, \\ x_3^* &= x_{\tilde{K}} & \mu_3 &= \lceil x_{\tilde{K}} \rceil - 2. \end{aligned}$$

With $N = 3$ points and interval length $L = \lceil x_{\tilde{K}} \rceil \geq 2$, all conditions (i) — (iii) are satisfied.

If $x_1 = 0$ and $x_2 > 1$, set $N = 2$, $N := 2$, $L := \lceil x_2 \rceil \geq 2$, and

$$x_2^* := x_2, \quad \mu_2 := \lceil x_2 \rceil - 2.$$

Properties (i) — (iii) are verified as in Step 6.

As we have covered all possible cases, this concludes the proof. $\square$

Theorem 6.2. *Let φ be a PEB-spline of order $m \in \mathbb{N}$ and $X \subseteq \mathbb{R}$ be a separated set with a multiplicity function $\mu_X : X \rightarrow \{0, \dots, m - 1\}$. If*

$$\mathbf{mg}(X, \mu_X) = \max \left\{ \sup_{j \in \mathbb{Z}} \frac{x_{j+1} - x_j}{1 + \mu_X(x_j)}, \sup_{j \in \mathbb{Z}} \frac{x_{j+1} - x_j}{1 + \mu_X(x_{j+1})} \right\} < 1,$$

then (X, μ) is a uniqueness set for $V^p(\varphi)$ for all $1 \leq p \leq \infty$.

⁶Here we use $\frac{x_{j+1} - x_j}{1 + \mu_X(x_{j+1})} \leq \mathbf{mg}(X, \mu_X) < 1$.

Proof. Assume that $f^{(s)}(x) = 0$ for all $0 \leq s \leq \mu_X(x)$ for all $x \in X$. We need to prove that $f \equiv 0$.

Step 1. Let $x \in X$ be arbitrary and set $a = \lfloor x \rfloor$. By Lemma 6.1, there exists an integer length $L \geq 2$ and points $a \leq x_1^* < \dots < x_N^* \leq a + L$ with multiplicities $\mu_1, \dots, \mu_N \in \{0, \dots, m-1\}$ satisfying properties (i) — (iii) (correct dimension count, Schoenberg-Whitney conditions and $x_N^* \in [L-1, L]$). By assumption on f , we know that

$$0 = f^{(s_j)}(x_j^*) = \left(\sum_{\ell \in \mathbb{Z}} c_\ell T_\ell \varphi \right)^{(s_j)}(x_j^*) = \left(\sum_{\ell=a-m+1}^{a+L-1} c_\ell T_\ell \varphi \right)^{(s_j)}(x_j^*), \quad 0 \leq s_j \leq \mu_j, \quad 1 \leq j \leq N.$$

Since the Schoenberg-Whitney conditions are satisfied for the points x_j^* with multiplicities $\mu_j, j = 1, \dots, N$, Corollary 3.6 asserts that $f = 0$ on $[a, a + L]$.

Step 2. By induction we construct a sequence of intervals $[a_k, a_k + L_k]$ such that

$$\begin{aligned} \bigcup_{k=1}^{\infty} [a_k, a_k + L_k] &= [\lfloor x \rfloor, \infty), \\ [a_k + L_k - 1, a_k + L_k] \cap X &\neq \emptyset \\ f &\equiv 0 \text{ on } [a_k, a_k + L_k] \text{ for all } k \in \mathbb{N}. \end{aligned}$$

We start with $a_1 = \lfloor x \rfloor$ and $L_1 = L$ from Step 1. Assume that we have already constructed K intervals with this property by induction. By the induction hypothesis, there exists a point

$$\tilde{x} \in [a_K + L_K - 1, a_K + L_K] \cap X.$$

Set $a_{K+1} := \lfloor \tilde{x} \rfloor$. By Step 1, there exists an $L_{K+1} \geq 2$ such that $f = 0$ on the interval $[a_{K+1}, a_{K+1} + L_{K+1}]$ and $[a_{K+1} + L_{K+1} - 1, a_{K+1} + L_{K+1}] \cap X \neq \emptyset$. The sequence $a_K + L_K$ is unbounded, because $a_{K+1} + L_{K+1} \geq L_K + a_K - 1 + 2$. Consequently,

$$\bigcup_{k=1}^K [a_k, a_k + L_k] = [\lfloor x \rfloor, a_{K+1} + L_{K+1}].$$

Since X is unbounded from below and $a_K + L_K$ is unbounded, f vanishes on the union $\bigcup_{x \in X, K \rightarrow \infty} [\lfloor x \rfloor, a_K + L_K] = \mathbb{R}$. This shows $f \equiv 0$. $\square$

We conclude the section with the proof of the sampling theorem.

Theorem 6.3 (Maximum Gap Theorem). *Let φ be a PEB-spline of order $m \geq 2$ and let $X \subseteq \mathbb{R}$ be a separated set with multiplicity function $\mu_X : X \rightarrow \{0, \dots, m-1\}$. If the multiplicity function satisfies*

$$\text{dist}(\{x \in X : \mu_X(x) = m-1\}, \mathbb{Z}) > 0$$

and the weighted maximum gap of (X, μ_X) satisfies

$$\text{mg}(X, \mu_X) < 1,$$

then (X, μ_X) is a sampling set for $V^p(\varphi)$, $1 \leq p \leq \infty$.

Proof. By Corollary 5.3, it suffices to prove that every weak limit of integer translates is a uniqueness set for $V^\infty(\varphi)$. Let $(Y, \mu_Y) \in W_{\mathbb{Z}}(X, \mu_X)$ be a weak limit of integer translates of (X, μ_X) . Lemma 5.4 asserts that $\mathbf{mg}(Y, \mu_Y) < 1$. By Theorem 6.2, (Y, μ_Y) is a uniqueness set for $V^\infty(\varphi)$, and we are done. $\square$

Constant multiplicity. Assume that $\mu_X \equiv s < m-1$, i.e., we sample the same number of derivatives at every point $x \in X$. In this case,

$$\mathbf{mg}(X, \mu_X) = \frac{\mathbf{mg}(X)}{s+1}.$$

We then obtain the following consequence.

Corollary 6.4. *Let φ be a PEB-spline of order $m \geq 2$ and let $X \subseteq \mathbb{R}$ be a separated set. Assume that the multiplicity function is constant $\mu_X(x) = s < m-1$ for all $x \in X$. If the maximum gap of X satisfies*

$$\mathbf{mg}(X) < s+1,$$

then (X, μ_X) is a sampling set for $V^p(\varphi)$, $1 \leq p \leq \infty$.

In particular, if $\mu_X \equiv 0$ and $\mathbf{mg}(X) < 1$, then X is a sampling set for $V^p(\varphi)$, $1 \leq p \leq \infty$.

For B -splines and multiplicity $\mu_X \equiv 0$, this sampling theorem was already proved in [4]. The method with weak limits yields a different proof.

For uniform sampling on $X = \alpha\mathbb{Z}$, the maximum gap is $\mathbf{mg}(\alpha\mathbb{Z}) = \alpha$, and $\alpha < s+1$ guarantees the sampling inequality

$$\sum_{k \in \mathbb{Z}} \sum_{j=0}^s |f^{(j)}(\alpha k)|^p \asymp \|f\|_p^p \quad \text{for all } f \in V^p(\varphi).$$

Note that the weighted density in this case is $D^-(\alpha\mathbb{Z}, \mu) = \frac{s+1}{\alpha} > 1$. As α approaches $s+1$, we obtain sampling sets of density arbitrarily close to the critical density. The condition $\alpha < s+1$ cannot be improved. The set $\alpha\mathbb{Z}$ for $\alpha > s+1$ fails to have the necessary density and is thus not a sampling set⁷. In this case, Theorem 6.3 is optimal and cannot be improved.

7. DISCUSSION OF THE METHODS

Comparison. Both theorems (Theorems 4.2 and 6.3) are based on similar ideas. We choose a suitable covering of $\mathbb{R}$ and study sampling locally on each interval of the partition by means of the active collocation matrix.

In Theorem 4.2 we control the distance of a sampling point to the boundary of the support of φ as a security margin. This parameter guarantees uniform bounds for the inversion of the local collocation matrices (Step 2 in the proof).

In Theorem 6.3, we omit that condition, and the Schoenberg-Whitney conditions imply only the invertibility of the local collocation matrices. As a consequence, we obtain only

⁷One can show that there is always some $x_0 \in \mathbb{R}$, such that $x_0 + (s+1)\mathbb{Z}$ fails to be sampling.

uniqueness sets. In this case, the theory of weak limits and subtle criteria for sampling sets help to derive stability estimates.

Both main theorems are optimal in some sense, but their assumptions are not comparable. To see this, let us consider a few simple examples. In all examples, the generator φ is a PEB-spline of order m with support in $[0, m]$.

Example 7.1. Let $X = \alpha\mathbb{Z}$ and consider $\mu_X(\alpha j) = \frac{1}{2}(1 + (-1)^j)$ so that $\mu_X(\alpha j) = 1$ for even j and $\mu_X(\alpha j) = 0$ for odd j . The weighted maximum gap is then $\mathbf{mg}(X, \mu_X) = \alpha$. Consequently, by Theorem 6.3 $\alpha < 1$ implies the sampling inequality

$$\sum_{j \in \mathbb{Z}} |f(\alpha j)|^p + \sum_{j \in \mathbb{Z}} |f'(2\alpha j)|^p \asymp \|f\|_p^p \quad \text{for all } f \in V^p(\varphi).$$

However, in an interval of length K there are $K/(2\alpha)$ samples with multiplicity 0 and $K/(2\alpha)$ samples with multiplicity 1 (up to a bounded error term), so the weighted Beurling density of (X, μ_X) is

$$D^-(X, \mu_X) = \frac{3}{2\alpha} > \frac{3}{2},$$

i.e., the maximum gap does not deliver an optimal estimate.

By contrast, by Theorem 4.4, for every $\varepsilon > 0$, using the conditions of Theorem 4.2, we can produce a set (X, μ_X) with weighted Beurling density $D^-(X, \mu_X) < 1 + \varepsilon$ that is sampling. Hereby, μ_X can be chosen to be an alternating function as above, and the explicit construction in the proof of Theorem 4.4 provides a subset of a lattice.

Example 7.2. We provide an example of a set uniformly bounded away from $\mathbb{Z}$ with a constant multiplicity function $\mu \equiv S-1 > 0$ and maximum gap $\mathbf{mg}(X) = S \mathbf{mg}(X, \mu_X) < S$ that does not satisfy the assumptions of Theorem 4.2. While the combinatorial theorem relied heavily on *overlaps*, Theorem 4.2 asks for a *partition* in intervals $I(k) = I_{M,L}(k)$. The last assumption (4.4) implies that there exists a sampling point, namely $x_1^k \in X \cap I_{M,L}(k)$, in the first subinterval of $I_{M,L}(k)$ of length 1;

$$\begin{aligned} & [M + kL, M + kL + L) \cap [0 - m + M + kL + 1 + \varepsilon, M + kL + 1 - \varepsilon) \\ &= [M + kL, M + kL + 1 - \varepsilon) \subseteq [M + kL, M + kL + 1) =: J_{M,L}(k). \end{aligned}$$

Note that M, L are supposed to be chosen globally. Furthermore, for any fixed L , one can always assume that $0 \leq M < L$, as $I_{M+L,L}(k) = I_{M,L}(k+1)$. Hence, to prove that the claimed set exists, it suffices to construct a separated set $X \subseteq \mathbb{R}$ with maximum gap $\mathbf{mg}(X) < 2$ and $\text{dist}(X, \mathbb{Z}) > 0$ such that for all $M \in \mathbb{Z}$, $L \in \mathbb{N}$, there exists an integer $k \in \mathbb{Z}$ with $X \cap J_{M,L}(k) = \emptyset$. The set can be constructed inductively by starting with a sufficiently dense lattice X_0 , e.g., $X_0 = \frac{1}{n}\mathbb{Z}$ for some $n \in \mathbb{N}$, $n \geq 2$, and for each $L \in \mathbb{N}$ and each $0 \leq M \leq L$ removing the first n points of X_0 in the interval $I_{M,L}((M+L)2^{2^L})$ (these are the points $X \cap J_{M,L}(k)$), resulting in a new sampling set X given by

$$X := X_0 \setminus \left(\bigcup_{L \in \mathbb{N}} \bigcup_{M=0}^{L-1} X \cap J_{M,L}((M+L)2^{2^L}) \right).$$

One can verify that each of the distinguished intervals is sufficiently far apart from the others so that the maximum gap is $\mathbf{mg}(X) = 1 + \frac{1}{n} < 2$. By construction of the set, there are no integers $M, L \in \mathbb{Z}$ that provide a suitable partition of the real axis where Theorem 4.2 is applicable. However, this is also a sampling set because the weighted maximum gap is $\mathbf{mg}(X, \mu_X) = \frac{1}{S} \mathbf{mg}(X) < \frac{2}{S} \leq 1$, so the conditions of Theorem 6.3 are satisfied.

8. IMPLICATIONS FOR GABOR SYSTEMS

In the final section, we exploit the connection between sampling in shift-invariant spaces and Gabor frames and improve a result in [24].

We denote with $\pi(x, \omega)$, $(x, \omega) \in \mathbb{R}^2$, the time-frequency shift (operator) acting on functions as $\pi(x, \omega)f(t) = e^{2\pi i \omega t} f(t - x)$. In terms of stable expansions, given a *window function* φ and a discrete set $\Lambda \subseteq \mathbb{R}^2$, one asks when a *Gabor system* $\mathcal{G}(\varphi, \Lambda)$, defined as

$$\mathcal{G}(g, \Lambda) = \{\pi(\lambda)\varphi : \lambda \in \Lambda\},$$

is a (*Gabor*) *frame*, i.e., it satisfies the frame inequality

$$A \|f\|_2^2 \leq \sum_{\lambda \in \Lambda} |\langle f, \pi(\lambda)\varphi \rangle|^2 \leq B \|f\|_2^2, \quad f \in L^2(\mathbb{R}),$$

with frame bounds $0 < A \leq B < \infty$ independent of $f \in L^2(\mathbb{R})$. From the general frame theory, the inequality implies a reconstruction formula

$$f = \sum_{\lambda \in \Lambda} \langle f, \pi(\lambda)\varphi \rangle \psi_\lambda = \sum_{\lambda \in \Lambda} \langle f, \psi_\lambda \rangle \pi(\lambda)\varphi, \quad f \in L^2(\mathbb{R})$$

with $(\langle f, \pi(\lambda)\varphi \rangle)_{\lambda \in \Lambda}, (\langle f, \psi_\lambda \rangle)_{\lambda \in \Lambda} \in \ell^2(\Lambda)$ and $\psi_\lambda \in L^2(\mathbb{R})$ for all $\lambda \in \Lambda$. Therefore, given φ , we are interested in determining those sets $\Lambda \subseteq \mathbb{R}^2$ whose Gabor system $\mathcal{G}(\varphi, \Lambda)$ is a frame.

While there is no manageable tool available for arbitrary point configurations, semi-regular sets of type $\Lambda = X \times \mathbb{Z}$ can be treated by the connection between sampling in shift-invariant spaces and Gabor frames [20, Thm. 3.1, Thm. 3.3].

Theorem 8.1. *Assume that $\varphi \in C(\mathbb{R})$ decays as $|\varphi(x)| \lesssim (1 + |x|)^{-1-\varepsilon}$ for an $\varepsilon > 0$, and has stable integer shifts. Let $X \subseteq \mathbb{R}$ be a separated set. Then the following are equivalent:*

- (i) *The family $\mathcal{G}(\varphi, (-X) \times \mathbb{Z})$ is a frame for $L^2(\mathbb{R})$.*
- (ii) *X is a sampling set of $V^p(\varphi)$ for some $p \in [1, \infty]$.*
- (iii) *X is a sampling set of $V^p(\varphi)$ for all $p \in [1, \infty]$.*

The following corollary is a small extension of a result in [4] from B-splines to PEB-splines.

Corollary 8.2. *Let φ be a PEB-spline of order $m \geq 2$. Assume $X \subseteq \mathbb{R}$ is a separated set satisfying the following property:*

- (*) *There exist integers $M, L > 0$ and $\varepsilon > 0$, such that for every $k \in \mathbb{Z}$, there exist points $x_1^k < x_2^k < \dots < x_{L+m-1}^k$ in $X \cap I_{M,L}(k)$ with*

$$x_j^k \in [M + kL + j - m + \varepsilon, M + kL + j - \varepsilon].$$

Then $\mathcal{G}(\varphi, (-X) \times \mathbb{Z})$ is a Gabor frame. In particular, if X is separated and $\mathbf{mg}(X) < 1$, then $\mathcal{G}(\varphi, X \times \mathbb{Z})$ is a Gabor frame.

Proof. We proved that φ has stable integer shifts in the proof of Theorem 4.2. The sampling property of X follows from Theorem 4.2, applied to $(X, 0)$, and Corollary 4.3. The claim follows from Theorem 8.1. $\square$

A direct consequence of the last corollary is the lattice case.

Corollary 8.3. *Let φ be a PEB-spline of order $m \geq 2$. Then $\mathcal{G}(\varphi, \alpha\mathbb{Z} \times \mathbb{Z})$ is a Gabor frame if and only if $0 < \alpha < 1$.*

Proof. The sufficient part is due to Corollary 8.2. The necessary part is due to the Balian-Low theorem, see, e.g., [9]. $\square$

9. APPENDIX

We sketch the proof of (an extended version of) Corollary 5.3. The only changes are required for the last, right-continuous $\varphi^{(s)}$.

Corollary 9.1. *Let $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ be a compactly supported, piecewise continuous function with finitely many jump discontinuities $\mathcal{J} \subseteq \mathbb{R}$. Further, assume that φ has stable integer translates. Let X be a separated set satisfying*

$$\varepsilon := \text{dist}(X, \mathcal{J} + \mathbb{Z}) > 0.$$

Then the following statements are equivalent:

- (i) X is a sampling set for $V^p(\varphi)$ for some $p \in [1, \infty]$.
- (ii) X is a sampling set for $V^p(\varphi)$ for all $p \in [1, \infty]$.
- (iii) Every weak limit $Y \in W_{\mathbb{Z}}(X)$ is a sampling set for $V^\infty(\varphi)$.
- (iv) Every weak limit $Y \in W_{\mathbb{Z}}(X)$ is a uniqueness set for $V^\infty(\varphi)$.

Proof sketch. The sampling property can be expressed in terms of the pre-Gramian matrices

$$P_X(\varphi) := (\varphi(x - k))_{x \in X, k \in \mathbb{Z}}.$$

Precisely, X is sampling for $V^p(\varphi)$ if and only if $P_X(\varphi)$ is bounded above and below on $\ell^p(\mathbb{Z})$. By Schur's test, $P_X : \ell^p(\mathbb{Z}) \rightarrow \ell^p(X)$ is a bounded linear operator for all $p \in [1, \infty]$. We make use of a non-commutative version of Wiener's Lemma [19, Prop. 8.1].

Proposition 9.2. *Let $X, Y \subseteq \mathbb{R}$ be relatively separated and $A \in \mathbb{C}^{X \times Y}$ be a matrix such that*⁸

$$|A_{x,y}| \lesssim (1 + |x - y|)^{-1-\varepsilon} \quad x \in X, y \in Y \quad \text{for some } \varepsilon > 0.$$

The operator A is bounded below on some $\ell^{p_0}(Y)$, $1 \leq p_0 \leq \infty$, i.e., $\|Ac\|_{p_0} \geq C_0 \|c\|_{p_0}$ for all $c \in \ell^{p_0}(Y)$, if and only if A is bounded below on all $\ell^p(Y)$, $1 \leq p \leq \infty$.

⁸The decay conditions can be weakened, e.g., $|A_{x,y}| \lesssim \Theta(x - y)$ for some continuous upper bound Θ in the Wiener-amalgam space $W(\mathbb{R})$.

Since φ is supported on some compact set $K \subseteq \mathbb{R}$ and bounded,

$$|(P_X(\varphi))_{x,k}| \lesssim (1 + |x - k|)^{-1-\varepsilon} \quad x \in X, k \in \mathbb{Z} \quad \text{for all } \varepsilon > 0.$$

This gives the equivalence (i) $\Leftrightarrow$ (ii).

(ii) $\Rightarrow$ (iii) Let $Y \in W_{\mathbb{Z}}(X)$, i.e., $X - k_n \rightarrow Y$ weakly for some $(k_n)_{n \in \mathbb{N}} \subseteq \mathbb{Z}$. Since X is a sampling set, $P_X(\varphi) : \ell^\infty(\mathbb{Z}) \rightarrow \ell^\infty(X)$ is bounded from below, so its pre-adjoint $P_X(g)' : \ell^1(X) \rightarrow \ell^1(\mathbb{Z})$, given by

$$P_X(\varphi)' = (\varphi(x - k))_{k \in \mathbb{Z}, x \in X},$$

is onto by the closed graph theorem. Fix now $c \in \ell^1(\mathbb{Z})$. For every sequence $(c_{k-k_n})_{k \in \mathbb{Z}}$ there exists a sequence $a^n \in \ell^1(X)$ with $\|a^n\|_1 \lesssim \|c\|_1$ and

$$c_{k-k_n} = \sum_{x \in X} a_x^n g(x - k), \quad k \in \mathbb{Z}.$$

The weak convergence can be rewritten by means of distributions. Set $b_{x-k_n}^n = a_x^n$ and $\mu_n := \sum_{z \in X-k_n} b_z^n \delta_z$. Then $\mu_n \in C_0^*(\mathbb{R})$ has support in $X - k_n$ and satisfies

$$\|\mu_n\|_{C_0^*(\mathbb{R})} = \|b^n\|_1 = \|a^n\|_1 \lesssim \|c\|_1.$$

By restricting to a subsequence and using the w^* -completeness of $C_0^*(\mathbb{R})$, μ_n converges to $\mu \in C_0^*(\mathbb{R})$ in $\sigma(C_0^*(\mathbb{R}), C_0(\mathbb{R}))$. Since $X - k_n \rightarrow Y$, the weak limit Y is relatively separated and $\text{supp}(\mu) \subseteq Y$ (see [19, Lem. 4.3 and 4.5]). Furthermore, $\mu = \sum_{y \in Y} b_y \delta_y$ for suitable $b \in \ell^1(Y)$,

$$\|b\|_1 = \|\mu\|_{C_0^*(\mathbb{R})} \lesssim \liminf_{n \rightarrow \infty} \|\mu_n\|_{C_0^*(\mathbb{R})}.$$

Here comes the decisive modification. By assumption, $\text{dist}(X - k_n, \mathcal{J} + \mathbb{Z}) = \text{dist}(X, \mathcal{J} + \mathbb{Z}) > 0$ for all $n \in \mathbb{Z}$ and therefore $\text{dist}(Y, \mathcal{J} + \mathbb{Z}) > 0$. Now we replace the generator φ (which has discontinuities $\mathcal{J}$) by a continuous compactly supported function φ_0 such that $\varphi = \varphi_0$ on $\mathbb{R} \setminus (\mathcal{J} + \mathbb{Z} + (-\varepsilon, \varepsilon))$ for an $\varepsilon \in (0, \text{dist}(X, \mathcal{J} + \mathbb{Z}))$. In particular, $\varphi = \varphi_0$ on Y and on $X - k_n$ for all $n \in \mathbb{N}$. Hence we obtain

$$\begin{aligned} \sum_{y \in Y} b_y \varphi(y - k) &= \int_{\mathbb{R}} \varphi_0(x - k) d\mu(x) = \lim_{n \rightarrow \infty} \int_{\mathbb{R}} \varphi_0(x - k) d\mu_n(x) \\ &= \lim_{n \rightarrow \infty} \sum_{z \in X - k_n} b_z^n \varphi_0(z - k) = \lim_{n \rightarrow \infty} \sum_{z \in X - k_n} b_z^n \varphi(z - k) = c_k. \end{aligned}$$

(iii) $\Rightarrow$ (iv) is trivial. To obtain (iv) $\Rightarrow$ (ii), one proceeds precisely as in [19], with the analogous modification with φ_0 when it comes to the limits. $\square$

The lift to sampling with multiplicities in Corollary 5.3 is the same as in [21].

REFERENCES

- [1] ADCOCK, B., GATARIC, M., AND HANSEN, A. C. Density theorems for nonuniform sampling of bandlimited functions using derivatives or bunched measurements. *J. Fourier Anal. Appl.* 23, 6 (2017), 1311–1347.
- [2] ADCOCK, B., AND SUI, Y. Compressive Hermite interpolation: sparse, high-dimensional approximation from gradient-augmented measurements. *Constr. Approx.* 50, 1 (2019), 167–207.
- [3] ALDROUBI, A., AND FEICHTINGER, H. G. Exact iterative reconstruction algorithm for multivariate irregularly sampled functions in spline-like spaces: the L^p -theory. *Proc. Amer. Math. Soc.* 126, 9 (1998), 2677–2686.
- [4] ALDROUBI, A., AND GRÖCHENIG, K. Beurling-Landau-type theorems for non-uniform sampling in shift invariant spline spaces. *J. Fourier Anal. Appl.* 6, 1 (2000), 93–103.
- [5] ALDROUBI, A., AND GRÖCHENIG, K. Nonuniform Sampling and Reconstruction in Shift-Invariant Spaces. *SIAM Rev.* 43, 4 (2001), 585–620.
- [6] BANNERT, S., GRÖCHENIG, K., AND STÖCKLER, J. Discretized Gabor Frames of Totally Positive Functions. *IEEE Trans. Inf. Theory* 60, 1 (2014), 159–169.
- [7] A. Ben-Artzi and A. Ron. Translates of exponential box splines and their related spaces. *Trans. Amer. Math. Soc.*, 309(2):683–710, 1988.
- [8] BENEDETTO, J. J., AND FERREIRA, P. J. S. G., Eds. *Modern sampling theory*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Inc., Boston, MA, 2001. Mathematics and applications.
- [9] BENEDETTO, J. J., HEIL, C., AND WALNUT, D. F. Differentiation and the Balian-Low theorem. *J. Fourier Anal. Appl.* 1, 4 (1995), 355–402.
- [10] BEURLING, A. *The collected works of Arne Beurling. Volume 1: Complex analysis. Volume 2: Harmonic analysis*. Ed. by Lennart Carleson, Paul Malliavin, John Neuberger, John Wermer. Contemp. Mathematicians. Birkhauser, 1989.
- [11] CHRISTENSEN, O., AND MASSOPUST, P. Exponential B-splines and the partition of unity property. *Adv. Comput. Math.* 37, 3 (2011), 301–318.
- [12] DE BOOR, C. *A practical guide to splines*, vol. 27 of *Applied Mathematical Sciences*. Springer-Verlag, New York-Berlin, 1978.
- [13] DE BOOR, C., DEVORE, R. A., AND RON, A. The structure of finitely generated shift-invariant spaces in $L_2(\mathbf{r}^d)$. *J. Funct. Anal.* 119, 1 (1994), 37–78.
- [14] A. Antony Selvan and R. Radha. An optimal result for sampling density in shift-invariant spaces generated by Meyer scaling function. *J. Math. Anal. Appl.*, 451(1):197–208, 2017.
- [15] GHOSH, R., AND SELVAN, A. A. Sampling and interpolation of periodic nonuniform samples involving derivatives. *Results Math.* 78, 2 (2023), Paper No. 52, 20.
- [16] GRÖCHENIG, K. Reconstruction algorithms in irregular sampling. *Math. Comp.* 59, 199 (1992), 181–194.
- [17] GRÖCHENIG, K., ROMERO, J. L., AND STÖCKLER, J. Sharp Results on Sampling with Derivatives in Shift-Invariant Spaces and Multi-Window Gabor Frames. *Constr. Approx.* 51, 1 (2020), 1–25.
- [18] GRÖCHENIG, K., AND SCHWAB, H. Fast local reconstruction methods for nonuniform sampling in shift-invariant spaces. *SIAM J. Matrix Anal. Appl.* 24, 4 (2003), 899–913 (electronic).
- [19] GRÖCHENIG, K., ORTEGA-CERDÀ, J., AND ROMERO, J. L. Deformation of Gabor systems. *Adv. Math.* 277 (2015), 388–425.
- [20] GRÖCHENIG, K., ROMERO, J. L., AND STÖCKLER, J. Sampling theorems for shift-invariant spaces, Gabor frames, and totally positive functions. *Invent. Math.* 211, 3 (2017), 1119–1148.
- [21] GRÖCHENIG, K., ROMERO, J. L., AND STÖCKLER, J. Sharp Results on Sampling with Derivatives in Shift-Invariant Spaces and Multi-Window Gabor Frames. *Constr. Approx.* 51, 1 (2019), 1–25.
- [22] JETTER, K., AND ZHOU, D. X. Order of linear approximation from shift-invariant spaces. *Constr. Approx.* 11, 4 (1995), 423–438.

- [23] JIA, R.-Q., AND MICCHELLI, C. A. Using the Refinement Equations for the Construction of Pre-Wavelets II: Powers of Two. In *Curves and Surfaces*. Elsevier, 1991, pp. 209–246.
- [24] KLOOS, T., AND STÖCKLER, J. Zak transforms and Gabor frames of totally positive functions and exponential B-splines. *J. Approx. Theory* 184 (2014), 209–237.
- [25] KYRIAZIS, G. C. Approximation from shift-invariant spaces. *Constr. Approx.* 11, 2 (1995), 141–164.
- [26] LANDAU, H. J. Necessary density conditions for sampling and interpolation of certain entire functions. *Acta Math.* 117 (1967), 37–52.
- [27] LANDAU, H. J. Necessary density conditions for sampling and interpolation of certain entire functions. *Acta Math.* 117 (1967), 37–52.
- [28] MALLAT, S. G. Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(\mathbb{R})$. *Trans. Am. Math. Soc.* 315, 1 (1989), 69.
- [29] PAWLOWSKI, A., GUZMAN, J. L., RODRIGUEZ, F., BERENGUEL, M., SANCHEZ, J., AND DORMIDO, S. Study of event-based sampling techniques and their influence on greenhouse climate control with wireless sensors network. In *Factory Automation*, J. Silvestre-Blanes, Ed. IntechOpen, Rijeka, 2010, ch. 14.
- [30] RAWN, M. D. A stable nonuniform sampling expansion involving derivatives. *IEEE Trans. Inform. Theory* 35, 6 (1989), 1223–1227.
- [31] RAZAFINJATOVO, H. N. Iterative reconstructions in irregular sampling with derivatives. *J. Fourier Anal. Appl.* 1, 3 (1995), 281–295.
- [32] A. Ron. Exponential box splines. *Constr. Approx.*, 4(4):357–378, 1988.
- [33] RON, A. Introduction to shift-invariant spaces. Linear independence. In *Multivariate Approximation and Applications*. Cambridge University Press, 2001, pp. 112–151.
- [34] SCHUMAKER, L. *Spline Functions: Basic Theory*, third ed. Cambridge University Press, 2007.
- [35] SEIP, K. *Interpolation and sampling in spaces of analytic functions*, vol. 33 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2004.
- [36] SELVAN, A. A., AND GHOSH, R. Sampling with derivatives in periodic shift-invariant spaces. *Numer. Funct. Anal. Optim.* 43, 13 (2022), 1591–1615.

3.2 On the frame property of Hermite functions and analytic extensions of their frame sets

Bibliographic Information

Faulhuber, M., Shafkulovska, I., Zlotnikov, I., On the frame property of Hermite functions and analytic extensions of their frame sets, *submitted* (2024)

Submission date

24 March 2024

Contribution

Markus Faulhuber, Irina Shafkulovska, and Ilia Zlotnikov contributed equally to the idea, concept, development, and writing of the paper.

E-Print

arXiv:2403.10503

ON THE FRAME PROPERTY OF HERMITE FUNCTIONS AND ANALYTIC EXTENSIONS OF THEIR FRAME SETS

MARKUS FAULHUBER, IRINA SHAFKULOVSKA, AND ILYA ZLOTNIKOV

ABSTRACT. We study Gabor frames with Hermite window functions. Gröchenig and Lyubarskii provided a sufficient density condition for their frame sets, which leads to what we call the “safety region”. For rectangular lattices and Hermite windows of order 4 and higher, we enlarge this safety region by providing new points on the boundary of this region. For this purpose, we employ the Janssen representation of the frame operator to compare its distance to the identity in the operator norm. The calculations lead to estimates on series involving Laguerre polynomials with Gaussian weight functions.

1. INTRODUCTION AND NOTATION

1.1. Introduction. The study of frame sets has seen breakthroughs and surprises in recent years. For Hermite functions, there was hope that the structure of their frame sets may be simple and can be fully understood soon. However, *The Mystery of Gabor frames*, after the eponymous article [17] by Gröchenig, is an appropriate description of our understanding of Gabor systems with Hermite functions. Our results contribute to this puzzle.

In this work, we are concerned with frame sets of Hermite functions. This means that we seek to find rectangular lattices $\Lambda \subset \mathbb{R}^2$ such that the resulting Gabor system is a frame. The Hermite function h_n of order n , scaled to unit norm, is (cf. [13, Chap. 1.7], [20])

$$h_n(t) = c_n (-1)^n e^{\pi t^2} \frac{d^n}{dt^n} e^{-2\pi t^2}, \quad t \in \mathbb{R}, \quad \text{so} \quad \|h_n\|_{L^2(\mathbb{R})} = 1, \quad n \in \mathbb{N}_0,$$

where $c_n = 2^{1/4} / \sqrt{n!(2\pi)^n 2^n}$. In particular, the n -th Hermite function is a Gaussian function multiplied with the n -th Hermite polynomial and $n = 0$ gives the Gaussian function

$$h_0(t) = 2^{1/4} e^{-\pi t^2}.$$

Gabor systems with Hermite window functions were studied, e.g., by Gröchenig and Lyubarskii [18], [19]. We recall one of the main results from [18].

Theorem (Gröchenig, Lyubarskii). *Consider the Gabor system $\mathcal{G}(h_n, \Lambda)$. If the density of the lattice Λ exceeds $n + 1$, then $\mathcal{G}(h_n, \Lambda)$ is a Gabor frame for $L^2(\mathbb{R})$.*

Existing results, which we will explain shortly, show in particular the following:

the Gabor system $\mathcal{G}_n = \mathcal{G}(h_n, (1/\sqrt{n+1})\mathbb{Z}^2)$ is *not* a frame for $n = 0, 1, 2, 3$.

It seems natural to expect that this pattern continues, but, again, the frame set of Hermite functions keeps being mysterious. This is underlined by our upcoming main result.

FACULTY OF MATHEMATICS, UNIVERSITY OF VIENNA

OSKAR-MORGENSTERN-PLATZ 1, 1090 VIENNA, AUSTRIA

E-mail addresses: markus.faulhuber@univie.ac.at, irina.shafkulovska@univie.ac.at, ilia.zlotnikov@univie.ac.at.

2020 *Mathematics Subject Classification.* 42C15; 33C45.

Key words and phrases. Gabor frame, Hermite function, Laguerre polynomial.

This research was funded in whole or in part by the Austrian Science Fund (FWF) [10.55776/P33217].

I. Z. is grateful to I. Krasikov for fruitful discussions.

Theorem 1.1 (Main result). *Denote by $\mathcal{G}_n = \mathcal{G}(h_n, (1/\sqrt{n+1})\mathbb{Z}^2)$ the Gabor system for the n -th Hermite function and square lattice of density $n+1$. Then $\mathcal{G}_n$ is a Gabor frame for $L^2(\mathbb{R})$ if and only if $n \geq 4$.*

As Theorem 1.1 shows, the opposite of the expected pattern holds true. So, at least for square lattices, equality can be achieved in the Theorem of Gröchenig and Lyubarskii once $n \geq 4$. The case $n = 4$ is also the first Gabor system of the form

$$\mathcal{G}_n = \mathcal{G}(h_n, (1/\sqrt{n+1})\mathbb{Z}^2) = \{\pi(\lambda)h_n \mid \lambda \in (1/\sqrt{n+1})\mathbb{Z}^2\},$$

which has not been treated in the literature so far. We will restrict our attention to the frame set of h_n for rectangular lattices of the form $\Lambda = a\mathbb{Z} \times b\mathbb{Z}$.

The case of the Gaussian function, i.e., $n = 0$, has been completely settled (for general point sets) independently by Lyubarskii [28] and in the articles of Seip [33] and Seip and Wallstén [34]. The case of the first Hermite function (and more generally odd functions) was treated by Lyubarskii and Nes [29] for rectangular lattices and extended to general lattices by one of the authors in [6].

In [17], Gröchenig came up with the following question, which is now called the frame set conjecture for Hermite functions:

- For h_n with $n \in 2\mathbb{N}$ is the Gabor system $\mathcal{G}(h_n, \Lambda)$ always a frame when $\text{vol}(\Lambda) < 1$?
- For h_n with $n \in 2\mathbb{N} - 1$ is the Gabor system $\mathcal{G}(h_n, \Lambda)$ always a frame when $\text{vol}(\Lambda) < 1$ and $\text{vol}(\Lambda) \neq \frac{N}{N+1}$, $N \in \mathbb{N}$?

The questions were originally posed for rectangular lattices and are based on the restriction found in [29]. Due to the extension given in [6] they may be formulated for arbitrary lattices. However, the frame set conjecture for Hermite functions was answered negatively by Lemvig [26], who provided first counter-examples. More exceptions by Horst, Lemvig, and Videbaek [20] followed. A sufficiency criterion with numerical investigations was presented by Ghosh and Selvan [14].

This work is structured as follows. In Sec. 1.2 we introduce the general notation used in this work. In Sec. 2, we recall essential results from Gabor analysis in Sec. 2.1, provide necessary details on Laguerre polynomials in Sec. 2.2 and introduce auxiliary results and outline the rough proof strategy in Sec. 2.3. Sec. 3 is devoted to new results on the frame set of Hermite functions. Our main result, Theorem 1.1, is proved in Sec. 3.1. The limitations of our approach are discussed in Sec. 3.2. Finally, in Sec. 3.3, we present more applications of the approach we used to prove Theorem 1.1. We enlarge what we call the safety region for h_{15} and provide a new neighborhood in the frame set of h_9 .

1.2. Notation. We mainly follow the notation from the textbook of Gröchenig [16]. In this work, as usual in time-frequency analysis, we use the following Fourier transform of a suitable function on $\mathbb{R}$:

$$\mathcal{F}f(\omega) = \int_{\mathbb{R}} f(t)e^{-2\pi i\omega t} dt.$$

This extends to a unitary operator on $L^2(\mathbb{R})$, hence we have Plancherel's theorem and Parseval's identity, respectively, given by

$$\|f\|_{L^2(\mathbb{R})} = \|\mathcal{F}f\|_{L^2(\mathbb{R})} \quad \text{and} \quad \langle f, g \rangle = \langle \mathcal{F}f, \mathcal{F}g \rangle.$$

The inner product and norm on $L^2(\mathbb{R})$ are, respectively, given by

$$\langle f, g \rangle = \int_{\mathbb{R}} f(t) \overline{g(t)} dt \quad \text{and} \quad \|f\|_{L^2(\mathbb{R})}^2 = \langle f, f \rangle.$$

Sometimes, we will pass an index to the inner product, denoting the specific Hilbert space, which will be either $L^2(\mathbb{R})$ or $L^2(\mathbb{R}^2)$. Note that the Hermite functions are eigenfunctions of the Fourier transform, in particular we have

$$(1.1) \quad \mathcal{F}h_n = (-i)^n h_n.$$

A Gabor system is a set of functions of the form

$$\mathcal{G}(g, \Lambda) = \{\pi(\lambda)g \mid \lambda \in \Lambda\},$$

where $g \in L^2(\mathbb{R})$ (of course, not the zero function throughout this work), $\Lambda \subset \mathbb{R}^2$ discrete and $\pi(z)$ is a time-frequency shift by z , which is a unitary operator, acting by the rule

$$\pi(z)g = M_\omega T_x g, \quad z = (x, \omega) \in \mathbb{R}^2.$$

Here, T_x and M_ω denote the usual operators of translation (time-shift) and modulation (frequency-shift), which act on functions by the rules

$$T_x g(t) = g(t - x) \quad \text{and} \quad M_\omega g(t) = e^{2\pi i \omega t} g(t),$$

with the commutation relation

$$(1.2) \quad M_\omega T_x = e^{2\pi i \omega x} T_x M_\omega.$$

Often Λ is assumed to be a lattice in the time-frequency plane, i.e., a discrete co-compact subgroup of $\mathbb{R}^2$. We denote the co-volume of Λ simply by $\text{vol}(\Lambda)$ in this work and will actually only deal with the case of rectangular lattices, i.e., $\Lambda = \Lambda_{a,b} = a\mathbb{Z} \times b\mathbb{Z}$, with $a, b \in \mathbb{R}_+$. We will also identify the lattice $\Lambda_{a,b}$ with the pair of parameters $(a, b) \in \mathbb{R}_+ \times \mathbb{R}_+$. Note that for a rectangular lattice we have

$$\text{vol}(\Lambda_{a,b}) = ab \quad \text{and its density is} \quad \delta = \text{vol}(\Lambda)^{-1} = 1/(ab).$$

For a general lattice $\Lambda = M\mathbb{Z}^2$, $M \in GL(2, \mathbb{R})$, we have $\text{vol}(\Lambda) = |\det(M)|$. For the lattice Λ we denote by Λ° its adjoint lattice, which is $J\Lambda^\perp$, where $\Lambda^\perp$ is the classical dual lattice and J the standard symplectic matrix;

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Note that while the classical dual lattice is defined by the condition

$$\Lambda^\perp = \{\lambda^\perp \mid \lambda^\perp \cdot \lambda \in \mathbb{Z} \forall \lambda \in \Lambda\},$$

the adjoint lattice may be defined by the commutation relations of time-frequency shifts

$$\Lambda^\circ = \{\lambda^\circ \mid \pi(\lambda)\pi(\lambda^\circ) = \pi(\lambda^\circ)\pi(\lambda) \forall \lambda \in \Lambda\}.$$

This is, by (1.2), equivalent to saying that $\lambda^\circ \in \Lambda^\circ$ if and only if $\sigma(\lambda^\circ, \lambda) \in \mathbb{Z}$, for all $\lambda \in \Lambda$, where $\sigma(\lambda^\circ, \lambda) = \lambda^\circ \cdot J\lambda$ denotes the standard symplectic form. Note that for a rectangular lattice $\Lambda_{a,b} = a\mathbb{Z} \times b\mathbb{Z}$ the adjoint lattice is given by $\Lambda_{a,b}^\circ = (1/a)\mathbb{Z} \times (1/b)\mathbb{Z}$.

The adjoint (or symplectic dual) lattice makes the usage of a symplectic Poisson summation formula possible. For a (suitable) function F on $\mathbb{R}^2$, this is

$$(1.3) \quad \sum_{\lambda \in \Lambda} F(\lambda + z) = \text{vol}(\Lambda)^{-1} \sum_{\lambda^\circ \in \Lambda^\circ} \mathcal{F}_\sigma F(\lambda^\circ) e^{2\pi i \sigma(z, \lambda^\circ)},$$

where $\mathcal{F}_\sigma$ is the symplectic Fourier transform [4]

$$\mathcal{F}_\sigma F(z) = \iint_{\mathbb{R}^2} F(w) e^{-2\pi i \sigma(w, z)} dw.$$

The symplectic Poisson summation formula (1.3) is implicitly present throughout the whole article. We mainly apply it to the short-time Fourier Transform (STFT) of Hermite functions. The STFT of a function $f \in L^2(\mathbb{R})$ with window $g \in L^2(\mathbb{R})$ is given by

$$V_g f(x, w) = \int_{\mathbb{R}} f(t) \overline{g(t-x)} e^{-2\pi i w t} dt = \langle f, \pi(z)g \rangle, \quad z = (x, w) \in \mathbb{R}^2.$$

A Gabor system $\mathcal{G}(g, \Lambda)$ is a frame for $L^2(\mathbb{R})$ if and only if there exist positive constants $0 < A \leq B < \infty$, such that the frame inequality holds:

$$A \|f\|_{L^2(\mathbb{R})}^2 \leq \sum_{\lambda \in \Lambda} |\langle f, \pi(\lambda)g \rangle|^2 \leq B \|f\|_{L^2(\mathbb{R})}^2, \quad \forall f \in L^2(\mathbb{R}).$$

The constants A and B are called lower and upper frame bound, respectively, and the sharpest possible bounds are the spectral bounds of the associated Gabor frame operator $S_{g, \Lambda}$, which acts on $f \in L^2(\mathbb{R})$ by

$$S_{g, \Lambda} f = \sum_{\lambda \in \Lambda} \langle f, \pi(\lambda)g \rangle \pi(\lambda)g.$$

The set of all (rectangular) lattices such that the Gabor system $\mathcal{G}(g, \Lambda)$ is a Gabor frame for $L^2(\mathbb{R})$, is called the frame set $\mathfrak{F}$ of g :

$$\mathfrak{F}(g) = \{\Lambda \subset \mathbb{R}^2 \mid \mathcal{G}(g, \Lambda) \text{ is a frame}\}.$$

This is also the terminology used by Gröchenig in [17]. An important result for functions in the modulation space $M^1(\mathbb{R})$, defined in Def. 2.1 just below, is that the frame set is an open subset of $\{\Lambda \mid \text{vol}(\Lambda) < 1\}$ and that it contains a neighborhood of 0 [17, Thm. 2.1] (see also [10], [11]). The latter statement guarantees that Gabor frames with a window in $M^1(\mathbb{R})$ exist if the density of Λ is large enough. For rectangular lattices $\Lambda_{a,b} = a\mathbb{Z} \times b\mathbb{Z}$, we can picture the frame set by identifying a lattice with the pair of parameters $(a, b) \in \mathbb{R}_+ \times \mathbb{R}_+$, as done in Fig. 1. Due to (1.1), the frame set for Hermite functions is symmetric in the sense that $(a, b) \in \mathfrak{F}(h_n)$ if and only if $(b, a) \in \mathfrak{F}(h_n)$. For the n -th Hermite function we call the part of the frame set $\text{vol}(\Lambda) < n+1$ the safety region, as we have a frame by the Theorem of Gröchenig and Lyubarskii [18]. For rectangular lattices $\Lambda_{a,b} = a\mathbb{Z} \times b\mathbb{Z}$, $a, b \in \mathbb{R}_+$, the safety region in $\mathfrak{F}(h_n)$ are all pairs of parameters such that $ab < n+1$.

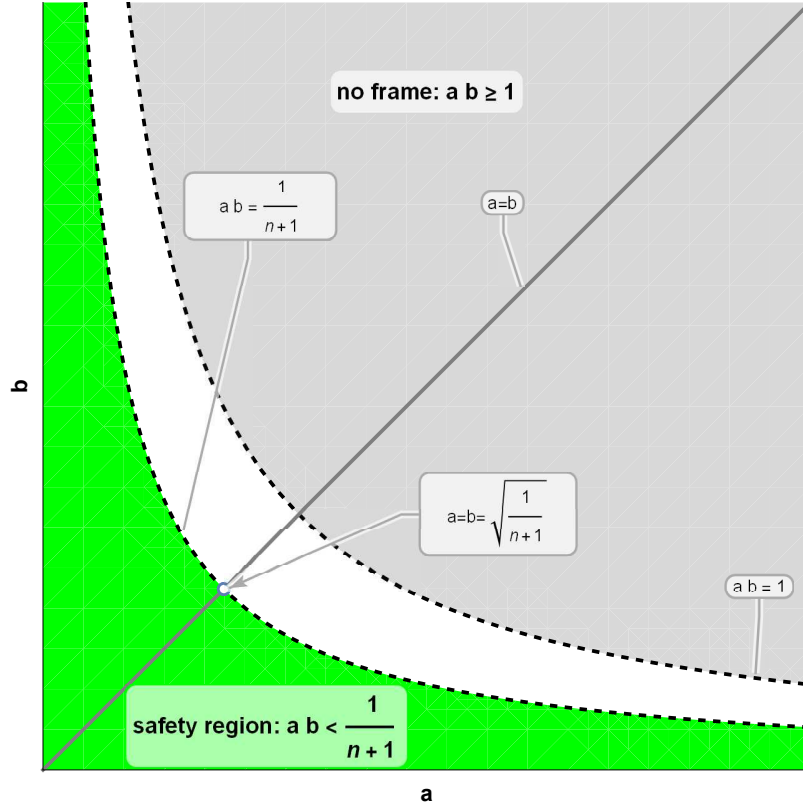


FIGURE 1. For the n -th Hermite function we know that the Gabor system $\mathcal{G}(h_n, a\mathbb{Z} \times b\mathbb{Z})$ is a frame if $ab < 1/(n+1)$ and that it is not a frame if $ab \geq 1$. The frame set structure of $\mathfrak{F}(h_n)$, $n \geq 1$, between the “safety region” and the “no frame” region is currently only partially understood.

2. PRELIMINARIES

2.1. Results from Gabor analysis. We recall some well-established results from Gabor analysis. The modulation space $M^1(\mathbb{R})$, which is also called Feichtinger’s algebra due to its discoverer Feichtinger [8], will provide a convenient set-up for our purposes. There are many equivalent definitions of this space and we refer to the treatise by Jakobsen [21] for details. For the definition of the modulation spaces one needs to fix a window in the Schwartz space $\mathcal{S}(\mathbb{R})$, but all windows lead to the same space with equivalent norms. Hence, we will pick the Gaussian window h_0 in our definition. More details can be found in the textbook of Gröchenig [16, Chap. 11–12].

Definition 2.1 (Modulation space). *The modulation space $M^p(\mathbb{R})$ consists of all tempered distributions $f \in \mathcal{S}'(\mathbb{R})$ such that*

$$\|f\|_{M^p(\mathbb{R})} = \|V_{h_0}f\|_{L^p(\mathbb{R}^2)} < \infty.$$

Note that $V_g g \in M^1(\mathbb{R}^2)$ if $g \in M^1(\mathbb{R})$ [21, Cor. 5.5]. Hence, Feichtinger's algebra is a convenient setting for us, as it contains h_n . Therefore, all manipulations in the sequel are justified and we do not run into any convergence issues (all series converge unconditionally).

The following result was first shown by Janssen [23] for rectangular lattices and for general lattices by Feichtinger and Luef [12]. We also refer to the work of Gröchenig and Koppensteiner [15], where all technical details are clarified.

Proposition 2.2 (Fundamental Identity of Gabor Analysis). *Assume we have functions $f_1, f_2, g_1, g_2 \in M^1(\mathbb{R})$, then the following identity holds*

$$(2.1) \quad \sum_{\lambda \in \Lambda} V_{g_1} f_1(\lambda) \overline{V_{g_2} f_2(\lambda)} = \text{vol}(\Lambda)^{-1} \sum_{\lambda^\circ \in \Lambda^\circ} V_{g_1} g_2(\lambda^\circ) \overline{V_{f_1} f_2(\lambda^\circ)}.$$

The proof uses symplectic Poisson summation (1.3) and the orthogonality relations

$$\langle V_{g_1} f_1, V_{g_2} f_2 \rangle_{L^2(\mathbb{R}^2)} = \langle f_1, f_2 \rangle_{L^2(\mathbb{R})} \overline{\langle g_1, g_2 \rangle_{L^2(\mathbb{R})}}.$$

This is the STFT-version of Moyal's identity for the Wigner distribution introduced in [30] (see also [4, Chap. 6], [16, Chap. 4.3]). We continue with the following observation, which is a slight variation of [16, Lem. 4.2.1] in the textbook of Gröchenig. The proof employs the Cauchy-Schwarz inequality, and when equality can be achieved.

Lemma 2.3. *For $g \in L^2(\mathbb{R})$ we note that $V_g g(0, 0) \in \mathbb{R}$. Moreover, we have*

$$V_g g(0, 0) = \|g\|_{L^2(\mathbb{R})}^2$$

and the following estimate for all $(x, \omega) \in \mathbb{R}^2 \setminus \{(0, 0)\}$:

$$|V_g g(x, \omega)| < V_g g(0, 0).$$

The following notion was introduced in [37]. A function g satisfies *Condition A* if

$$(Condition\ A) \quad \sum_{\lambda^\circ \in \Lambda^\circ} |V_g g(\lambda^\circ)| < \infty.$$

The next result was shown by Tolimieri and Orr [37] under the assumption that the window function g indeed satisfies (Condition A). We only add the convenient choice that we have $g \in M^1(\mathbb{R})$, so that we can directly use the result.

Lemma 2.4. *Let $g \in M^1(\mathbb{R})$, then it satisfies (Condition A) and, consequently, the Gabor system $\mathcal{G}(g, \Lambda)$ has a finite upper frame bound B , which is at most*

$$(2.2) \quad B \leq \text{vol}(\Lambda)^{-1} \sum_{\lambda^\circ \in \Lambda^\circ} |V_g g(\lambda^\circ)| < \infty.$$

The proof of the previous result follows easily from (2.1) and Lemma 2.3.

Lemma 2.5 (Janssen representation). *For a window $g \in M^1(\mathbb{R})$ and a lattice Λ , the frame operator $S = S_{g, \Lambda}$ can be expressed as*

$$S = \text{vol}(\Lambda)^{-1} \sum_{\lambda^\circ \in \Lambda^\circ} \langle g, \pi(\lambda^\circ) g \rangle \pi(\lambda^\circ).$$

The result is well-known and it follows basically from (2.1). The last result in this section is the following sufficiency criterion from [39], also known as the Janssen test. As this will be the basis for all our proofs in Sec. 3, we provide the proof for this result.

Lemma 2.6 (Janssen test). *Let Λ be a lattice and g a window with $\|g\|_{L^2(\mathbb{R})} = 1$. If*

$$(2.3) \quad \sum_{\lambda^\circ \in \Lambda^\circ} |V_g g(\lambda^\circ)| < 2,$$

then $\mathcal{G}(g, \Lambda)$ constitutes a Gabor frame for $L^2(\mathbb{R})$.

Proof. We use the proof given in [39] and compare the scaled frame operator $\text{vol}(\Lambda)S$, with $g \in M^1(\mathbb{R})$ and $\Lambda \subset \mathbb{R}^2$ a lattice, to the identity operator $I = I_{L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})}$. Noting that $\pi(0) = I$, and using the Janssen representation of S , we get

$$\begin{aligned} \|\text{vol}(\Lambda)S - I\|_{L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})} &= \left\| \sum_{\lambda^\circ \in \Lambda^\circ \setminus \{0\}} \langle g, \pi(\lambda^\circ)g \rangle \pi(\lambda^\circ) \right\|_{L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})} \\ &\leq \sum_{\lambda^\circ \in \Lambda^\circ \setminus \{0\}} |\langle g, \pi(\lambda^\circ)g \rangle| = \tilde{B}. \end{aligned}$$

By our assumption (2.3), we see that $\text{vol}(\Lambda)S$ is less than 1 away from the identity operator. Therefore we can build the Neumann series of the inverse and estimate its norm by

$$\|(\text{vol}(\Lambda)S)^{-1}\|_{L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})} \leq \sum_{k=0}^{\infty} \tilde{B}^k = \frac{1}{1 - \tilde{B}},$$

which shows that $\|(\text{vol}(\Lambda)S)^{-1}\| \leq 1/(1 - \tilde{B})$ and so the frame operator is invertible. It follows that the Gabor system has a lower frame bound. The upper frame bound is finite due to (2.2), hence $\mathcal{G}(g, \Lambda)$ is a frame. $\square$

To the best of our knowledge, this tool has so far not been used in rigorous analytic investigations of frame sets. It was introduced in [38] for finite Gabor systems in $\mathbb{C}^L$, $L \in \mathbb{N}$ and then presented in [39, Chap. 3.1, Cor. 3.1] for Gabor systems in $L^2(\mathbb{R})$. In both cases it was combined with numerical investigations to have heuristic arguments for certain optimal sampling strategies. It was again discussed in [7, Sec. 3] for heuristic reasons and in [35, Sec. 6] an extension of the result was presented for Gabor g -frames. It was called “Janssen condition” in [38, Chap. 4.1] as the criterion arises from the Janssen representation of the frame operator and we also refer to [9, Sec. 6] for further reading.

2.2. Laguerre polynomials. Recall that the n -th Laguerre polynomial, denoted by $\mathcal{L}_n$, is defined on the positive real axis $\mathbb{R}_+$ (but can be extended to $\mathbb{R}$) by

$$\mathcal{L}_n(x) = \sum_{k=0}^n \binom{n}{k} \frac{(-x)^k}{k!}.$$

We remark that there is a generalized version $\mathcal{L}_n^\alpha$, which does not concern us here, and that $\mathcal{L}_n = \mathcal{L}_n^0$. For details we refer to the textbook by Szegő [36] or to [1, Chap. 22]. The Laguerre polynomials are orthogonal on $\mathbb{R}_+$ with respect to the exponential weight

$w(x) = e^{-x}$ [36, Sec. 5.1, eq. (5.1.1)] and $\mathcal{L}_n(x)$ has n distinct roots in $\mathbb{R}_+$ [36, Chap. 3.3, Chap. 6.31]. The largest root x_n of $\mathcal{L}_n$ satisfies [36, Thm.6.31.2]

$$x_n < 2n + 1 + \sqrt{(2n + 1)^2 + 1/4}, \quad \text{and} \quad x_n \sim 4n.$$

Our interest in Laguerre polynomials comes from the following result, which may be looked up in [13, Chap. 1.9, Thm. 1.104]. Note that $V_g f$ is defined slightly different in [13, Chap. 1.8, eq. (1.88)] (where it is more or less the ambiguity function), which makes the appearance of a phase factor explicit in our work.

Lemma 2.7. *For the n -th Hermite function h_n , the STFT with window h_n is given by*

$$(2.4) \quad V_{h_n} h_n(x, \omega) = e^{-\pi i x \omega} \mathcal{L}_n(\pi(x^2 + \omega^2)) e^{-\frac{\pi}{2}(x^2 + \omega^2)}.$$

In the sequel, our goal will always be to find Λ such that

$$\sum_{\lambda^\circ \in \Lambda^\circ} |V_{h_n} h_n(\lambda^\circ)| = \sum_{\lambda^\circ \in \Lambda^\circ} |\mathcal{L}_n(\pi|\lambda^\circ|^2)| e^{-\frac{\pi}{2}|\lambda^\circ|^2} < 2.$$

2.3. Auxiliary results and proof strategy. In this section, we present some needed auxiliary results and present the general proof strategy.

Upper bounds for Laguerre polynomials were investigated in a number of papers, see, e.g., [24], [25], [27]. There is a classical estimate, known as Szegő bound, which is

$$(2.5) \quad |\mathcal{L}_n(x)| \leq e^{x/2} \quad \text{for } x \geq 0,$$

proved by Szegő [36, Sec. 7.21]. However, (2.5) will not be sufficient for our purposes as, using (2.4), it only gives that $|V_{h_n} h_n|$ is bounded by 1, which we also know from Lemma 2.3. It is worth mentioning that, in the other direction, Lemma 2.3 in combination with Lemma 2.4 implies the Szegő bound. We need more delicate estimates in the sequel.

First, observe that it is easy to obtain a crude polynomial bound for $\mathcal{L}_n$, as we have

$$(2.6) \quad |\mathcal{L}_n(x)| \leq \sum_{k=0}^n \binom{n}{k} \frac{x^k}{k!} \leq (n+1) \binom{n}{\lfloor n/2 \rfloor} x^n, \quad \forall x \geq 1.$$

Here, the notation $\lfloor m \rfloor$ stands for the greatest integer less than or equal to m . Note that the binomial coefficient may be huge, but this estimate becomes useful once x is, say, larger than x_n , the largest root of $\mathcal{L}_n$. Our estimate (2.6) will be useful in the calculations to prove Theorem 1.1 for n sufficiently small, which in our case is $n \leq 36$.

In the other direction, i.e., for large n , we use the following estimates, obtained from contemporary bounds on the values of Laguerre polynomials by Krasikov [24], [25].

Lemma 2.8. *For $n \geq 11$, the following estimates are simultaneously true.*

$$(2.7) \quad \mathcal{L}_n(-\pi(n+1)) \leq e^{(\pi+1)(n+1)}.$$

$$(2.8) \quad |\mathcal{L}_n(\pi(n+1))| \leq e^{\frac{\pi}{2}(n+1)} \sqrt{\frac{2}{n}}.$$

$$(2.9) \quad |\mathcal{L}_n(2\pi(n+1))| \leq e^{\pi(n+1)} \sqrt{\frac{2}{\pi}} e^{-0.1n-0.86}.$$

In addition, for generalized Laguerre polynomials we noticed that we can painlessly prove the following estimate:

$$(2.10) \quad |\mathcal{L}_n^\alpha(t)| \leq \sqrt{\frac{(n+\alpha)!}{n!}} e^{t/2} t^{-\alpha/2}, \quad n, \alpha \in \mathbb{N}_0, t > 0.$$

We don't actually need relation (2.10), but it may be of independent interest.

Proof. The proof of (2.7) is simple:

$$\mathcal{L}_n(-\pi(n+1)) = \sum_{k=0}^n \binom{n}{k} \frac{(\pi(n+1))^k}{k!} \leq \left(\binom{n}{\lfloor \frac{n}{2} \rfloor} \right) e^{\pi(n+1)} \leq (2e)^{n/2} e^{\pi(n+1)} \leq e^{(\pi+1)(n+1)}.$$

To prove (2.8), we invoke the following estimate from [24] (see also [25, Thm. 1]): for $n \geq 2$ and $x \in [q^2, s^2]$, where $q = \sqrt{n+1} - \sqrt{n}$ and $s = \sqrt{n+1} + \sqrt{n}$, the inequality

$$(2.11) \quad |\mathcal{L}_n(x)| \leq e^{x/2} \sqrt{\frac{s^2 - q^2}{(x - q^2)(s^2 - x)}}$$

holds true. Note that $x = \pi(n+1)$ belongs to $[q^2, s^2]$, since

$$q^2 = (\sqrt{n+1} - \sqrt{n})^2 \leq 2n+1 \leq \pi(n+1) \leq 4n \leq (\sqrt{n+1} + \sqrt{n})^2 = s^2.$$

Next, for $x = \pi(n+1)$ we estimate the value inside the square root as follows:

$$\begin{aligned} \frac{s^2 - q^2}{(x - q^2)(s^2 - x)} &= \frac{(\sqrt{n+1} + \sqrt{n})^2 - (\sqrt{n+1} - \sqrt{n})^2}{(\pi(n+1) - (\sqrt{n+1} - \sqrt{n})^2)(\sqrt{n+1} + \sqrt{n})^2 - \pi(n+1))} \\ &= \frac{4\sqrt{n(n+1)}}{(4\pi - \pi^2)n^2 + (6\pi - 2\pi^2)n - (1 - \pi)^2} \leq \frac{2}{n}, \end{aligned}$$

provided n is large enough. We estimate $\sqrt{n}$ by $\sqrt{n+1}$ in the numerator. Then, by cross-multiplying the ratios, it suffices to establish that

$$\tau(n) := (4\pi - \pi^2 - 2)n^2 + (6\pi - 2\pi^2 - 2)n - (1 - \pi)^2 > 0$$

holds true for $n \geq 11$. Note that $4\pi - \pi^2 - 2 > 0$, and the largest root n_+ of $\tau(n)$ satisfies

$$n_+ = \frac{(\pi^2 - 3\pi + 1) + \sqrt{(3\pi - \pi^2 - 1)^2 + (1 - \pi)^2(4\pi - \pi^2 - 2)}}{(4\pi - \pi^2 - 2)} \leq 6.$$

Therefore, using (2.11), we arrive at the desired estimate

$$|\mathcal{L}_n(\pi(n+1))| \leq e^{\frac{\pi}{2}(n+1)} \sqrt{\frac{2}{n}}.$$

It remains to prove (2.9). We recall that $q = \sqrt{n+1} - \sqrt{n} \leq s = \sqrt{n+1} + \sqrt{n}$. Clearly, $2\pi(n+1) \geq s^2$. Therefore, we can use Theorem 5 from [25] that asserts the estimate

$$(2.12) \quad |\mathcal{L}_n(2\pi(n+1))| \leq \frac{s}{\sqrt{2\pi(n+1)}} |\mathcal{L}_n(s^2)| e^{\pi(n+1) - \frac{s^2}{2}} e^{-R(s^2, 2\pi(n+1))/2},$$

where

$$R(s^2, 2\pi(n+1)) = \int_{s^2}^{2\pi(n+1)} \frac{\sqrt{(q^2 - t)(s^2 - t)}}{t} dt.$$

We start with estimating the function R from below. Note that

$$(2.13) \quad \begin{aligned} R(s^2, 2\pi(n+1)) &= \int_{s^2}^{2\pi(n+1)} \frac{\sqrt{(t - q^2)(t - s^2)}}{t} dt \\ &\geq \int_{s^2}^{2\pi(n+1)} \frac{(t - s^2)}{t} dt = 2\pi(n+1) - s^2 \left(1 + \ln \frac{2\pi(n+1)}{s^2} \right). \end{aligned}$$

Since Clearly, $4n+1 \leq s^2 \leq 4n+3 \leq 2\pi(n+1)$. Therefore, the logarithmic expression in (2.13) is positive and we get

$$(2.14) \quad R(s^2, 2\pi(n+1)) \geq 2\pi(n+1) - (4n+3) \left(1 + \ln \frac{2\pi(n+1)}{4n+1} \right).$$

Next, one can check that the logarithmic expression in (2.14) is monotonically decreasing in n for $n \geq 11$, and by evaluating at 11, we have $\ln(2\pi(n+1)/(4n+1)) \leq 0.52$. Hence,

$$(2.15) \quad R(s^2, 2\pi(n+1)) \geq n(2\pi - 4 \cdot 1.52) + 2\pi - 3 \cdot 1.52 \geq 0.2n + 1.72.$$

Now, we return to relation (2.12). Using the classical Szegő estimate (2.5), and estimate (2.15), we obtain our desired inequality:

$$|\mathcal{L}_n(2\pi(n+1))| \leq e^{\pi(n+1)} \frac{\sqrt{4(n+1)}}{\sqrt{2\pi(n+1)}} e^{-0.1n-0.86} \leq e^{\pi(n+1)} \frac{\sqrt{2}}{\sqrt{\pi}} e^{-0.1n-0.86}.$$

□

Remark 2.9. As can be seen in the proof, we only need the assumption $n \geq 11$ in our final estimate and the other inequalities already hold for smaller n . The estimates will be employed for $n \geq 36$ and so we did not try to find the optimal n .

We conclude the section with the following estimate.

Lemma 2.10. Let $a \geq 2$, $\gamma > \frac{2n+2}{a^2\pi}$, and $n \in \mathbb{N}$. Then

$$(2.16) \quad \sum_{\substack{(k,l) \in \mathbb{Z}^2, \\ k^2+l^2 \geq a^2}} (k^2 + l^2)^n e^{-\frac{\pi\gamma}{2}(k^2+l^2)} \leq \frac{4}{\gamma} a^{2n} e^{-\frac{a^2\pi\gamma}{2}}.$$

Proof. The series at the right-hand side of (2.16) is majorized by the integral

$$\begin{aligned} \int_0^{2\pi} \int_a^\infty r^{2n+1} e^{-\frac{\pi}{2}\gamma r^2} dr d\varphi &= 2\pi \int_a^\infty r^{2n+1} e^{-\frac{\pi}{2}\gamma r^2} dr \\ &= \pi \int_{a^2}^\infty \rho^n e^{-\frac{\pi}{2}\gamma \rho} d\rho = \pi \left(\frac{2}{\pi\gamma} \right)^{n+1} \Gamma \left(n+1, \frac{a^2\pi\gamma}{2} \right), \end{aligned}$$

where the notation $\Gamma(s, x)$ stands for the incomplete Gamma function defined by

$$\Gamma(s, x) = \int_x^\infty t^{s-1} e^{-t} dt.$$

To estimate $\Gamma(s, x)$ from above, we need the inequality

$$|\Gamma(s, x)| \leq Bx^{s-1}e^{-x} \quad \text{for } x > \frac{B}{(B-1)}(s-1), \quad s > 0, \quad B > 1.$$

established in [31]. Using this estimate with $B = 2$ (recall that $a^2\pi\gamma \geq 2n+2$), we get

$$\Gamma \left(n+1, \frac{a^2\pi\gamma}{2} \right) \leq 2 \left(\frac{a^2\pi\gamma}{2} \right)^n e^{-\frac{a^2\pi\gamma}{2}}$$

and conclude that

$$\sum_{\substack{(k,l) \in \mathbb{Z}^2, \\ |k|^2 + |l|^2 \geq a^2}} (k^2 + l^2)^n e^{-\frac{\pi\gamma}{2}(k^2 + l^2)} \leq 2\pi \left(\frac{2}{\pi\gamma} \right)^{n+1} \left(\frac{a^2\pi\gamma}{2} \right)^n e^{-\frac{a^2\pi\gamma}{2}} = \frac{4}{\gamma} a^{2n} e^{-\frac{a^2\pi\gamma}{2}}.$$

□

Proof strategy. In Section 3, we will use the Janssen test (2.3) to show that certain Gabor systems with Hermite functions over a square lattice form a frame. By Lemma 2.3 and (2.4) this leads to the verification that

$$\sum_{k,l \in \mathbb{Z}} |\mathcal{L}_n(\pi\delta(k^2 + l^2))| e^{-\frac{\pi}{2}\delta(k^2 + l^2)} < 2,$$

where δ is the density of the square lattice. The rough strategy is always to split the series for $|k|, |l|$ relatively small and the rest. After all, we have a Gaussian weight which guarantees that the contributions for $|k|, |l|$ large will be tiny. For small orders n , the bound of the terms with a small index will be achieved by a finite number of computations, i.e., explicitly computing values of a finite sum, while for large orders n we employ the bounds based on Lemma 2.8. The estimates on the remainder (large index) come from the auxiliary results in this section. The splitting will either involve squares, containing a finite number of points, or computing values for layers close to the origin, as shown in Fig. 2.

A layer of a lattice Λ is a set of points satisfying

$$\ell_r = \{\lambda \in \Lambda \mid |\lambda|^2 = r\}.$$

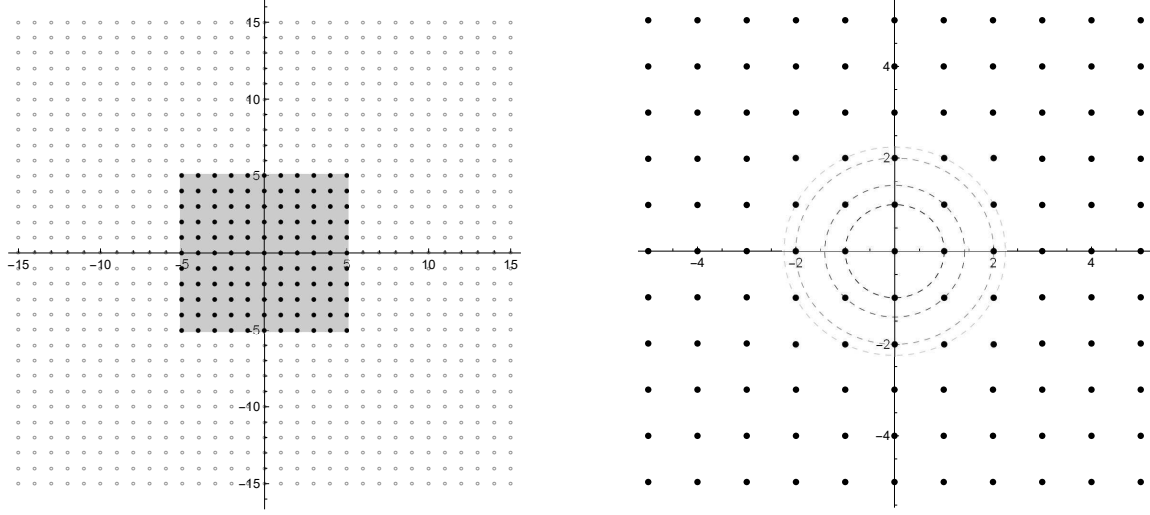


FIGURE 2. We split $\mathbb{Z}^2$ into a finite section, where we perform a finite number of computations, and a complementary region, where the series can be bounded by $0 < \varepsilon \ll 1$ sufficiently small (left). At some instances, instead of splitting $\mathbb{Z}^2$ by means of the max-norm, we use the Euclidean distance and $\mathbb{Z}^2$ splits into layers $\ell_m = \{(k, l) \in \mathbb{Z}^2 \mid k^2 + l^2 = m\}$ (right). This also shows the connection to decomposing an integer into a sum of two squares.

For most r the set will, of course, be empty. The set with the smallest index $r > 0$ which is not empty is called the first layer (it contains the closest neighbors of the origin), the non-empty set with the second smallest index is the second layer, and so on. The decomposition of the square lattice $\mathbb{Z}^2$ into layers naturally leads to the question of how to write an integer as the sum of two squares. Denote by $r_2(n)$ the number of ways of writing $n \in \mathbb{N}_0$ as the sum of two squares, allowing zeros and distinguishing signs and order. For example,

$$\begin{aligned} r_2(0) &= 1 && \text{as } 0 = 0^2 + 0^2, \\ r_2(1) &= 4 && \text{as } 1 = 0^2 + 1^2 = 0^2 + (-1)^2 = 1^2 + 0^2 = (-1)^2 + 0^2, \\ r_2(2) &= 4 && \text{as } 2 = 1^2 + 1^2 = 1^2 + (-1)^2 = 1^2 + (-1)^2 = (-1)^2 + (-1)^2, \\ r_2(3) &= 0 && \text{as 3 factors containing primes } p \equiv 3 \pmod{4} \text{ of odd power.} \end{aligned}$$

Note that $|\ell_m| = r_2(m)$ and that we have the trivial estimate

$$(2.17) \quad r_2(m) \leq 4m, \quad m \in \mathbb{N},$$

as we have at most $\sqrt{m}$ possible choices for each summand and 4 choices to place signs. This bound is quite rough, but sufficient for our purposes.

The last ingredient we need are truncated geometric series and their derivatives. This will include, e.g., for $|q| < 1$,

$$\begin{aligned} f_N(q) &:= \sum_{m \geq N} q^m = q^N \sum_{m \geq 0} q^m = \frac{q^N}{1-q}, \\ q f'_N(q) &= \sum_{m \geq N} m q^m = \frac{q^N}{(1-q)^2} (N - (N-1)q), \\ q^2 f''_N(q) &= \sum_{m \geq N} (m^2 q^m - m q^m) = \sum_{m \geq N} m^2 q^m - q^2 f'_N(q) \\ &= \frac{q^N}{(1-q)^3} ((N^2 - N) - (2N^2 + 4N)q + (N^2 - 3N + 2)q^2), \end{aligned}$$

and so on. As these results are obtained by elementary means, but are rather laborious for high orders, we will not always provide all steps and only give the final result instead.

3. RESULTS

3.1. Proof of Theorem 1.1. We will now prove Theorem 1.1 by settling the cases for $4 \leq n \leq 36$ first, which we solve by a finite number of computations and showing that the cut-off remainders cannot contribute enough mass such that the finite sums would violate the Janssen test in Lemma 2.6. The case $n \geq 36$ will be proved separately by providing uniform bounds (in n) on the terms in the first and second layer of the adjoint lattice and bounding the remainder.

Proposition 3.1. *Let $n \in \mathbb{N}$ with $4 \leq n \leq 36$. Then, the Gabor system*

$$\mathcal{G}_n = \mathcal{G} \left(h_n, \frac{1}{\sqrt{n+1}} \mathbb{Z}^2 \right)$$

constitutes a frame for $L^2(\mathbb{R})$.

Proof. We use the Janssen test to show that $\mathcal{G}_n$ is a frame for $n \in \mathbb{N}$ with $4 \leq n \leq 36$. So, we have to show that

$$\sum_{(k,l) \in \mathbb{Z}^2} |\mathcal{L}_n(\pi(n+1)(k^2 + l^2))| e^{-\frac{\pi}{2}(n+1)(k^2 + l^2)} < 2.$$

As h_n is in $M^1(\mathbb{R})$, $n \in \mathbb{N}_0$, (Condition A) holds and the series converges unconditionally. So, we may split the series in two parts: a finite part consisting of lattice points with $|k|, |l| \leq 5$ and a remainder with $\max\{|k|, |l|\} \geq 6$.

By (2.6), (2.17), we bound the remainder in a generous way, as we expect it to be small:

$$\begin{aligned}
& \sum_{\max\{|k|, |l|\} \geq 6} |\mathcal{L}_n(\pi(n+1)(k^2 + l^2))| e^{-\frac{\pi}{2}(n+1)(k^2 + l^2)} \\
& \leq (n+1) \binom{n}{\lfloor n/2 \rfloor} \sum_{\max\{|k|, |l|\} \geq 6} (\pi(n+1)(k^2 + l^2))^n e^{-\frac{\pi}{2}(n+1)(k^2 + l^2)} \\
& = \pi^n (n+1)^{n+1} \binom{n}{\lfloor n/2 \rfloor} \sum_{m \geq 36} r_2(m) m^n e^{-\frac{\pi}{2}(n+1)m} \\
& \leq 4\pi^n (n+1)^{n+1} \binom{n}{\lfloor n/2 \rfloor} \sum_{m \geq 36} m^{n+1} e^{-\frac{\pi}{2}(n+1)m}.
\end{aligned}$$

We first estimate the constant in front of the series, which is clearly increasing in n :

$$4\pi^n (n+1)^{n+1} \binom{n}{\lfloor n/2 \rfloor} \leq 4\pi^{36} \cdot 37^{37} \cdot \binom{36}{18} < 4 \cdot 10^{86} < 10^{87}, \quad n \leq 36.$$

To estimate the series we need the following auxiliary result.

Claim. For $x, y > 0$, the function $f(y) := x^y e^{-\frac{\pi}{2}yx}$ is decreasing in y , for any fixed x . Indeed, the claim follows easily from

$$f'(y) = x^y e^{-\frac{\pi}{2}yx} \left(\log(x) - \frac{\pi}{2}x \right) < 0, \quad x > 0, y > 0.$$

The series can now be estimated by using the claim and plugging in $n = 4$:

$$\begin{aligned}
\sum_{m \geq 36} m^{n+1} e^{-\frac{\pi}{2}(n+1)m} & \leq \sum_{m \geq 36} m^5 q^m = \frac{q^{36}}{(1-q)^6} \cdot \left(60\,466\,176 - 293\,453\,099\,q \right. \\
& \quad \left. + 570\,164\,066\,q^2 - 554\,350\,974\,q^3 \right. \\
& \quad \left. + 269\,695\,826\,q^4 - 52\,521\,875\,q^5 \right).
\end{aligned}$$

Here, we set $q = e^{-5\pi/2}$ and used *Mathematica* [40] to symbolically compute the infinite series. We remark that this can also be achieved by hand by using the truncated geometric series and finite linear combinations of its derivatives as indicated in Sec. 2.3. Plugging in numerical values for $q = e^{-5\pi/2}$ finally leads to the estimate

$$\sum_{m \geq 36} m^{n+1} e^{-\frac{\pi}{2}(n+1)m} \leq 10^{-115}, \quad n \geq 4.$$

Thus, for $4 \leq n \leq 36$, the remainder can be estimated by

$$\sum_{\max\{|k|, |l|\} \geq 6} |\mathcal{L}_n(\pi(n+1)(k^2 + l^2))| e^{-\frac{\pi}{2}(n+1)(k^2 + l^2)} < 10^{-115+87} = 10^{-28}.$$

Now, we are left with a finite number of computations. Using *Mathematica*, we compute the values of the finite part to 5 significant digits after the decimal point (always rounded

up) to see that indeed we have

$$(3.1) \quad \sum_{|k|, |l| \leq 5} |\mathcal{L}_n(\pi(n+1)(k^2 + l^2))| e^{-\frac{\pi}{2}(n+1)(k^2 + l^2)} < 1.99999 = 2 - 10^{-5}, \quad 4 \leq n \leq 36.$$

Again, these calculations could be carried out to the necessary precision by hand. The precise values of (3.1) for $n \in \{0, \dots, 36\}$ are given in Fig. 3. For $4 \leq n \leq 36$ they are all significantly (at least 10^{-3}) smaller than 2. With our estimates on the remainder, this gives a complete proof of the frame property for the Gabor systems $\mathcal{G}_n$ and $4 \leq n \leq 36$. $\square$

n	0	1	2	3	4	5	6	7
value	2.01497	2	2.00003	2	1.99390	1.97889	1.95381	1.91844
n	8	9	10	11	12	13	14	15
value	1.87308	1.81835	1.75515	1.68451	1.60760	1.52567	1.44006	1.35211
n	16	17	18	19	20	21	22	23
value	1.26320	1.17470	1.08793	1.00419	1.07535	1.14951	1.21732	1.27788
n	24	25	26	27	28	29	30	31
value	1.33044	1.37440	1.40932	1.43490	1.45101	1.45770	1.45517	1.44376
n	32	33	34	35	36			
value	1.42396	1.39642	1.36187	1.32118	1.27528			

FIGURE 3. Values of the finite computations part of the Janssen test for the Gabor systems $\mathcal{G}(h_n, (1/\sqrt{n+1})\mathbb{Z}^2)$. The values for $n = 1$ and $n = 3$ are exact and the other values are given to 5 digits after the decimal point (rounded up).

As the values in Fig. 3 did not show a monotonic behavior, we carried out further numerical investigations for n up to 120. The behavior is curious and visualized in Fig. 4.

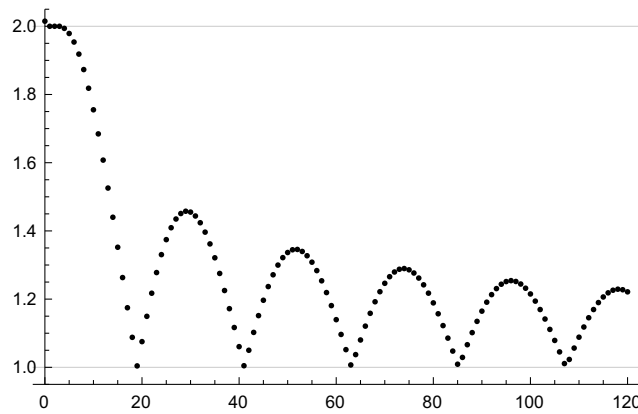


FIGURE 4. Values of the Janssen test for $\mathcal{G}(h_n, 1/\sqrt{(n+1)}\mathbb{Z}^2)$ and $n \leq 120$.

We will now treat the case $n \geq 36$. The proof, again, naturally splits into bounding the mass of the resulting series on a finite set, consisting of the origin and the first two layers of the adjoint lattice, and a small remainder. Our bounds are not the best possible and there is room to improve on n , i.e., make it smaller, but in any case we would be left with finitely many cases which need to be checked by hand (or computer), unless we take more layers into account for the finite set.

Proposition 3.2. *Let $n \geq 36$, then the Gabor system $\mathcal{G}(h_n, \frac{1}{\sqrt{n+1}}\mathbb{Z} \times \frac{1}{\sqrt{n+1}}\mathbb{Z})$ is a frame.*

Proof. We will use the Janssen test again and show that

$$\sum_{\lambda^\circ \in \Lambda^\circ} |V_{h_n} h_n(\lambda^\circ)| < 2 \quad \text{for } n \geq 36,$$

which by (2.4), is equivalent to showing

$$(3.2) \quad \sum_{k,l \in \mathbb{Z}} |\mathcal{L}_n(\pi(n+1)(k^2 + l^2))| e^{-\frac{\pi(n+1)}{2}(k^2 + l^2)} < 2, \quad n \geq 36.$$

Denote by $K = \{(0, 0), (0, \pm 1), (\pm 1, 0), (\pm 1, \pm 1)\}$ the set which contains the origin, its 4 closest neighbors ℓ_1 and its 4 second closest neighbors ℓ_2 . We split the sum at the left-hand side of (3.2) into the sum over K and $\mathbb{Z}^2 \setminus K$. Since $\mathcal{L}_n(0) = 1$, it suffices to prove

$$\ell_1(n) + \ell_2(n) + \ell_{3+}(n) < 1,$$

where $\ell_1(n)$ is the value of the sum over points in the first layer, $\ell_2(n)$ the value for the second layer and $\ell_{3+}(n)$ the value of the sum over all remaining layers, i.e.,

$$\begin{aligned} \ell_1(n) &= 4\mathcal{L}_n(\pi(n+1)) e^{-\frac{\pi}{2}(n+1)}, & \ell_2(n) &= 4\mathcal{L}_n(2\pi(n+1)) e^{-\pi(n+1)}, \\ \ell_{3+}(n) &= \sum_{(k,l) \in \mathbb{Z}^2 \setminus K} |\mathcal{L}_n(\pi(n+1)(k^2 + l^2))| e^{-\frac{\pi(n+1)}{2}(k^2 + l^2)}. \end{aligned}$$

By (2.8) and (2.9), we have

$$(3.3) \quad |\ell_1(n)| < 4\sqrt{\frac{2}{n}} \quad \text{and} \quad |\ell_2(n)| < 4\sqrt{\frac{2}{\pi}} e^{-0.1n-0.86}.$$

Since the coefficients of Laguerre polynomials have alternating signs, we have

$$\ell_{3+}(n) \leq \sum_{(k,l) \in \mathbb{Z}^2 \setminus K} (k^2 + l^2)^n |\mathcal{L}_n(-\pi(n+1))| e^{-\frac{\pi(n+1)}{2}(k^2 + l^2)}.$$

Using (2.7) and the fact that $n+1 \leq \frac{\pi(n+1)}{2}$, we arrive at

$$\begin{aligned} \ell_{3+}(n) &\leq \sum_{(k,l) \in \mathbb{Z}^2 \setminus K} (k^2 + l^2)^n e^{(\pi+1)(n+1) - \frac{\pi(n+1)}{2}(k^2 + l^2)} \\ (3.4) \quad &< \sum_{(k,l) \in \mathbb{Z}^2 \setminus K} (k^2 + l^2)^n e^{-\frac{\pi(n+1)}{2}(k^2 + l^2 - 3)} = e^{\frac{3\pi}{2}(n+1)} \sum_{(k,l) \in \mathbb{Z}^2 \setminus K} (k^2 + l^2)^n e^{-\frac{\pi(n+1)}{2}(k^2 + l^2)}. \end{aligned}$$

To estimate the series at the right-hand side of (3.4), we use Lemma 2.10 with $\gamma = n + 1$ and $a = 2$:

$$\sum_{(k,l) \in \mathbb{Z}^2 \setminus K} (k^2 + l^2)^n e^{-\frac{\pi(n+1)}{2}(k^2+l^2)} \leq \frac{4^{2(n+1)}}{n+1} e^{-2\pi(n+1)}.$$

Therefore,

$$(3.5) \quad l_{3+}(n) \leq e^{-0.184n}.$$

Combining (3.3) and (3.5), we finally arrive at

$$\ell_1(n) + \ell_2(n) + \ell_{3+}(n) < 4\sqrt{\frac{2}{n}} + 4\sqrt{\frac{2}{\pi}} e^{-0.1n-0.86} + e^{-0.184n} < 1$$

as long as n is sufficiently large. Clearly, the expression above is decreasing in n and the inequality is satisfied for $n = 36$. This finishes the proof. $\square$

3.2. Failure of the Janssen test.

3.2.1. The case of small n . For the Gabor systems $\mathcal{G}(h_n, \delta^{-1/2}\mathbb{Z}^2)$ and $n \in \{0, 1, 2, 3\}$ we know that we do not have a frame for $\delta = n + 1$. In particular, the lower frame bound vanishes and the bound in the Janssen test must yield a value of at least 2. For $\delta > n + 1$ we are in the safety region and have a Gabor frame. For small n , in particular up to 3, the Janssen test does not shed any new light on the frame set of

$$\mathcal{G}_n(\delta) = \mathcal{G}(h_n, \delta^{-1/2}\mathbb{Z}^2).$$

In accordance with the previous sections, we set $\mathcal{G}_n = \mathcal{G}_n(n + 1)$.

For $\mathcal{G}_0$, i.e., the Gaussian function h_0 and the integer lattice $\mathbb{Z}^2$, the Janssen test fails:

$$\sum_{k,l \in \mathbb{Z}} |V_{h_0} h_0(k, l)| = \sum_{k,l \in \mathbb{Z}} e^{-\frac{\pi}{2}(k^2+l^2)} > \sum_{k,l \in \{-1,0,1\}} e^{-\frac{\pi}{2}(k^2+l^2)} > 2.004.$$

So, the Janssen test is inconclusive, but it is a well-known fact that the lower frame bound vanishes in this case [28], [33], [34] and we do not have a frame. Similarly, for $\mathcal{G}_2$ we have

$$\sum_{k,l \in \mathbb{Z}} |V_{h_2} h_2(\sqrt{3}k, \sqrt{3}l)| > \sum_{k,l \in \{-1,0,1\}} |\mathcal{L}_3(3\pi(k^2 + l^2))| e^{-\frac{3\pi}{2}(k^2+l^2)} > 2.00001.$$

The lower frame bound vanishes, as shown by J. Lemvig [26, Thm. 7], and we do not have a frame with the Gabor system $\mathcal{G}_2$.

Next, we consider the Gabor systems $\mathcal{G}_n$ for $n \in \{1, 3\}$. The results of Lyubarskii and Nes show that the lower bound vanishes if $n = 1$. For $n = 3$ we have again a result by Lemvig [26, Thm. 8] which shows that the lower frame bound is zero. We remark that Boon, Zak, and Zucker [3] already studied the series considered in [26] in the context of coherent states and provided zeros (see also [2]).

It is quite remarkable that for n being 1 or 3 the Janssen test yields exactly 2. In this regard, these Gabor systems are closest possible to being a frame, but they still just fail.

Proposition 3.3. *Let $n \in \{1, 3\}$. Then, the following hold:*

$$\sum_{k,l \in \mathbb{Z}} |V_{h_n} h_n(\sqrt{n+1} k, \sqrt{n+1} l)| = \sum_{k,l \in \mathbb{Z}} |\mathcal{L}_n((n+1)\pi(k^2 + l^2))| e^{-\frac{(n+1)\pi}{2}(k^2 + l^2)} = 2$$

and

$$\sum_{k,l \in \mathbb{Z}} V_{h_n} h_n(\sqrt{n+1} k, \sqrt{n+1} l) = \sum_{k,l \in \mathbb{Z}} \mathcal{L}_n((n+1)\pi(k^2 + l^2)) e^{-\frac{(n+1)\pi}{2}(k^2 + l^2)} = 0.$$

Proof. Note that for any $n \in \mathbb{N}$, the term $(k, l) = (0, 0)$ contributes 1. Hence, our claim is

$$\sum_{(k,l) \in \mathbb{Z}^2 \setminus \{0,0\}} |\mathcal{L}_n((n+1)\pi(k^2 + l^2))| e^{-\frac{(n+1)\pi}{2}(k^2 + l^2)} = 1, \quad \text{for } n = 1, 3.$$

By the results of Lyubarskii and Nes [29] and Lemvig [26] we know that the Gabor system $\mathcal{G}(h_n, (n+1)^{-1/2}\mathbb{Z}^2)$ is not a frame as the lower frame bound vanishes.

Now, on the one hand, we know that this implies (the Janssen test must fail) that

$$(3.6) \quad \sum_{(k,l) \in \mathbb{Z}^2 \setminus \{(0,0)\}} |\mathcal{L}_n((n+1)\pi(k^2 + l^2))| e^{-\frac{(n+1)\pi}{2}(k^2 + l^2)} \geq 1.$$

On the other hand, we know from the results by Janssen [22] (see also [5]) that the minimum of the Fourier series

$$F_{g,\Lambda}(z) = \text{vol}(\Lambda)^{-1} \sum_{\lambda^\circ \in \Lambda^\circ} V_g g(\lambda^\circ) e^{2\pi i \sigma(\lambda^\circ, z)}, \quad z = (x, \omega) \in \mathbb{R}^2,$$

gives the sharp lower frame bound of the Gabor system $\mathcal{G}(g, \Lambda)$ for $\text{vol}(\Lambda)^{-1} \in \mathbb{N}$. Thus, we have $F_{g,\Lambda}(x, \omega) \geq 0$. Setting $\Lambda_n = 1/\sqrt{n+1} \mathbb{Z}^2$ for the moment, this implies

$$\begin{aligned} -1 \leq \text{vol}(\Lambda) F_{h_n, \Lambda_n}(0, 0) - 1 &= \sum_{(k,l) \in \mathbb{Z}^2 \setminus \{0,0\}} V_{h_n} h_n(\sqrt{n+1} k, \sqrt{n+1} l) \\ &= \sum_{(k,l) \in \mathbb{Z}^2 \setminus \{0,0\}} \mathcal{L}_n((n+1)\pi(k^2 + l^2)) e^{-\frac{(n+1)\pi}{2}(k^2 + l^2)}. \end{aligned}$$

Note that the phase factor, imposed by (2.4), is 1 as the lattice density is $n+1$, which is even in this case. Next, we look at the Laguerre polynomials $\mathcal{L}_1(x)$ and $\mathcal{L}_3(x)$. They are

$$\mathcal{L}_1(x) = 1 - x \quad \text{and} \quad \mathcal{L}_3(x) = \frac{1}{6} (6 - 18x + 9x^2 - x^3).$$

The distance of the closest lattice point to the origin is $(n+1)\pi$, so we see that $\mathcal{L}_1(x)$ and $\mathcal{L}_3(x)$ are negative when evaluated at lattice points (except for the origin). Thus, for all $(k, l) \in \mathbb{Z}^2 \setminus \{(0, 0)\}$ we have

$$|\mathcal{L}_n((n+1)\pi(k^2 + l^2))| e^{-\frac{(n+1)\pi}{2}(k^2 + l^2)} = -\mathcal{L}_n((n+1)\pi(k^2 + l^2)) e^{-\frac{(n+1)\pi}{2}(k^2 + l^2)}.$$

Hence, relation (3.6) can equivalently be written as

$$\sum_{(k,l) \in \mathbb{Z}^2 \setminus \{0,0\}} \mathcal{L}_n((n+1)\pi(k^2 + l^2)) e^{-\frac{(n+1)\pi}{2}(k^2 + l^2)} \leq -1$$

which ultimately gives the equality

$$\sum_{(k,l) \in \mathbb{Z}^2 \setminus \{0,0\}} \mathcal{L}_n((n+1)\pi(k^2+l^2)) e^{-\frac{(n+1)\pi}{2}(k^2+l^2)} = -1.$$

The claims follow by adding the value at the origin, which contributes 1, to the sum. $\square$

For the proof of this result, we made use of the fact that considered systems do not form frames. It would be nice to find a proof which does not need this fact.

As h_1 remains the only case where the frame set conjecture is still open (no further restrictions than the ones from [29] are known), we considered it worthwhile to apply the Janssen test in this case. The result is that we do not get any new insights.

Lemma 3.4. *For the Gabor system $\mathcal{G}(h_1, \delta^{-1/2}\mathbb{Z}^2)$, $\delta \geq 1$, the Janssen test decreases, i.e.,*

$$\delta \mapsto \sum_{k,l} |\mathcal{L}_1(\pi\delta(k^2+l^2))| e^{-\frac{\pi}{2}\delta(k^2+l^2)}$$

is a decreasing function in δ .

Proof. We introduce the function j_1 , which we define, for $\delta \geq 1$, to be

$$j_1(\delta) = \sum_{k,l \in \mathbb{Z}} |\mathcal{L}_1(\pi\delta(k^2+l^2))| e^{-\frac{\pi}{2}\delta(k^2+l^2)} = \sum_{k,l \in \mathbb{Z}} |1 - \pi\delta(k^2+l^2)| e^{-\frac{\pi}{2}\delta(k^2+l^2)}.$$

All terms have a negative contribution, except for the constant term, which is 1, provided by the origin. So, we have

$$j_1(\delta) = 1 + \sum_{k,l \in \mathbb{Z} \setminus \{0\}} (\pi\delta(k^2+l^2) - 1) e^{-\frac{\pi}{2}\delta(k^2+l^2)},$$

which is clearly smooth on $(1, \infty)$ and continuous at 1. Taking the derivative, we obtain

$$\begin{aligned} j_1'(\delta) &= \sum_{k,l \in \mathbb{Z} \setminus \{0\}} \left(\pi(k^2+l^2) e^{-\frac{\pi}{2}\delta(k^2+l^2)} - \frac{\pi(k^2+l^2)}{2} (\pi\delta(k^2+l^2) - 1) e^{-\frac{\pi}{2}\delta(k^2+l^2)} \right) \\ &= \frac{\pi}{2} \sum_{k,l \in \mathbb{Z} \setminus \{0\}} (k^2+l^2) (3 - \pi\delta(k^2+l^2)) e^{-\frac{\pi}{2}\delta(k^2+l^2)} < 0. \end{aligned}$$

$\square$

Proposition 3.3 together with Lemma 3.4 in particular shows that the Janssen test cannot be used to enlarge the frame set of h_1 for Gabor systems of the form $\mathcal{G}(h_1, \delta^{-1/2}\mathbb{Z}^2)$ for $1 < \delta < 2$. We remark that it also seems to be true that the Janssen test is inconclusive for rectangular Gabor systems of the form $\mathcal{G}(h_1, \delta^{-1/2}(a\mathbb{Z} \times (1/a)\mathbb{Z}))$, $a \in \mathbb{R}_+$ and $1 < \delta < 2$.

3.3. Outside the safety region. In this section, we show that the frame set of the Hermite function h_n on the main diagonal $a = b$ may have a rather surprising structure. In particular, $\mathfrak{F}(h_n)$ may contain a large segment on the main diagonal outside of the safety region. Our results do not rule out the option that all pairs (a, a) , $a < 1$, belong to $\mathfrak{F}(h_n)$, $n \geq 4$, except for the points ruled out by Lemvig [26, Thm. 7, Thm. 8] and Lyubarskii and Nes [29], but we do not dare to come up with a new conjecture.

3.3.1. *The 15-th Hermite function.* For illustrative reasons, we decided to pick the particular case $n = 15$. We prove the following result.

Proposition 3.5. *The Gabor system $\mathcal{G}(h_{15}, 1/\sqrt{\delta} \mathbb{Z}^2)$ is a frame for $L^2(\mathbb{R})$ for $\delta \geq 11$.*

Proof. We follow our approach in the proof of Theorem 1.1 and use the Janssen test with sharper estimates on the Laguerre polynomials for the particular case $n = 15$. By Lemma 2.6 and (2.4), it suffices to show that

$$\sum_{k,l \in \mathbb{Z}} |\mathcal{L}_{15}(\pi\delta(k^2 + l^2))| e^{-\frac{\pi\delta}{2}(k^2 + l^2)} < 2 \quad \text{for } 11 \leq \delta \leq 16.$$

We split the summation and, for $\varepsilon \in (0, 1)$, reduce the problem to the verification of

$$4|\mathcal{L}_{15}(\pi\delta)|e^{-\frac{\pi\delta}{2}} + 4|\mathcal{L}_{15}(2\pi\delta)|e^{-\pi\delta} + 4|\mathcal{L}_{15}(4\pi\delta)|e^{-2\pi\delta} + 4|\mathcal{L}_{15}(5\pi\delta)|e^{-\frac{5\pi}{2}\delta} < 1 - \varepsilon$$

$$\sum_{k^2 + l^2 \geq 8} |\mathcal{L}_{15}(\pi\delta(k^2 + l^2))| e^{-\frac{\pi\delta}{2}(k^2 + l^2)} < \varepsilon.$$

First, we deal with the tail estimate. By (2.6), we have

$$\sum_{k^2 + l^2 \geq 8} |\mathcal{L}_{15}(\pi\delta(k^2 + l^2))| e^{-\frac{\pi\delta}{2}(k^2 + l^2)} \leq 16 \binom{15}{7} \pi^{15} \delta^{15} \sum_{k^2 + l^2 \geq 8} (k^2 + l^2)^{15} e^{-\frac{\pi\delta}{2}(k^2 + l^2)}.$$

Next, we apply Lemma 2.10 with $\gamma = \delta$, $a = \sqrt{8}$, and get

$$\sum_{k^2 + l^2 \geq 8} (k^2 + l^2)^{15} e^{-\frac{\pi\delta}{2}(k^2 + l^2)} \leq 4 \frac{8^{15}}{\delta} e^{-4\pi\delta}.$$

Since

$$2^{4+2+45} \binom{15}{7} \pi^{15} \leq 10^{27} \quad \text{and} \quad \delta^{14} e^{-4\pi\delta} < 10^{-39} \quad \text{for } 11 \leq \delta \leq 16,$$

we deduce that

$$(3.7) \quad \sum_{k^2 + l^2 \geq 8} |\mathcal{L}_{15}(\pi\delta(k^2 + l^2))| e^{-\frac{\pi\delta}{2}(k^2 + l^2)} \leq 10^{-12}.$$

For the remaining layers we use the following estimates for $\delta \in [11, 16]$:

$$|\mathcal{L}_{15}(\pi\delta)e^{-\frac{\pi\delta}{2}}| \leq 0.15, \quad |\mathcal{L}_{15}(2\pi\delta)e^{-\pi\delta}| \leq 0.04,$$

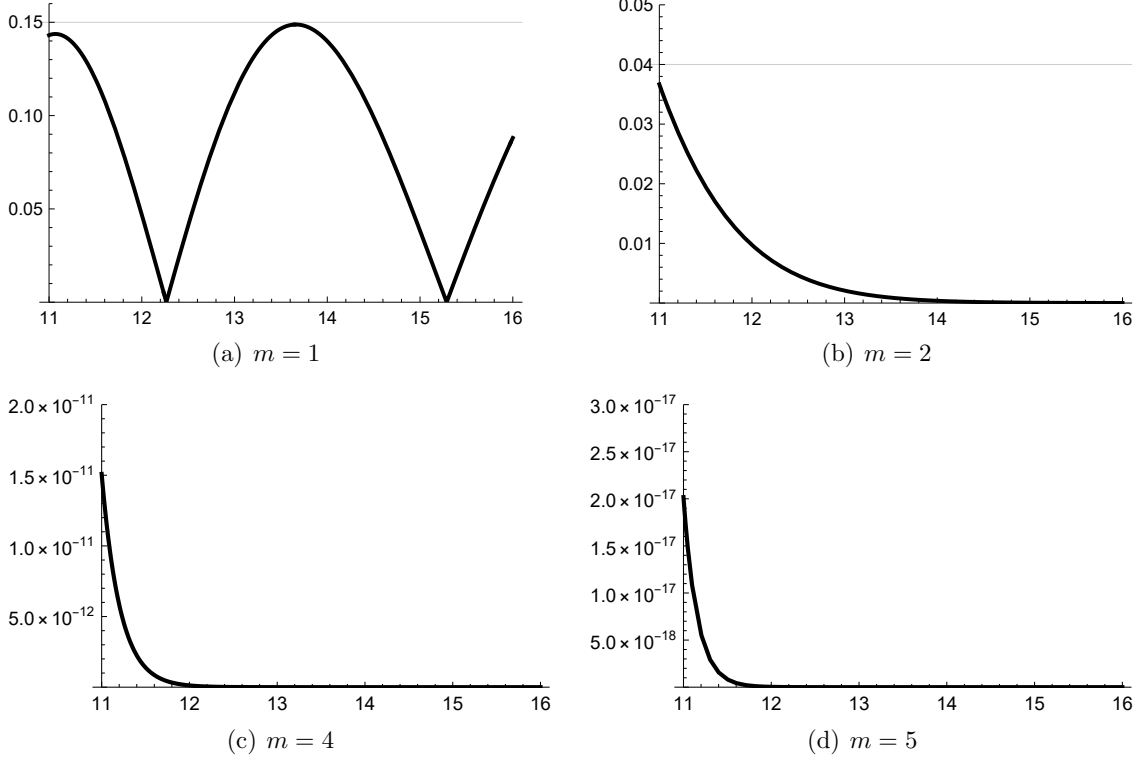
$$|\mathcal{L}_{15}(4\pi\delta)e^{-2\pi\delta}| \leq 10^{-10}, \quad |\mathcal{L}_{15}(5\pi\delta)e^{-\frac{5\pi}{2}\delta}| \leq 10^{-16}.$$

We support these bounds with Fig. 5. However, they can be verified by hand, as $\mathcal{L}_{15}$ is a fixed polynomial and evaluated on a compact interval, and its roots (given, e.g., in [32]) and critical points can be calculated with any required precision.

Combining these estimates with (3.7), we arrive at

$$\sum_{k^2 + l^2 \geq 8} |\mathcal{L}_{15}(\pi\delta(k^2 + l^2))| e^{-\frac{\pi\delta}{2}(k^2 + l^2)} \leq 1 + 4 \cdot 0.15 + 4 \cdot 0.04 + 4 \cdot 10^{-10} + 4 \cdot 10^{-16} < 1.8.$$

□

FIGURE 5. The functions $|\mathcal{L}_{15}(m\pi\delta)|e^{-m\pi\delta/2}$ for $m = 1, 2, 4, 5$.

Remark 3.6. Numerical investigations (see Fig. 6) suggest that the Gabor system constitutes a frame in $L^2(\mathbb{R})$ in a neighborhood of $\delta = 3.06$, which is very far from the safety region. This can be checked with the methods used in Proposition 3.1. We have

$$\sum_{|k|, |l| \leq 5} |\mathcal{L}_{15}(3.06 \pi(k^2 + l^2))| e^{-\frac{3.06}{2} \pi(k^2 + l^2)} \leq 1.96933$$

and the remainder term is again small:

$$\begin{aligned} \sum_{\max |k|, |l| \geq 6} |\mathcal{L}_{15}(3.06 \pi(k^2 + l^2))| e^{-\frac{3.06}{2} \pi(k^2 + l^2)} &\leq 16 \binom{16}{8} (3.06\pi)^{15} \sum_{m \geq 36} m^{16} \left(e^{-\frac{3.06\pi}{2}} \right)^m \\ &< 10^{21-50} = 10^{-29}. \end{aligned}$$

Hence, the Gabor system $\mathcal{G}(h_{15}, 1/\sqrt{3.06} \mathbb{Z}^2)$ is a frame and there is an open neighborhood U around $(1/\sqrt{3.06}, 1/\sqrt{3.06})$ such that $\mathcal{G}(h_{15}, a_U \mathbb{Z} \times b_U \mathbb{Z})$ is a frame for $(a_U, b_U) \in U$.

Remark 3.7. As can be seen from Fig. 6, the safety region for h_{15} and $\delta^{-1/2} \mathbb{Z}^2$ can be extended to, say, $\delta = 9.6$. As the frame set of a function in $M^1(\mathbb{R})$ is open [17], this also means that there is a neighborhood around the pairs $(1/\sqrt{\delta}, 1/\sqrt{\delta})$ for $\delta \geq 9.6$ which

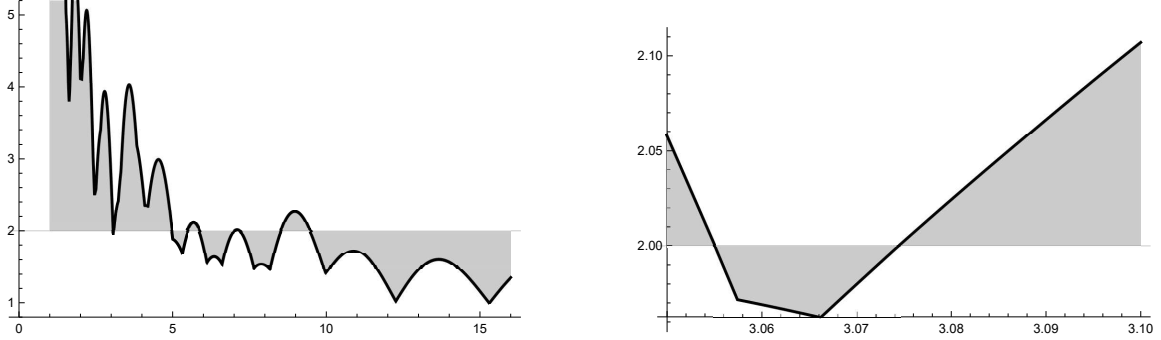


FIGURE 6. The Janssen test for the Gabor system $\mathcal{G}(h_{15}, \delta^{-1/2}\mathbb{Z}^2)$. For the square lattice of density $\delta > 16$, we are in the safety region. However, the Janssen test provides a sufficient criterion way earlier, e.g., from $\delta \geq 10$. It is inconclusive around $\delta = 7.1$, but suffices around $\delta = 3.06$. The gray area is for visual aid where the value of the test is above or below 2.

belongs to $\mathfrak{F}(h_{15})$. The above estimates, however, need refinements for $\delta \in [9.6, 11]$ as $\mathcal{L}_{15}(\delta\pi)$ has a zero close to $\delta = 10$ ($\mathcal{L}_{15}(x)$ has a root at $x = 31.407519169754$ [32] which gives $\delta \approx 9.99732$). The mass of the first layer is small, while the second layer contributes more mass. By dividing the interval $[9.6, 11]$ into a finite number of sub-intervals and performing estimates similar to the ones above, we get the desired extension. Moreover, with a finite number of computations, it seems possible to enlarge the safety region for h_{15} for rectangular lattices $a\mathbb{Z} \times b\mathbb{Z}$ to, say, $ab < 1/11$, for $a, b \in [1, 1.3]$, as shown in Fig. 7.

3.3.2. The 9-th Hermite function. In [20], the existence of pairs (a, b) with $ab = 1/3$ which do not belong to the frame set of h_n , $n \geq 3$ was shown. We consider the example of h_9 and the square lattice of density 3, i.e., $1/\sqrt{3}\mathbb{Z}^2$ and show that $\mathcal{G}(h_9, 1/\sqrt{3}\mathbb{Z}^2)$ is a frame.

Proposition 3.8. *The Gabor system $\mathcal{G}(h_9, 1/\sqrt{3}\mathbb{Z}^2)$ is a frame.*

Proof. We start with a finite number of computations, which gives

$$\sum_{|k|, |l| \leq 5} \mathcal{L}_9(3\pi(k^2 + l^2)) e^{-\frac{3\pi}{2}(k^2 + l^2)} \leq 1.76496.$$

The bound on the remainder is derived analogously to Proposition 3.1. We estimate

$$\begin{aligned} \sum_{\max |k|, |l| \geq 6} \mathcal{L}_9(3\pi(k^2 + l^2)) e^{-\frac{3\pi}{2}(k^2 + l^2)} &\leq \sum_{\max |k|, |l| \geq 6} 10 \binom{10}{5} (3\pi(k^2 + l^2))^9 e^{-\frac{3\pi}{2}(k^2 + l^2)} \\ &< 10^{13-58} = 10^{-45}. \end{aligned}$$

□

For h_0 and h_1 we know that density 3 is sufficient as we are in the safety region [18]. For h_2 we know that $1/\sqrt{3}\mathbb{Z}^2$ does not provide a frame [26]. Curiously, the only cases $n \geq 3$

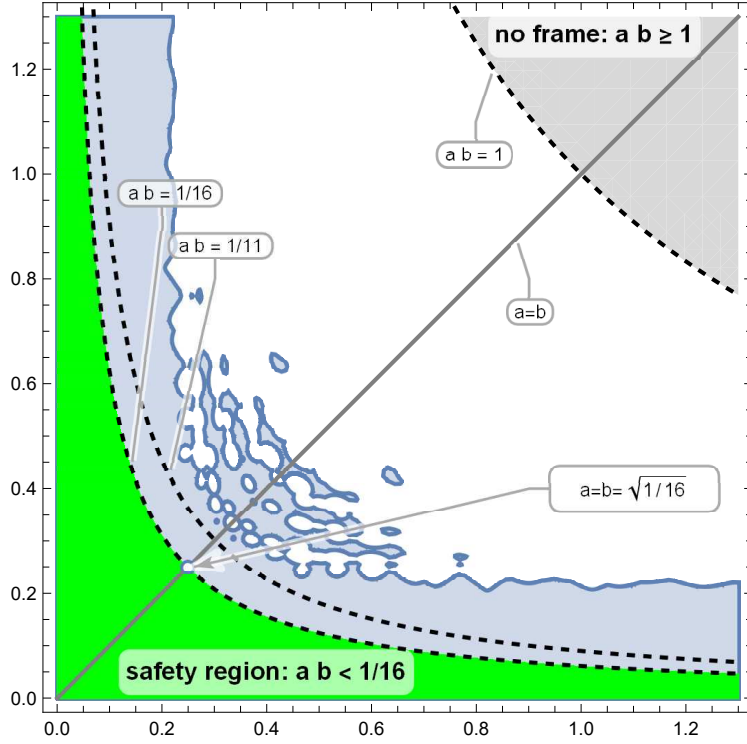


FIGURE 7. The safety region for h_{15} and a numerical extension provided by the Janssen test. For say $a, b \in [0, 1.3]$, which are the depicted parameters, parts of this extensions should be verifiable by hand, e.g., for $ab \leq 1/11$.

where the Janssen test seems to imply that $\mathcal{G}(h_n, 1/\sqrt{3} \mathbb{Z}^2)$ is a frame (see Fig. 8), are $n = 4$ (where the value does not exceed 1.59338) and $n = 9$, which we just proved. Due to [26, Thm. 7, Thm. 8], only Hermite functions of order $4m$ and $4m + 1$ can constitute a frame with a square lattice of density 3, in particular $\mathcal{G}(h_{15}, 1/\sqrt{3} \mathbb{Z}^2)$ is not a frame.

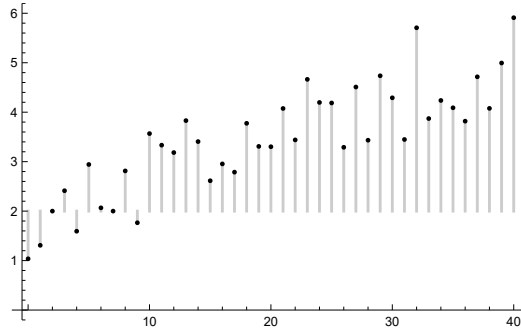


FIGURE 8. Values of the Janssen test for $\mathcal{G}(h_n, 1/\sqrt{3} \mathbb{Z}^2)$, $n = 0, \dots, 40$. Only for $n \in \{0, 1, 4, 9\}$ the Janssen test suffices to see that we have a frame.

REFERENCES

- [1] M. Abramowitz and I. A. Stegun, editors. *Handbook of mathematical functions, with formulas, graphs, and mathematical tables*. Dover Publications, Inc., New York, 1966.
- [2] M. Boon and J. Zak. Discrete coherent states on the von Neumann lattice. *Physical Review B*, 18(12):6744–6751, 1978.
- [3] M. Boon, J. Zak, and I. J. Zucker. Rational von Neumann lattices. *Journal of Mathematical Physics*, 24(2):316–323, 1983.
- [4] M. A. de Gosson. *The Wigner Transform*. World Scientific, Singapore, 2017.
- [5] M. Faulhuber. A short note on the frame set of odd functions. *Bulletin of the Australian Mathematical Society*, 98(3):481–493, December 2018.
- [6] M. Faulhuber. On the Parity under Metaplectic Operators and an Extension of a Result of Lyubarskii and Nes. *Results in Mathematics*, 75(8), 2020.
- [7] M. Faulhuber. Some curious results related to a result of Strohmer and Beaver. *Analysis and Applications*, 19:845–873, 2021.
- [8] H. G. Feichtinger. On a new Segal algebra. *Monatshefte für Mathematik*, 92(4):269–289, 1981.
- [9] H. G. Feichtinger. Gabor Expansions of Signals: Computational Aspects and Open Questions. In *Landscapes of Time-Frequency Analysis*, volume ATFA17 of *Applied and Numerical Harmonic Analysis*, page 173–206. Springer, 2019.
- [10] H. G. Feichtinger and N. Kaiblinger. Varying the time-frequency lattice of Gabor frames. *Trans. Am. Math. Soc.*, 356(5):2001–2023, 2004.
- [11] H. G. Feichtinger and W. Kozek. Quantization of TF lattice-invariant operators on elementary LCA groups. In *Gabor Analysis and Algorithms*, page 233–266. Springer, 1998.
- [12] H. G. Feichtinger and F. Luef. Wiener Amalgam Spaces for the Fundamental Identity of Gabor Analysis. *Collectanea Mathematica*, 57:233–253, 2006.
- [13] G. B. Folland. *Harmonic analysis in phase space*. Number 122 in Annals of Mathematics Studies. Princeton University Press, 1989.
- [14] Ghosh, Riya and Selvan, Antony. On Gabor frames generated by B-splines, totally positive functions, and Hermite functions. *arXiv*, 2304.11322, 2023.
- [15] K. Gröchenig and S. Koppensteiner. Gabor Frames: Characterizations and Coarse Structure. In A. Aldroubi, C. Cabrelli, S. Jaffard, and U. Molter, editors, *New Trends in Applied Harmonic Analysis, Volume 2*, Applied and Numerical Harmonic Analysis, pages 93–120. Springer, 2019.
- [16] K. Gröchenig. *Foundations of Time-Frequency Analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser, Boston, MA, 2001.
- [17] K. Gröchenig. The Mystery of Gabor Frames. *Journal of Fourier Analysis and Applications*, 20(4):865–895, 2014.
- [18] K. Gröchenig and Y. Lyubarskii. Gabor frames with Hermite functions. *Comptes Rendus Mathématique*, 344(3):157–162, 2007.
- [19] K. Gröchenig and Y. Lyubarskii. Gabor (super)frames with Hermite functions. *Mathematische Annalen*, 345(2):267–286, 2009.
- [20] A. Horst, J. Lemvig, and A. E. Videbaek. On the non-frame property of Gabor systems with Hermite generators and the frame set conjecture, 2023.
- [21] M. S. Jakobsen. On a (No Longer) New Segal Algebra: A Review of the Feichtinger Algebra. *Journal of Fourier Analysis and Applications*, 24(6):1579–1660, 2018.
- [22] A. J. E. M. Janssen. Some Weyl-Heisenberg frame bound calculations. *Indagationes Mathematicae*, 7(2):165–183, 1996.
- [23] A. J. E. M. Janssen. The duality condition for Weyl-Heisenberg frames. In H. G. Feichtinger and T. Strohmer, editors, *Gabor Analysis and Algorithms: Theory and Applications*, page 33–84. Birkhäuser Boston, Boston, MA, 1998.
- [24] I. Krasikov. Inequalities for Laguerre polynomials. *East J. Approx.*, 11(3):257–268, 2005.

- [25] I. Krasikov. Inequalities for orthonormal Laguerre polynomials. *J. Approx. Theory*, 144(1):1–26, 2007.
- [26] J. Lemvig. On some Hermite series identities and their applications to Gabor analysis. *Monatshefte für Mathematik*, 182(4):889–912, 2017.
- [27] Z. Lewandowski and J. Szynal. An upper bound for the Laguerre polynomials. *J. Comput. Appl. Math.*, 99(1-2):529–533, 1998.
- [28] Y. Lyubarskii. Frames in the Bargmann space of entire functions. In *Entire and Subharmonic Functions*, page 167–180. American Mathematical Society, Providence, RI, 1992.
- [29] Y. Lyubarskii and P. G. Nes. Gabor frames with rational density. *Applied and Computational Harmonic Analysis*, 34(3):488–494, 2013.
- [30] J. E. Moyal. Quantum mechanics as a statistical theory. *Proceedings of the Cambridge Philosophical Society*, 45:99–124, 1949.
- [31] P. Natalini and B. Palumbo. Inequalities for the incomplete gamma function. *Mathematical Inequalities and Applications*, 3(1):69–77, 2000.
- [32] H. E. Salzer and R. Zucker. Table of the zeros and weight factors of the first fifteen Laguerre polynomials. *Bulletin of the American Mathematical Society*, 55(10):1004–1012, 1949.
- [33] K. Seip. Density theorems for sampling and interpolation in the Bargmann – Fock space I. *Journal für die Reine und Angewandte Mathematik*, 429:91–106, 1992.
- [34] K. Seip and R. Wallstén. Density theorems for sampling and interpolation in the Bargmann – Fock space II. *Journal für die Reine und Angewandte Mathematik*, 429:107–114, 1992.
- [35] E. Skrettingland. On Gabor g-frames and Fourier series of operators. *Studia Mathematica*, 259(1):25–78, 2021.
- [36] G. Szegő. *Orthogonal Polynomials*, volume 23 of *Colloquium Publications*. American Mathematical Society, 4. edition, 1975.
- [37] R. Tolimieri and R. S. Orr. Poisson summation, the ambiguity function, and the theory of Weyl-Heisenberg frames. *Journal of Fourier Analysis and Applications*, 1(3):233–247, 1995.
- [38] T. Tschurtschenthaler. The Gabor Frame Operator (its Structure and Numerical Consequences). Master’s thesis, University of Vienna, 2000.
- [39] C. Wiesmeyr. *Construction of frames by discretization of phase space*. PhD thesis, University of Vienna, 2013.
- [40] Wolfram Research, Inc., Champaign, IL. *Mathematica, Version 13.3*, 2023.

3.3 Gabor frame bound optimizations

Bibliographic Information

Faulhuber, M., Shafkulovska, I., Gabor frame bound optimization. *Applied and Computational Harmonic Analysis* **67**, 101574 (2023)

Contribution

Markus Faulhuber and Irina Shafkulovska contributed equally to the idea, concept, development, and writing of the paper.

E-Print

arXiv:2204.02917

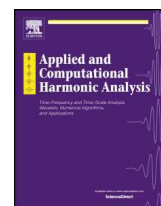
Supplementary material

Two supplementary materials are related to the paper. The first material is a collection of used technical results. The second material is an accompanying Mathematica code.



Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

journal homepage: www.elsevier.com/locate/achaGabor frame bound optimizations

Markus Faulhuber, Irina Shafkulovska *

NuHAG, Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria



ARTICLE INFO

Article history:

Received 11 April 2022

Received in revised form 24 January 2023

Accepted 3 July 2023

Available online 13 July 2023

Communicated by Thomas Strohmer

MSC:

primary 42C15

secondary 26A06, 33B10

Keywords:

Exponential functions

Frame bounds

Gabor frame

Hyperbolic functions

Lattice

ABSTRACT

We study sharp frame bounds of Gabor systems over rectangular lattices for different windows and integer oversampling rate. In some cases we obtain optimality results for the square lattice, while in other cases the lattices optimizing the frame bounds and the condition number are rectangular lattices which are different for the respective quantities. Also, in some cases optimal lattices do not exist at all and a degenerated system is optimal.

© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

We aim to find extremal lattices for the spectral bounds of Gabor systems with specific windows. The quantities which we aim to optimize are the lower and upper frame bound as well as their ratio, which is the condition number of the associated frame operator. We study the cases provided in [34], where Janssen computed sharp spectral bounds for Gabor frames over rectangular lattices of the form $a\mathbb{Z} \times b\mathbb{Z}$ for several different window functions and $(ab)^{-1} \in \mathbb{N}$, and we will build upon this work. Our findings are for rectangular lattices of integer density and for different windows. The main results can be summed up as follows:

* Markus Faulhuber was supported by the Austrian Science Fund (FWF) projects TAI6 and P33217. Irina Shafkulovska was supported by the Austrian Science Fund (FWF) project P33217. Parts of this work were established while Irina Shafkulovska was associated with the Acoustics Research Institute (ARI) at the Austrian Academy of Sciences (ÖAW).

* Corresponding author.

E-mail addresses: markus.faulhuber@univie.ac.at (M. Faulhuber), irina.shafkulovska@univie.ac.at (I. Shafkulovska).

- *Hyperbolic secant*: the square lattice optimizes the lower and upper frame bound simultaneously and, hence, also the condition number.
- *Cut-off exponentials*: we find that a lattice optimizing the frame bounds does not exist for cut-off exponentials supported on $[0, 1/b]$. For support $[0, 2/b]$, optimal lattices may or may not exist, depending on the lattice density and which quantity we seek to optimize. If they exist, then they depend on the density and on the decay parameter of the exponential.
- *One-sided exponential*: the optimizing lattices for the condition number and the frame bounds are different from each other and depend on the (over)sampling rate.
- *Two-sided exponentials*: each quantity has a unique optimizer. They depend on the (over)sampling rate and they all differ from each other.

Until now, Gaussians were the only family of window functions for which optimality results have been known [4], [10], [14]. We see that many of the situations are quite different from what we know about the Gaussian case. Therefore, our results emphasize that in general we may not expect that an optimal lattice exists at all (for the frame bounds or the condition number) and that a lattice optimizing one quantity does not necessarily optimize the others. Indeed, assume that the lower frame bound A and the upper frame bound B each have a unique critical point with respect to one free lattice parameter (others are fixed), so there exists a unique optimizing lattice. Critical points of the condition number B/A need to fulfill

$$\left(\frac{B}{A}\right)' = 0 \iff \frac{A'}{A} = \frac{B'}{B}, \quad (1.1)$$

where the differentiation is with respect to the free parameter. If the critical points of A and B do not coincide, then no critical point of B/A can coincide with either of the other two. As can be seen from the example of the two-sided exponential, this is not necessarily a pathology of the window not being in the modulation space $M^1(\mathbb{R})$, also known as Feichtinger's algebra $S_0(\mathbb{R})$. However, the situation for the hyperbolic secant in the rectangular case is comparable to the Gaussian case (see also [37] for a connection).

The rigorous analytic study of optimal lattices for Gabor systems is relatively new and, to the best of our knowledge, has only been carried out numerically prior to [14]. We are choosing a pedestrian approach for the analysis of the frame bounds' behavior, mainly using a critical point approach combined with careful interval estimates. The computations turn out to be rather technical and to avoid an overload of elementary (though delicate) calculations and estimates, we do not always give all details. However, the manuscript is self-contained and we also provide a supplement to this manuscript. We put numbered markers which tell where to find the respective calculation in the supplement.

This work is structured as follows. We present our results in Section 2. In Section 3 we settle the notation and provide some background information and motivation. Section 4 contains explanation on auxiliary techniques used in the manuscript. The proofs of our results follow in Section 5 for the hyperbolic secant, Section 6 for cut-off exponentials, Section 7 for one-sided exponentials and, finally, in Section 8 we present the proofs for two-sided exponentials.

2. The results

We will now present an overview of the results and also provide some background information. The complete proofs are given in the later sections of this work. All parameters are always assumed to be positive, in particular $(a, b) \in \mathbb{R}_+ \times \mathbb{R}_+$.

2.1. Hyperbolic secant

The frame property of the hyperbolic secant and lattices of the form $a\mathbb{Z} \times b\mathbb{Z}$ has been studied by Janssen and Strohmer [37]. By relating its Zak transform to the one of the Gaussian they were able to show that

$$\mathcal{G}(\text{sech}, a\mathbb{Z} \times b\mathbb{Z}) \text{ is a frame} \iff (ab)^{-1} > 1.$$

As the (properly scaled) hyperbolic secant is invariant under the Fourier transform, it follows by the general theory of symplectic matrices and metaplectic operators [19,20,22] that the two Gabor systems

$$\mathcal{G}(\text{sech}(\pi t), a\mathbb{Z} \times b\mathbb{Z}) \quad \text{and} \quad \mathcal{G}(\text{sech}(\pi t), b\mathbb{Z} \times a\mathbb{Z}) \quad \text{are unitarily equivalent.} \quad (2.1)$$

In particular, (2.1) says that the lattice parameters (a, b) and (b, a) yield the same frame bounds for the hyperbolic secant window $\text{sech}(\pi t)$ and that the associated frame operator is invariant under rotations of the lattice by integer multiples of 90 degrees. Therefore, for any fixed density n , the frame bounds as functions on $\{\Lambda_{a,b} \mid (ab)^{-1} = n\}$ are symmetric about the pair $(1/\sqrt{n}, 1/\sqrt{n})$, so the square lattice has to be a *local* extremum for both frame bounds. Theorem 2.1 states that it is actually the unique *global* optimum for integer oversampling.

Theorem 2.1. *Let $g(t) = (\frac{\pi}{2})^{1/2} \text{sech}(\pi t)$ be the normalized and Fourier-invariant hyperbolic secant and consider the rectangular lattice $\Lambda_{a,b} = a\mathbb{Z} \times b\mathbb{Z}$ of density n , i.e. $(ab)^{-1} = n$, with $2 \leq n \in \mathbb{N}$. By $A(a, b)$ and $B(a, b)$ we denote the sharp lower and upper frame bound of the Gabor system $\mathcal{G}(g, \Lambda_{a,b})$, respectively. Then, we have that*

$$A\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right) \geq A(a, b) \quad \text{and} \quad B\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right) \leq B(a, b),$$

with equality if and only if $a = b = \frac{1}{\sqrt{n}}$, i.e., if the lattice is the square lattice of density n .

A simple corollary of Theorem 2.1 can be obtained by a symplectic deformation result, which is in particular a dilation result (see Proposition 3.1)

Corollary 2.2. *Let $g_\gamma(t) = (\frac{\pi}{2\gamma})^{1/2} \text{sech}(\frac{\pi t}{\gamma})$ be the hyperbolic secant dilated by γ . Then, the optimizing lattice is $\frac{\gamma}{\sqrt{n}}\mathbb{Z} \times \frac{1}{\gamma\sqrt{n}}\mathbb{Z}$, i.e., the by γ dilated version of the square lattice.*

We remark that in the case of critical sampling, i.e., $(ab)^{-1} = 1$, we obtain

$$B(\gamma, \frac{1}{\gamma}) \leq B(a\gamma, \frac{b}{\gamma}) < \infty,$$

with equality if and only if $a = b = 1$. This can be deduced from the explicit expressions for B and combining results on Laplace-Beltrami operators on a flat torus [12] and results on Gaussian Gabor systems [14]. Since we have a finite upper frame bound, the lower frame bound must vanish identically in this case, as imposed by the Balian-Low theorem.

2.2. Cut-off exponentials

The frame set of this class of functions is not known and so we focus only on the cases from [34]. In these situations we also know that we have a frame. We consider the following families of functions

$$g_m(t) = C_{b,\gamma} e^{-\gamma t} \chi_{[0, m/b]}(t), \quad m \in \{1, 2\}, \quad \gamma \geq 0, \quad C_{b,\gamma} \text{ chosen such that } \|g_m\|_2 = 1.$$

2.2.1. The case $m = 1$

This particular family of windows is actually adapted to the lattice and we may perform an analysis depending on the decay parameter $\gamma > 0$.

Proposition 2.3. *Let $A(\gamma)$ and $B(\gamma)$ be the frame bounds of the Gabor system $\mathcal{G}(g_1, a\mathbb{Z} \times b\mathbb{Z})$ with $(ab)^{-1} = n \in \mathbb{N}$. Then, for any fixed lattice (hence also fixed n), the behavior of the frame bounds with respect to the decay parameter $\gamma > 0$ is*

$$A(\gamma) < n, \quad A'(\gamma) < 0 \quad \text{and} \quad B(\gamma) > n, \quad B'(\gamma) > 0.$$

The frame condition number is independent of n and always given by

$$\frac{B(\gamma)}{A(\gamma)} = e^{2a\gamma}.$$

From the calculations in the proof it will be obvious that the optimal window (for fixed a) is actually obtained for $\gamma \rightarrow 0$, which yields a tight frame with the box function. Moreover, the roles of γ and a are exchangeable. We may fix $\gamma > 0$ and then optimize with respect to a and the above results hold verbatim with γ replaced by a . Note that we then change the window and the lattice at the same time. The optimal system is degenerated ($a \rightarrow 0, b \rightarrow \infty$) and the window tends to a Dirac delta.

2.2.2. The case $m = 2$

We start with results for fixed lattices and optimization with respect to the decay parameter.

Proposition 2.4. *Let $A(\gamma)$ and $B(\gamma)$ be the frame bounds of the Gabor system $\mathcal{G}(g_2, a\mathbb{Z} \times b\mathbb{Z})$ with $(ab)^{-1} = n \in \mathbb{N}$. Then, for any fixed lattice (hence also fixed n), the behavior of the frame bounds with respect to the decay parameter $\gamma > 0$ is as follows:*

- (i) *The lower frame bound $A(\gamma)$ has a unique maximum which depends on n and $a = 1/(bn)$. The unique maximum is attained for the unique positive solution of the equation*

$$-1 + \frac{1}{\gamma a} - \coth(\gamma a) = n \coth\left(\frac{\gamma na}{2}\right) \operatorname{sech}(\gamma na).$$

- (ii) *For $n \in \{1, 2\}$, we have $B(\gamma) > 2$ and $B'(\gamma) > 0$.*
 (iii) *For $3 \leq n \in \mathbb{N}$, we have that B has a local maximum and a local minimum, which may be global (depending on n and a). Denoting the point of the local maximum and minimum by γ^* and γ_* , respectively, we have*

$$\gamma^* < \frac{\operatorname{arccosh}\left(\frac{1+\sqrt{5}}{2}\right)}{na} < \gamma_*.$$

If $B(\gamma_*) < 2$, then the minimum is global. Moreover, we have $B'(\gamma) > 0$ for all $\gamma > \gamma_*$.

From the computations in the proofs we will see that the parameters γ and a are exchangeable.

Proposition 2.5. *The frame condition number of the Gabor system $\mathcal{G}(g_2, a\mathbb{Z} \times b\mathbb{Z})$ with $(ab)^{-1} = n \in \mathbb{N}$ depends on n and is given by*

$$\frac{B}{A} = e^{2a\gamma} \coth\left(\frac{\gamma na}{2}\right)^2.$$

The condition number is minimal if and only if

$$\gamma a = \frac{\log(n + \sqrt{n^2 + 4}) - \log(2)}{n}.$$

We see that optimality for the decay parameter γ depends on the lattice parameter a , and vice versa, as well as on the (over)sampling rate n .

2.3. One-sided exponential

The one-sided exponential function is given by

$$g(t) = \sqrt{2\gamma} e^{-\gamma t} \chi_{\mathbb{R}_+}(t), \quad \gamma > 0.$$

We note that it is known that the Gabor system $\mathcal{G}(g, a\mathbb{Z} \times b\mathbb{Z})$ is a frame whenever $(ab)^{-1} \geq 1$ [34, Sec. 4] (see also [26]). Note that g is not continuous (thus $g \notin M^1(\mathbb{R})$), so having a frame at critical sampling rate is possible. Also, the one-sided exponential plays a central role when it comes to finding examples, other than the Gaussian, of zero-free ambiguity functions [24].

Theorem 2.6. Consider the Gabor system $\mathcal{G}(g, a\mathbb{Z} \times b\mathbb{Z})$ with $(ab)^{-1} = n \in \mathbb{N}$. For fixed n , the frame condition number is minimal only for the lattice $(a_0/n)\mathbb{Z} \times (1/a_0)\mathbb{Z}$ with

$$a_0 = \frac{\operatorname{arcsinh}(n)}{\gamma}.$$

The lower frame bound has a unique maximizing lattice $(a_+/n)\mathbb{Z} \times (1/a_+)\mathbb{Z}$ with

$$a_+ < \frac{\operatorname{arcsinh}(n)}{\gamma}.$$

The upper frame bound has a unique minimizing lattice $(a^-/n)\mathbb{Z} \times (1/a^-)\mathbb{Z}$ with

$$a^- > \frac{\operatorname{arcsinh}(n)}{\gamma}.$$

2.4. Two-sided exponential

Lastly, consider the two-sided exponential, which is given by

$$g(t) = \sqrt{\gamma} e^{-\gamma|t|}, \quad \gamma > 0.$$

The Gabor system $\mathcal{G}(g, a\mathbb{Z} \times b\mathbb{Z})$ is a frame whenever $(ab)^{-1} > 1$ [35, Sec. 5]. Note that the case of critical sampling is excluded this time. This is a consequence of the Balian-Low theorem as $g \in M^1(\mathbb{R})$. We also refer to [36] where windows with certain convexity properties on $\mathbb{R}_+$ have been studied, and the two-sided exponential falls into this class.

Theorem 2.7. Consider the Gabor system $\mathcal{G}(g, a\mathbb{Z} \times b\mathbb{Z})$ with $(ab)^{-1} = n \in \mathbb{N}$. For fixed n , the frame condition number is minimal only for the lattice $(a_0/n)\mathbb{Z} \times (1/a_0)\mathbb{Z}$ where a_0 is contained in some open interval depending on n :

$$a_0 \in \frac{1}{\gamma} I_n \subset \frac{1}{\gamma} \left(\frac{\eta_n}{2}, 2\eta_n \right), \quad \eta_n = 2 \operatorname{arccosh}(n).$$

The lower frame bound has a unique maximum, dependent on the oversampling rate $2 \leq n \in \mathbb{N}$ ($A = 0$ for $n = 1$). The maximizing lattice $(a_+/n)\mathbb{Z} \times (1/a_+)\mathbb{Z}$ satisfies $a_+ \in 1/\gamma I_n$.

The upper frame bound has a unique minimum, dependent on the oversampling rate $n \in \mathbb{N}$. The minimizing lattice $(a^-/n)\mathbb{Z} \times (1/a^-)\mathbb{Z}$ satisfies $a^- \in 1/\gamma I_n$.

We show analytically that $a^- < a_0 < a_+$, and for $n \geq 4$ we have $a^- < 2 \operatorname{arccosh}(n)/\gamma < a_+$.

3. Notation and preliminaries

We briefly clarify the setting. Our notation is more or less the same as in the textbook of Gröchenig [22] and the reader familiar with Gabor systems and frames may simply skip this section. For two functions $f, g \in L^2(\mathbb{R})$, the inner product and induced norm are given by

$$\langle f, g \rangle = \int_{\mathbb{R}} f(t) \overline{g(t)} dt \quad \text{and} \quad \|f\|_2^2 = \langle f, f \rangle.$$

Here $\bar{g}$ is the complex conjugation of g . The Fourier transform of a (suitable) function f is

$$\mathcal{F}f(\omega) = \int_{\mathbb{R}} f(t) e^{-2\pi i \omega t} dt.$$

The operator $\mathcal{F}$ extends to a unitary operator on $L^2(\mathbb{R})$. After having settled the notation, we come to defining the central objects of this work. We consider Gabor systems for the Hilbert space $L^2(\mathbb{R})$ over rectangular lattices:

$$\mathcal{G}(g, \Lambda_{a,b}) = \{\pi(\lambda)g \mid \lambda \in \Lambda_{a,b}\}, \quad \text{with} \quad \Lambda_{a,b} = a\mathbb{Z} \times b\mathbb{Z}.$$

Here $\pi(\lambda)$ is a unitary operator on $L^2(\mathbb{R})$, usually called a time-frequency shift by $\lambda \in \mathbb{R}^2$;

$$\pi(\lambda)g(t) = M_\omega T_x g(t) = g(t - x) e^{2\pi i \omega t}, \quad \lambda = (x, \omega).$$

The operators T_x and M_ω are the familiar translation and modulation operator, respectively. A Gabor system is a frame if and only if the frame inequality is fulfilled

$$A \|f\|_{L^2}^2 \leq \sum_{\lambda \in \Lambda_{a,b}} |\langle f, \pi(\lambda)g \rangle|^2 \leq B \|f\|_{L^2}^2, \quad \forall f \in L^2(\mathbb{R}), \quad (3.1)$$

for some positive constants $0 < A \leq B < \infty$, which we call frame bounds. Note that these bounds depend crucially on g as well as $\Lambda_{a,b}$. The middle expression in (3.1) is derived by sampling the short-time Fourier transform $V_g f$ on the lattice $\Lambda_{a,b}$:

$$V_g f(x, \omega) = \langle f, \pi(\lambda)g \rangle = \int_{\mathbb{R}} f(t) \overline{g(t - x)} e^{-2\pi i \omega t} dt.$$

Being a frame usually requires some redundancy of the system. The number $(ab)^{-1}$ yields the density of the lattice, which is the average number of lattice points per unit area. In the context of Gabor systems, we also speak of the (over)sampling rate. Comparably to the Nyquist rate for band-limited functions, we need $(ab)^{-1} \geq 1$. The best achievable constants in (3.1) are the spectral bounds of the associated Gabor frame operator which is given by

$$S_{g, \Lambda_{a,b}} f = \sum_{\lambda \in \Lambda_{a,b}} \langle f, \pi(\lambda)g \rangle \pi(\lambda)g.$$

In case of a Gabor frame, any element in our Hilbert space has a stable expansion of the form

$$f = \sum_{\lambda \in \Lambda_{a,b}} c_\lambda \pi(\lambda)g,$$

with $(c_\lambda) \in \ell^2(\Lambda_{a,b})$ (see also [31], [40]). Note that $(ab)^{-1} \geq 1$ is a necessary condition which may be far from sufficient for expansions of the above type.

We recall the notion of the frame set of a window function g [23] (see also [11]) for rectangular lattices. The rectangular (or reduced) frame set of $g \in L^2(\mathbb{R})$ is given by

$$\mathfrak{F}_r(g) = \{\Lambda_{a,b} = a\mathbb{Z} \times b\mathbb{Z} \mid \mathcal{G}(g, \Lambda_{a,b}) \text{ is a frame}\}.$$

Note that $\Lambda_{a,b} \in \mathfrak{F}_r(g)$ does not necessarily imply $\Lambda_{b,a} \in \mathfrak{F}_r(g)$ (counterexamples may easily be constructed by, e.g., using the box function), but for all $\Lambda_{a,b} \subseteq \mathbb{R}^2$ and all $g \in L^2(\mathbb{R})$ holds

$$\Lambda_{a,b} \in \mathfrak{F}_r(g) \iff \Lambda_{b,a} \in \mathfrak{F}_r(\mathcal{F}g).$$

This is a special case of the *symplectic invariance of Gabor systems* [19,20,22]. We formulate another (simplified) version only for dilations, which is sufficient for our purposes.

Proposition 3.1. *Let $g \in L^2(\mathbb{R})$ be given. We denote by D_γ ($\gamma > 0$) the unitary dilation $D_\gamma g(t) := \gamma^{-1/2}g(\gamma^{-1}t)$. Then for all $\Lambda_{a,b} \subseteq \mathbb{R}^2$ holds*

$$\Lambda_{a,b} \in \mathfrak{F}_r(g) \iff \Lambda_{\gamma a, \gamma^{-1}b} \in \mathfrak{F}_r(D_\gamma g).$$

Furthermore, the Gabor systems $\mathcal{G}(g, \Lambda_{a,b})$ and $\mathcal{G}(D_\gamma g, \Lambda_{\gamma a, b/\gamma})$ possess the same frame bounds.

A necessary condition, mentioned already several times, on $\Lambda_{a,b}$ to be contained in $\mathfrak{F}_r$ of any function in the Hilbert space $L^2(\mathbb{R})$ is given by the density theorem for Gabor systems (see also [28]) and can be summed up in the following way:

$$\mathfrak{F}_r(g) \subset \{\Lambda_{a,b} \mid (ab)^{-1} \geq 1\}.$$

Assuming that $g \in S_0(\mathbb{R}) = \{f \in L^2(\mathbb{R}) \mid V_f f \in L^1(\mathbb{R}^2)\}$, which is known as Feichtinger's algebra [15], [16] (see also [3], [29]), it is known that the frame set is open and contains a neighborhood of $\mathbf{0}$ [23]. This means that for any non-zero window from $S_0(\mathbb{R})$ there exists a sufficiently high oversampling rate $(ab)^{-1}$, depending on the window, such that Gabor frames exist. The case of density 1 is called critical density and only at this level Gabor orthonormal bases exist. However, for windows in Feichtinger's algebra the Balian-Low theorem ([1], [38], see also [22, Chap. 8.3]) shows that there exist no Gabor frames at critical density. For the hyperbolic secant, Janssen and Strohmer [37] could show that the rectangular frame set is the largest possible;

$$\mathfrak{F}_r(\text{sech}) = \{\Lambda_{a,b} \mid (ab)^{-1} > 1\}.$$

There are other functions for which we now know the rectangular frame set. Among these are totally positive functions of finite type (greater than or equal to 2) which all possess a full rectangular frame set [27]. Note that the two-sided exponential belongs to this class of functions and the hyperbolic secant is a totally

positive function of infinite type, and both are also in $S_0(\mathbb{R})$. Despite the fact that they both belong to the nice function space $S_0(\mathbb{R})$, they exhibit very different optimality properties. The one-sided exponential is a totally positive function of (finite) type 1 and not in $S_0(\mathbb{R})$. Its frame set is the largest possible, also containing all rectangular lattices at critical density. Another example is the indicator function of a finite interval and its frame set has an extremely complicated structure [6,7]. The first and still only window for which a full characterization of Gabor frames has been known is the Gaussian window [39], [44], [45]. For more general density results we refer to [30]. For more details on frames and Gabor systems we refer to the textbooks of Christensen [5] and Gröchenig [22] and for related results we also refer to the textbook of Folland [19].

After having determined the frame set of a function g , a natural follow-up question is on the optimality of a lattice $\Lambda_{a,b} \in \mathfrak{F}_r(g)$ of fixed density n . Several things can be meant by that. We could look for a maximizer of the lower bound $A(a,b)$, a minimizer of the upper bound $B(a,b)$ or a minimizer of the condition number $\kappa(a,b) = B(a,b)/A(a,b)$. Neither the existence nor the uniqueness of optimizing pairs (a,b) is an obvious matter. Even in the uniqueness case, the optimal lattices will not necessarily be the same for all quantities.

There are several theoretical ways to investigate the behavior of the frame bounds. In the case of integer or rational density, a popular method is the Zak transform [32]. Alternatively, one can turn to duality theory, where the ground work is due to Daubechies, Landau and Landau [9], Janssen [33], Ron and Shen [42] and Wexler and Raz [47]. For a more thorough treatise of the topic we refer to [25].

In [34], Janssen computed the optimal frame bounds for 6 different windows. We built upon those computations and determine the critical points with respect to the lattice parameters. The case of the Gaussian and rectangular lattices has already been treated by Faulhuber and Steinerberger in [14], where optimality of the square lattice (in all three senses) under rectangular lattices of integer density has been shown. An optimality result for the upper bound and the hexagonal lattice (even density) followed [10] (see also [41]) and lastly Bétermin, Faulhuber and Steinerberger [4] proved the optimality of the lower bound for the hexagonal lattice (even density), which also implies the optimality of the hexagonal lattice for the condition number. In what follows, we will see that this does not seem to be a general pattern.

4. Preparations

4.1. The hyperbolic functions

Hyperbolic functions will be omnipresent throughout the manuscript. We will only consider them restricted to the domain $[0, \infty)$. The monotonicity properties of hyperbolic functions can be read easily from their Taylor expansions or the relation to the exponential function. Truncations of the Taylor series also provide good upper and lower bounds. We mention these rather simple facts because the proofs which follow require evaluating and bounding these functions multiple times at various points. To verify the estimates, it would often suffice to evaluate the corresponding tenth Taylor polynomial. Of course, the reader is also welcome to use their preferred software. For numerical checks, we used Mathematica [48], which evaluates these functions to arbitrary precision (we used a precision to the 24th digit)¹. Still, our arguments are elementary and can also be checked by hand, but the bookkeeping is delicate at several points and makes the proofs non-trivial and rather technical.

Besides the various elementary identities involving the derivatives or squares of the hyperbolic functions², which can all be found in [21], we will also use the following equation [2, eq. 86]

¹ The Mathematica notebook is provided for download as an ancillary file.

² Some identities and explanations are provided in the supplement.

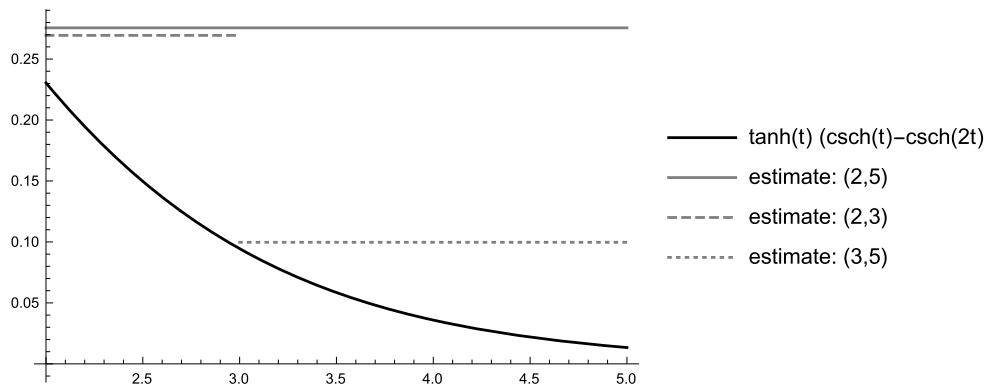


Fig. 1. Estimate of the above type for $\tanh(t)(\operatorname{csch}(t) - \operatorname{csch}(2t))$ on the interval $(2, 5)$. In order to get more appropriate estimates we may use the sub-intervals $(2, 3)$ and $(3, 5)$.

$$x \sum_{k=-\infty}^{\infty} \operatorname{sech}(kx)^2 = 2 + \frac{2}{x} \sum_{k=0}^{\infty} \operatorname{csch} \left(\left(k + \frac{1}{2} \right) \frac{1}{x} \right)^2. \quad (4.1)$$

This and many other equations in [2] illuminate the deeper connection of the hyperbolic functions to elliptic functions and elliptic integrals, which also appear to be closely linked to Gabor frames with the Gaussian window (see [34], compare also [13]).

4.2. Interval estimate

We will often speak of interval estimates. By that we mean bounding an expression consisting of monotonic functions (or otherwise well-known bounded functions) on an interval $[x_0, y_0]$ by the worst possible value. In some of our calculations we will use the symbols $\nearrow$ and $\searrow$ to denote that an expression is increasing or decreasing, respectively. For example, $\tanh$ is strictly monotonically increasing and csch is strictly monotonically decreasing, both strictly positive. Then $\operatorname{csch}(t) - \operatorname{csch}(2t) > 0$ and

$$\underbrace{\tanh(t)}_{\nearrow} \underbrace{(\operatorname{csch}(t) - \operatorname{csch}(2t))}_{\searrow} < \tanh(y_0) (\operatorname{csch}(x_0) - \operatorname{csch}(2y_0)), \quad t \in (x_0, y_0),$$

which gives us as a rough upper bound as in Fig. 1 (if the end point is 0 or ∞ , we would take the limit, presuming it exists). The arrows indicate the monotonicity properties we applied for the estimate. Of course, the sign of each factor also goes into the overall estimate.

Proofs

5. The hyperbolic secant

In this section we will prove Theorem 2.1. By Proposition 3.1, we only need to prove it for the case $\gamma = 1$ in order to obtain Corollary 2.2. We will show that the lower frame bound, the upper frame bound and the condition number of the Gabor system $\mathcal{G}(\sqrt{\pi/2} \operatorname{sech}(\pi t), a\mathbb{Z} \times b\mathbb{Z})$ are optimal if and only if $a = b$, assuming $2 \leq (ab)^{-1} \in \mathbb{N}$. We begin with a transformation of the problem. We use the formula provided by Janssen [34, eq. (7.8)] (with the additional normalizing factor $\frac{\pi}{2}$) to express the frame bounds. To simplify the expressions, we set

$$f_A(t) = t \sum_{k=0}^{\infty} \operatorname{sech}(\alpha_k t)^2 \quad \text{and} \quad f_B(t) = t \sum_{k=0}^{\infty} \operatorname{csch}(\alpha_k t)^2, \quad t \in (0, \infty),$$

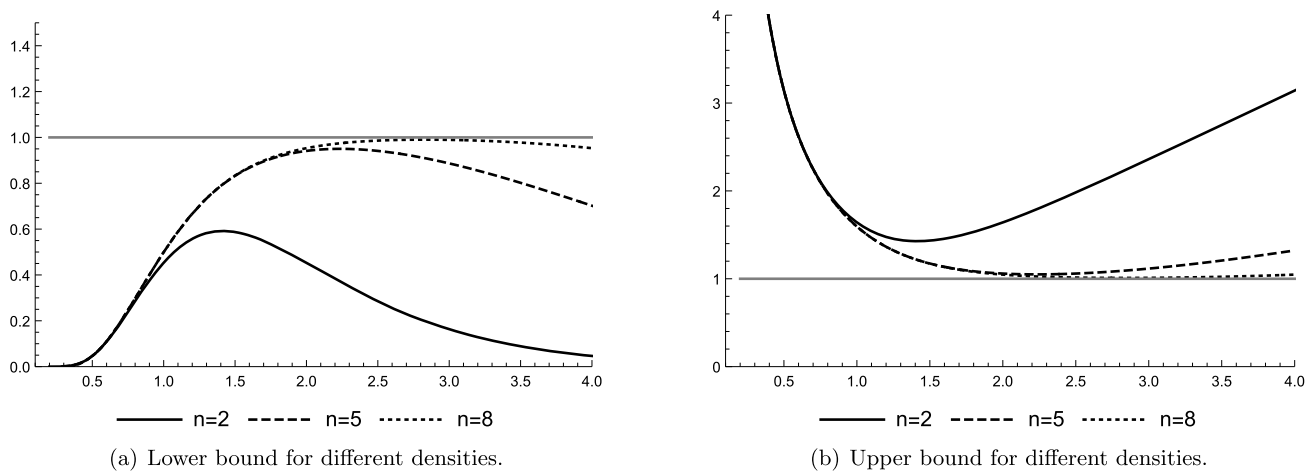


Fig. 2. The re-normalized optimal bounds $n^{-1}A(\eta/n, 1/n)$ and $n^{-1}B(\eta/n, 1/n)$ for densities $n \in \{2, 5, 8\}$. As n grows, the bounds become extremely flat around the optimum. For both cases the optimizing parameter η (depending on n) yields a square lattice of density n .

where $\alpha_k = \pi(k + 1/2)$. We use a change of variables which respects the fact that $(ab)^{-1} = n$:

$$(a, b) \mapsto \left(\frac{\eta}{n}, \frac{1}{\eta}\right). \quad (5.1)$$

The exact frame bounds can now be expressed by (compare [34, eq. (7.8)])

$$\begin{aligned} A\left(\frac{\eta}{n}, \frac{1}{\eta}\right) &= \frac{n\pi}{2} \sum_{k=-\infty}^{\infty} \frac{\eta}{n} \operatorname{sech}\left(\pi\left(k + \frac{1}{2}\right) \frac{\eta}{n}\right)^2 + \frac{n\pi}{2} \sum_{k=-\infty}^{\infty} \frac{1}{\eta} \operatorname{sech}\left(\pi\left(k + \frac{1}{2}\right) \frac{1}{\eta}\right)^2 - n \\ &= n\pi f_A\left(\frac{\eta}{n}\right) + n\pi f_A\left(\frac{1}{\eta}\right) - n, \\ B\left(\frac{\eta}{n}, \frac{1}{\eta}\right) &= \frac{n\pi}{2} \sum_{k=-\infty}^{\infty} \frac{\eta}{n} \operatorname{sech}\left(\pi k \frac{\eta}{n}\right)^2 + \frac{n\pi}{2} \sum_{k=-\infty}^{\infty} \frac{1}{\eta} \operatorname{sech}\left(\pi k \frac{1}{\eta}\right)^2 - n \\ &= n\pi f_B\left(\frac{\eta}{n}\right) + n\pi f_B(\eta) - n \end{aligned}$$

where we used the symmetry of sech and Identity (4.1).

The series defining f_A and f_B converge locally uniformly. By first taking formal derivatives (taking term-wise derivatives within the series) we see that also these series converge locally uniformly for all formal derivatives. Hence, f_A and f_B are infinitely continuously differentiable and we are allowed to simply differentiate term-wise (see [43, Chap. 7] for technical details). The quantity η describes the lattice geometry and we have a square lattice if $\eta = n^{1/2}$. We note that optimizing the bounds gets more and more difficult as the density grows since the curves get flatter and flatter as indicated in Fig. 2.

Our strategy is the following: From the algebraic structure of the problem, it follows immediately that $\eta = n^{1/2}$ gives a critical point, which by (5.1) shows that the square lattice is critical in the set of rectangular lattices. Then, we will prove certain monotonicity properties of the involved terms and their derivatives to exclude further critical points.

Determining the critical points is achieved by finding all values $\eta \in \mathbb{R}_+$ such that

$$\begin{aligned} \frac{\eta}{n\pi} \frac{d}{d\eta} A\left(\frac{\eta}{n}, \frac{1}{\eta}\right) &= \frac{\eta}{n} f'_A\left(\frac{\eta}{n}\right) - \frac{1}{\eta} f'_A\left(\frac{1}{\eta}\right) = h_A\left(\frac{\eta}{n}\right) - h_A\left(\frac{1}{\eta}\right) = 0, \\ \frac{\eta}{n\pi} \frac{d}{d\eta} B\left(\frac{\eta}{n}, \frac{1}{\eta}\right) &= -\frac{\eta}{\eta} f'_B\left(\frac{\eta}{n}\right) + \eta f'_B(\eta) = -h_B\left(\frac{\eta}{n}\right) + h_B(\eta) = 0. \end{aligned} \quad (5.2)$$

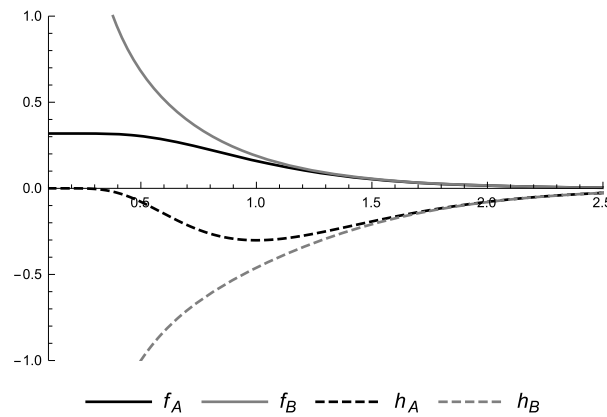


Fig. 3. Comparison of the behavior of f_A and f_B as well as h_A and h_B .

Here we used the abbreviation

$$h_A(t) = t f'_A(t) \quad \text{and} \quad h_B(t) = t f'_B(t).$$

Obviously, the expressions in (5.2) vanish for $\eta = n^{1/2}$, which corresponds to the square lattice of density n . We seek to show that there are no other critical points to prove Theorem 2.1. So far, the lower and upper bound's dependence on f_* and h_* , as well as the structure of f_* and h_* seem to be nearly the same. Yet, the accumulation of derivatives of sech and csch secretly amplifies several differences (see also Fig. 3).

5.1. The lower bound

In [2], we find the identity (abbreviated with our notation above)

$$\pi f_A(t) + \pi f_A(t^{-1}) = 1, \quad (5.3)$$

which implies the point symmetry $h_A(t) = h_A(t^{-1})$. Rather remarkably, we could have concluded (5.3) with only our knowledge about Gabor frames at critical density (a similar remark is given in [34]): $A(\eta, 1/\eta) = \pi f_A(\eta) + \pi f_A(\eta^{-1}) - 1 = 0$ because of the Balian-Low theorem [1], [38] (see also [22, Chap. 8.4]). We will now prove that the following equivalence.

Lemma 5.1. *For all $x, y \in (0, \infty)$ holds*

$$h_A(x) = h_A(y) \quad \Longleftrightarrow \quad x = y \vee x = y^{-1}. \quad (5.4)$$

Proof. One implication is rather simple: the case $x = y$ is trivial and the case $x = y^{-1}$ follows from the already established point symmetry of h_A . We start the proof of the other implication with the observation that due to the symmetry, it suffices to prove the strict monotonicity on $(1, \infty)$ to exclude further solutions for (5.4). We split our computations in two parts. The monotonicity is straightforward on the interval $(1.03, \infty)$. On the part $(1, 1.04)$, we prove that h_A is strictly convex, i.e., h'_A is strictly increasing. Taking into account that $h'_A(1) = 0$, due to the symmetry, $h'_A(t) > 0$ for all $t > 1$.

Monotonicity away from 1: Firstly, we compute the derivative³ of $h_a(t) = t f'_A(t)$.

$$h'_A(t) = \sum_{k=0}^{\infty} \operatorname{sech}(\alpha_k t)^2 \left(1 - 2\alpha_k^2 t^2 \operatorname{sech}(\alpha_k t)^2 + 4\alpha_k^2 t^2 \tanh(\alpha_k t)^2 - 6\alpha_k t \tanh(\alpha_k t) \right).$$

³ The full computation is provided in the supplement. Computation 3 is used twice.

We split the expression in the brackets in two. The goal is to determine points which satisfy

$$1 - 2x^2 \operatorname{sech}(x)^2 > 0 \quad \text{and} \quad 4x^2 \tanh(x)^2 - 6x \tanh(x) \geq 0.$$

The second inequality is fulfilled as long as $x \tanh(x) \geq \frac{3}{2}$, which is a strictly monotonically increasing function in $x > 0$. By evaluating, one can check that the solution of $x \tanh(x) = \frac{3}{2}$ lies in $(1.6218, 1.6219)$. For the former inequality, we observe

$$\begin{aligned} \lim_{x \rightarrow 0} 1 - 2x^2 \operatorname{sech}(x)^2 &= \lim_{x \rightarrow \infty} 1 - 2x^2 \operatorname{sech}(x)^2 = 1, \\ \frac{d}{dx} [1 - 2x^2 \operatorname{sech}(x)^2] &= 4x \operatorname{sech}(x)^2 (x \tanh(x) - 1). \end{aligned}$$

Since the fixed point of $\coth$ lies in $(1.19, 1.2)$, this yields the only critical point of $x \mapsto 1 - 2x^2 \operatorname{sech}(x)^2$ for $x > 0$. This must yield the minimum. Therefore, it holds that

$$1 - 2x^2 \operatorname{sech}(x)^2 > 1 - 2 \cdot 1.2^2 \cdot \operatorname{sech}(1.19)^2 > 0.1 > 0, \quad \forall x > 0.$$

As $\alpha_k t \geq \alpha_0 t$, for all $t \geq 1.6219 \alpha_0^{-1} = \frac{3.2438}{\pi} \in (1.03, 1.04)$, we see that $h'_A(t) > 0$ for $t > 1.03$.

Strict convexity close to 1: Since $h'_A(1) = 0$, trying to find positive lower bounds for h'_A close to 1 is playing a losing game, so we turn to the second derivative³ of h_A .

$$\begin{aligned} h''_A(t) &= 2 \sum_{k=0}^{\infty} \alpha_k \operatorname{sech}(\alpha_k t)^2 \times \left(8\alpha_k^2 t^2 \operatorname{sech}(\alpha_k t)^2 \tanh(\alpha_k t) - 5\alpha_k t \operatorname{sech}(\alpha_k t)^2 \right. \\ &\quad \left. - 4 \tanh(\alpha_k t) + 10\alpha_k t \tanh(\alpha_k t)^2 - 4\alpha_k^2 t^2 \tanh(\alpha_k t)^3 \right). \end{aligned}$$

We do a similar separation as above.

Part 1: Recall that $\alpha_k \geq \frac{\pi}{2}$ for $k \geq 0$. For all $t \in [1, 1.04]$, holds (due to $x \tanh(x) \nearrow$)

$$\begin{aligned} &8\alpha_k^2 t^2 \operatorname{sech}(\alpha_k t)^2 \tanh(\alpha_k t) - 5\alpha_k t \operatorname{sech}(\alpha_k t)^2 \\ &\geq \operatorname{sech}(\alpha_k t)^2 \left(8 \frac{\pi}{2} \tanh\left(\frac{\pi}{2}\right) - 5 \right) > 6 \operatorname{sech}(\alpha_k t)^2 > 0. \end{aligned}$$

Part 2: We compute

$$\begin{aligned} &\alpha_k \operatorname{sech}(\alpha_k t)^2 \left(-4 \tanh(\alpha_k t) + 10\alpha_k t \tanh(\alpha_k t)^2 - 4\alpha_k^2 t^2 \tanh(\alpha_k t)^3 \right) \\ &= 2\alpha_k \operatorname{sech}(\alpha_k t)^2 \tanh(\alpha_k t) p(\alpha_k t \tanh(\alpha_k t)), \end{aligned}$$

where $p(x) = -2x^2 + 5x - 2 = -2(x - \frac{1}{2})(x - 2)$. First of all, $p(x) > -3x^2$ on $(1, \infty)$. By examining the graph of p and taking into account $t \in (1, 1.04)$, one can verify⁴

$$\begin{aligned} p(\alpha_0 t) \tanh(\alpha_0 t) &> 0, \\ p(\alpha_k t) \tanh(\alpha_k t) &< 0, \quad k \in \mathbb{N}. \end{aligned}$$

We estimate

$$0 \geq 2\alpha_k \tanh(\alpha_k t) \operatorname{sech}(\alpha_k t) p(\alpha_k t \tanh(\alpha_k t)) > -12(\alpha_k t)^3 e^{-\alpha_k t} > -16.14,$$

⁴ The full computation is provided in the supplement.

where the last estimate is due to the fact that $|\tanh(x)| < 1$, $e^{-x} \leq \operatorname{sech}(x)$ and the maximum of $x^3 e^{-x}$ is attained at $x = 3$. To sum up, for all $t \in (1, 1.04)$ we have⁵

$$\frac{1}{2}h_a''(t) > 3 \cdot 0.07 - 16.14 \cdot \frac{0.04}{2.46} + 0.19 > 0.13 > 0.$$

All in all, we conclude that h_a is strictly monotonically increasing on $(1, \infty)$ and due to the symmetry it has to be strictly monotonically decreasing on $(0, 1)$. Let now $x, y \in (0, \infty)$ be two distinct points satisfying $h_A(x) = h_A(y)$. Without loss of generality, we can assume $x < y$. By the established monotonicity, $x < 1 < y$, implying also $1 < x^{-1}$. So, we can write

$$h_A\left(\frac{1}{x}\right) = h_A(x) = h_A(y).$$

The monotonicity on $(1, \infty)$ now implies $x^{-1} = y$. $\square$

So, by (5.4) of the above lemma we have

$$h_A\left(\frac{\eta}{n}\right) = h_A\left(\frac{1}{\eta}\right) \iff \frac{\eta}{n} = \frac{1}{\eta} \vee \frac{\eta}{n} = \eta \iff \eta = \sqrt{n} \vee n = 1.$$

The case of critical density ($n = 1$) is trivial and A vanishes identically in this case. For all other densities $n \in \mathbb{N}$, $\eta = \sqrt{n}$ is the only critical point. To conclude the statement on the lower frame bound, recall that for $(ab)^{-1} = n$, we used the map $(a, b) \mapsto (\eta/n, 1/\eta)$. Hence, the square lattice $n^{-1/2}\mathbb{Z} \times n^{-1/2}\mathbb{Z}$ yields the unique maximum of A for density $2 \leq n \in \mathbb{N}$.

5.2. The upper bound

Contrary to h_A , the function h_B possesses no symmetries, but is strictly monotonically increasing. We will split $(0, \infty)$ in multiple intervals. There is no denial that there are far better estimates than the following ones. In fact, numerical inspection of the function ψ suggest $\psi > 1$, while we only show its strict positivity. However, the estimates here are very simple and easy to verify. Trying to improve them would likely make the conditions under which the interval splitting is obtained significantly more complicated. At least two cases are probably unavoidable, since the function under consideration ψ behaves rather differently for very small and for very big arguments. Our main strategy is similar.

Lemma 5.2. *For all $x, y \in (0, \infty)$ holds*

$$h_B(x) = h_B(y) \iff x = y.$$

Proof. We show that $h_B'(t) > 0$ for all $t \in (0, \infty)$, which then gives the proof.

We begin with computing the first derivative⁶:

$$h_B'(t) = \sum_{k=0}^{\infty} \operatorname{csch}(\alpha_k t)^2 \psi(\alpha_k t),$$

where $\psi(x) = 1 - 2x^2 + 6x^2 \coth(x)^2 - 6x \coth(x)$. If we prove that ψ is strictly positive on $(0, \infty)$, then so is h_B' .

⁵ The full computation is provided in the supplement.

⁶ The full computation is provided in the supplement.

Part 1: $x > \log(\sqrt{21})$. Due to the monotonicity of $\exp$, one can easily convince themselves that $1.5 < \log(\sqrt{21}) < 1.6$. Since $\coth$ is strictly monotonically decreasing, and satisfies $\coth(x) = 1 + 2/(\exp(2x) - 1)$, we estimate as follows for $x > \log(\sqrt{21})$:

$$\psi(x) > 1 - 2x^2 + 6x^2 \cdot 1^2 - 6x \cdot 1 + \frac{2}{21-1} > 4 \cdot (1.5 - 0.825)^2 - 1.7225 = 0.1 > 0.$$

Part 2: Close to zero, $\coth(x)$ is unbounded, so we have to take into account the interaction with x . We rewrite ψ on $(0, 1.6)$ as

$$\psi(x) = q_1(x) + 6q_2(x \coth(x)), \quad q_1(x) = 1 - 2x^2 \quad \text{and} \quad q_2(x) = x(x - 1).$$

Clearly, q_2 is strictly monotonically decreasing on $(0, \infty)$, whereas q_1 is strictly monotonically increasing and strictly positive on $(1, \infty)$. To that, $x \coth(x)$ is strictly monotonically increasing. The interval estimate of ψ on an interval $(x_0, y_0) \subseteq (0, 1.6)$ will satisfy

$$\psi(x) > q_1(y_0) + 6q_2(x_0 \coth(x_0)) > 0,$$

if we choose x_0, y_0 well. One can convince themselves of the inequality⁷ with interval estimates on $(0, \sqrt{0.5})$, $(0.7, 1.01)$, $(1, 1.3)$ and $(1.29, 1.6)$ (as described in Subsection 4.2). All in all, h'_B is strictly positive on $(0, \infty)$. $\square$

We now consider what Lemma 5.2 means for the optimality of the upper bound B . We have established

$$h_B\left(\frac{\eta}{n}\right) = h_B\left(\frac{1}{\eta}\right) \iff \frac{\eta}{n} = \frac{1}{\eta} \iff \eta = \sqrt{n}.$$

Furthermore, for all $\eta > \sqrt{n}$ we have

$$\frac{d}{d\eta} B\left(\frac{\eta}{n}, \frac{1}{\eta}\right) > 0.$$

Due to the symmetry about $\sqrt{n}$, the square lattice $n^{-1/2}\mathbb{Z} \times n^{-1/2}\mathbb{Z}$ is the unique minimizing lattice of B among rectangular lattices of density $n \in \mathbb{N}$.

As A is maximized by the square lattice $n^{-1/2}\mathbb{Z} \times n^{-1/2}\mathbb{Z}$ and B is minimized by the square lattice $n^{-1/2}\mathbb{Z} \times n^{-1/2}\mathbb{Z}$, also the condition number $\kappa = B/A$ has to be minimal in this case.

In contrast to the Gaussian case, we do not know sufficient density conditions for the hyperbolic secant and general lattices to yield Gabor frames. So, before speculating about an optimality result for general lattices one should work out the lattice frame set of the hyperbolic secant. Nonetheless, we note that the conjecture of Strohmer and Beaver [46] was based on the radial symmetry of the Gaussian in the time-frequency plane and connections to sphere packings (see also [4]). By the characterization of radially symmetric ambiguity functions, which can only come from Gaussians and Hermite functions [19], it is clear that the hyperbolic secant does not possess a radial symmetry in the time-frequency plane.

⁷ The full computation is provided in the supplement.

6. Cut-off exponentials

6.1. Support $1/b$

In this section we study frame bounds of cut-off exponentials with support in $[0, 1/b]$ (or any interval of length $1/b$) for the lattice $\Lambda_{a,b} = a\mathbb{Z} \times b\mathbb{Z}$ with density $(ab)^{-1} \geq 1$. We denote the cut-off exponential with decay parameter $\gamma > 0$ by

$$g_{b,\gamma}(t) = C_{b,\gamma} e^{-\gamma t} \chi_{[0,1/b]}(t),$$

where $C_{b,\gamma}$ is a normalizing factor, which is explicitly given by $C_{b,\gamma} = \sqrt{2\gamma/(1 - e^{-2\gamma/b})}$. The bounds have been computed by Janssen [34] without the factor $C_{b,\gamma}$, so we only need a small adjustment compared to his result. We also refer to [8, Sec. 3.4.4] where similar computations have been carried out. After proper re-normalization we end up with

$$A = n \frac{2\gamma a}{1 - e^{-2\gamma a}} e^{-2\gamma a} \quad \text{and} \quad B = n \frac{2\gamma a}{1 - e^{-2\gamma a}}, \quad \gamma > 0. \quad (6.1)$$

We may include the case of the box function ($\gamma = 0$) by a limiting procedure and by l'Hôpital's rule we obtain $A = B = n$. As we always fix $(ab)^{-1}$, we usually have one free parameter to optimize the frame bounds. This time we fix both lattice parameters, a and b , and consider a family of windows parametrized by $\gamma > 0$.

6.1.1. The lower frame bound

We fix the oversampling rate n and consider the lower frame bound A as a function of γ . Taking the derivative yields the expression

$$A'(\gamma) = -2an \frac{e^{2a\gamma}(2a\gamma - 1) + 1}{(e^{2a\gamma} - 1)^2} < 0.$$

Verifying that the above expression is indeed negative is a simple exercise: to show that the numerator is positive, just observe that, for $x \in \mathbb{R} \setminus \{0\}$, $e^x > (1 + x)$, which is its linearization at 0. Thus, numerator and denominator are positive and there is a minus sign in front.

6.1.2. The upper frame bound

With a very similar observation we also obtain that

$$B'(\gamma) = 2an \frac{e^{2a\gamma}(e^{2a\gamma} - 1 - 2a\gamma)}{(e^{2a\gamma} - 1)^2} > 0.$$

From Fig. 4 one may get the impression that the bounds deteriorate slower for the square lattice than for rectangular lattices. This is, however, only the case as long as $a > b$, which we deduce from (6.1).

6.1.3. The condition number

Finally, we easily observe that the condition number is

$$\frac{B}{A} = e^{2a\gamma}.$$

We give some concluding remarks for the parameters of the family of the window $g_{b,\gamma}$. Fixing any rectangular lattice of density n , we see that the box function ($\gamma = 0$) is actually optimal within the family of localization

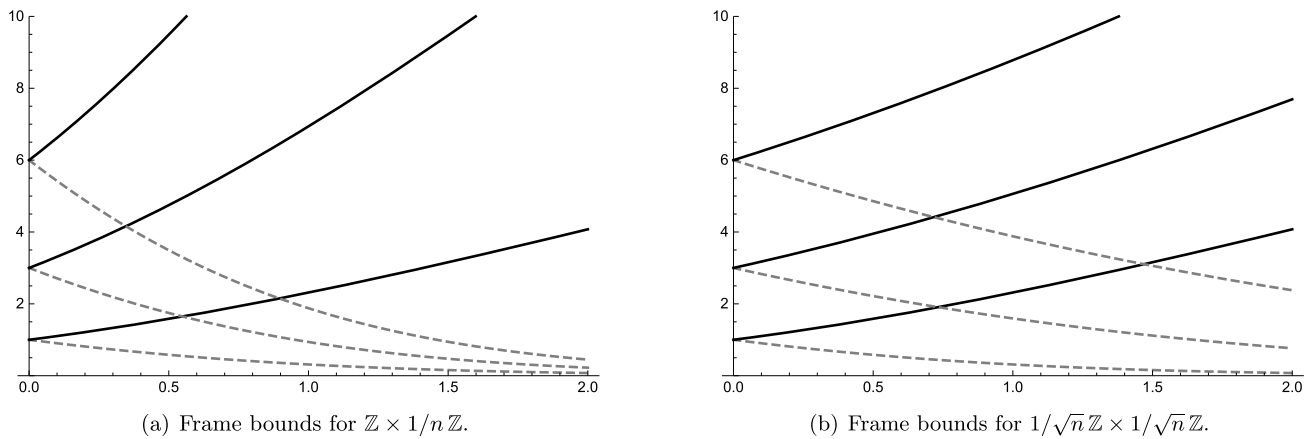


Fig. 4. Frame bounds of the Gabor systems $\mathcal{G}(g_\gamma, a\mathbb{Z} \times b\mathbb{Z})$ for $(ab)^{-1} = n \in \{1, 3, 6\}$ and $a = 1$, $b = 1/n$ and $a = b = \sqrt{1/n}$. For $\gamma = 0$, we always have a tight frame with bounds $A = B = n$. The upper frame bounds (black) are increasing with γ whereas the lower frame bounds (gray, dashed) are decreasing.

windows $g_{b,\gamma}$. Thus, the additional decay without additional smoothness in the time variable yields to even worse behavior in the frequency variable. Therefore, the frame condition is only made worse by increasing the decay parameter γ .

From (6.1) it is obvious that the decay parameter γ and the lattice parameter a are exchangeable for our computations. Note, however, that we change the window and the lattice at the same time. For any fixed γ the lower bound decreases and the upper bound increases as a grows. For $a \rightarrow 0$ we obtain a tight frame. This may be interpreted in a distributional sense: For $b \rightarrow \infty$ ($\Leftrightarrow a \rightarrow 0$), the function $g_{b,\gamma}$ tends to (a multiple of) the Dirac delta. Hence, even though we use an infinitely wide step size in the frequency domain, we know the function at every instance in time. The frame operator is therefore (a multiple of) the identity operator. In particular, an optimal lattice in terms of the frame bounds (or the frame condition number) does not exist.

6.2. Support $2/b$

Similarly to the previous case, we study frame bounds of cut-off exponentials, but this time with support in $[0, 2/b]$ for the lattice $\Lambda_{a,b} = a\mathbb{Z} \times b\mathbb{Z}$ of density $(ab)^{-1} \geq 1$. We use (more or less) the same notation as in the previous section and denote the window by

$$g_{b/2,\gamma}(t) = C_{b/2,\gamma} e^{-\gamma t} \chi_{[0,2/b]}(t),$$

where $C_{b/2,\gamma}$ is the normalizing factor, which is explicitly given by $C_{b/2,\gamma} = \sqrt{2\gamma/(1 - e^{-4\gamma/b})}$. We skip the computational details again and refer to Janssen [34, Sec. 3] where the bounds have been computed explicitly. With the proper normalization, the frame bounds are then

$$A = C_{b/2,\gamma}^2 \frac{a(1 - e^{-2\gamma/b})(1 - e^{-\gamma/b})^2}{1 - e^{-2\gamma a}} e^{-2a\gamma} \quad \text{and} \quad B = C_{b/2,\gamma}^2 \frac{a(1 - e^{-2\gamma/b})(1 + e^{-\gamma/b})^2}{1 - e^{-2\gamma a}}.$$

By using the explicit expression for $C_{b/2,\gamma}$ and the fact that $(ab)^{-1} = n$ ($\in \mathbb{N}$), we get

$$A = \frac{2\gamma a(1 - e^{-2\gamma na})(1 - e^{-\gamma na})^2}{(1 - e^{-2\gamma a})(1 - e^{-4\gamma na})} e^{-2a\gamma} \quad \text{and} \quad B = \frac{2\gamma a(1 - e^{-2\gamma na})(1 + e^{-\gamma na})^2}{(1 - e^{-2\gamma a})(1 - e^{-4\gamma na})}.$$

For $\gamma \rightarrow 0$, the bounds' behavior is $A \rightarrow 0$ and $B \rightarrow 2$. In the other direction, i.e., for $\gamma \rightarrow \infty$ we have $A \rightarrow 0$ and $B \rightarrow \infty$. So, in both limiting cases we do not have a frame. This stands in contrast to the previous case,

where we had a tight frame (orthonormal basis) with $A = B = n$ for the integer lattice $n^{-1/2}\mathbb{Z} \times n^{-1/2}\mathbb{Z}$ and $\gamma = 0$ (the box function).

6.2.1. The lower bound

Our next goal is to determine optimizing parameters for the lower bound. We start with re-writing the expression as

$$A = \underbrace{\gamma a (\coth(\gamma a) - 1)}_{f(\gamma)} \underbrace{(1 - \operatorname{sech}(\gamma na))}_{h(\gamma)}.$$

Determining critical points with respect to γ is equivalent to finding points where the sum of the logarithmic derivatives of f and h vanishes (because f and h are positive) or where:

$$\frac{f'(\gamma)}{f(\gamma)} = -\frac{h'(\gamma)}{h(\gamma)}. \quad (6.2)$$

We compute

$$\frac{f'(\gamma)}{f(\gamma)} = -a + \frac{1}{\gamma} - a \coth(\gamma a) = -a + \frac{1}{\gamma} \underbrace{(1 - \gamma a \coth(\gamma a))}_{<0, \searrow},$$

which is a negative and strictly decreasing function of γ (in fact it is less than $-a$). For the other side of (6.2), we compute

$$\frac{h'(\gamma)}{h(\gamma)} = na \coth\left(\frac{\gamma na}{2}\right) \operatorname{sech}(\gamma na).$$

This is the product of two positive, strictly decreasing functions and hence positive and strictly decreasing. It follows that the right-hand side of (6.2) is negative and strictly increasing and therefore (6.2) has at most one solution. By considering the (directional) limits

$$\lim_{x \rightarrow 0^+} \coth(x/2) \operatorname{sech}(x) = \infty \quad \text{and} \quad \lim_{x \rightarrow \infty} \coth(x/2) \operatorname{sech}(x) = 0,$$

we see that (6.2) must have exactly one solution and so the lower bound has a unique maximum (see Fig. 5). The optimal parameter γ , which is the unique solution of (6.2), depends on the density n of the lattice and the lattice geometry (the parameter a fixes b). We note that, from a computational point of view, the decay parameter γ and the lattice parameter a are exchangeable. Hence, we also find that there is a unique lattice maximizing the lower frame bound. The lattice geometry, i.e., the parameter a , depends on the density and the decay parameter γ and is (for any fixed n and γ), the unique solution of $f'(a)/f(a) = -h'(a)/h(a)$.

6.2.2. The upper bound

We will repeat the same kind of calculations for the upper frame bound. We re-write it as

$$B = \underbrace{\gamma a (\coth(\gamma a) + 1)}_{\tilde{f}(\gamma)} \underbrace{(1 + \operatorname{sech}(\gamma na))}_{\tilde{h}(\gamma)}.$$

Any critical points must satisfy the equation

$$\frac{\tilde{f}'(\gamma)}{\tilde{f}(\gamma)} = -\frac{\tilde{h}'(\gamma)}{\tilde{h}(\gamma)}. \quad (6.3)$$

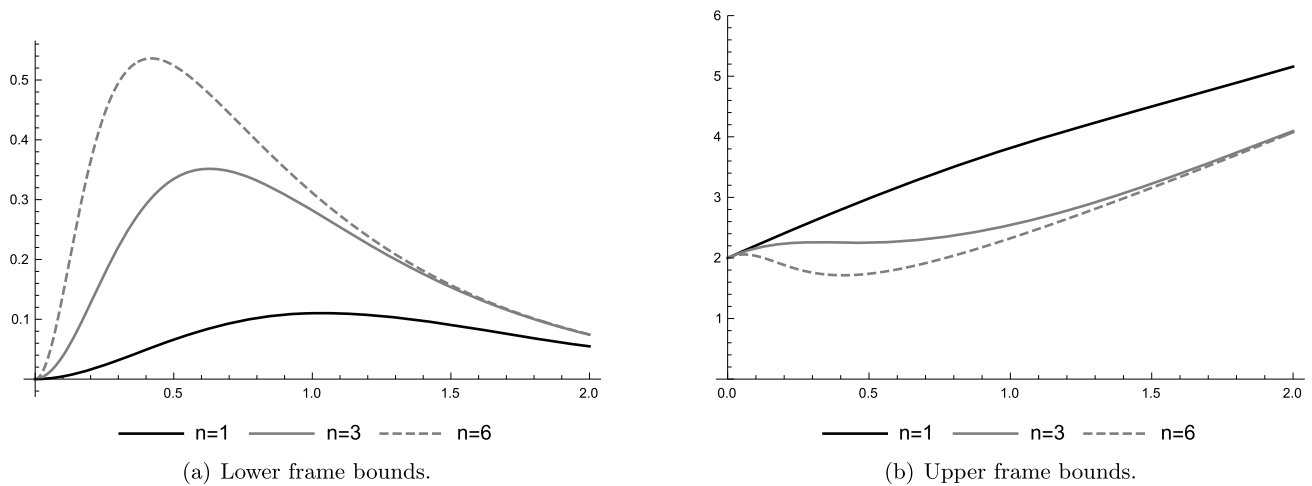


Fig. 5. Lower frame bounds and upper frame bounds in dependence of the lattice parameter a for $\gamma = 1$ and $n \in \{1, 3, 6\}$. We see that A has a unique maximum, whereas B may not have a minimizing lattice (the lattice is degenerated for $a = 0$).

The left-hand side of (6.3) is independent of n and is given by

$$\frac{\tilde{f}'(\gamma)}{\tilde{f}(\gamma)} = a + \frac{1}{\gamma} - a \coth(\gamma a) = a + \underbrace{a \left(\frac{1}{\gamma a} - \coth(\gamma a) \right)}_{0 >, \searrow, > -1},$$

which is positive and decreasing (in fact, it is less than a). The right-hand side is given by

$$-\frac{\tilde{h}'(\gamma)}{\tilde{h}(\gamma)} = na \operatorname{sech}(\gamma na) \tanh\left(\frac{\gamma na}{2}\right).$$

This function is positive and we may use the logarithmic derivative to check the sign of the derivative. We obtain, after some simplifications by using identities for hyperbolic functions,

$$\frac{d}{d\gamma} \left[\log \left(na \operatorname{sech}(\gamma na) \tanh\left(\frac{\gamma na}{2}\right) \right) \right] = na (\operatorname{csch}(\gamma na) - \tanh(\gamma na)),$$

which has a unique 0. Since $-\tilde{h}'(\gamma)/\tilde{h}(\gamma)$ has only 1 critical point, we conclude that this is the unique global maximum by the asymptotic behavior. As the right-hand side of (6.3) decays by an exponential factor faster than the left-hand side, we conclude that, depending on n , (6.3) has either 0, 1 or 2 solutions (see Fig. 6). We show that for $n \in \{1, 2\}$, there is no solution and for $n \geq 3$ we find 2 solutions: dividing (6.3) by $a > 0$ leaves us with the equation

$$1 + \frac{1}{\gamma a} - \coth(\gamma a) = n \operatorname{sech}(\gamma na) \tanh\left(\frac{\gamma na}{2}\right), \quad (6.4)$$

where the left-hand side is strictly decreasing and at most 1 and where the right-hand side has a unique global maximum. The point of the maximum is attained where the logarithmic derivative vanishes. This brings us to solving the equation

$$\operatorname{csch}(\gamma na) = \tanh(\gamma na) \iff \cosh(\gamma na)^2 - \cosh(\gamma na) - 1 = 0.$$

The quadratic polynomial in $\cosh$ has exactly one positive solution, namely:

$$\gamma na = \operatorname{arccosh}\left(\frac{1+\sqrt{5}}{2}\right) \approx 1.06128 \dots$$

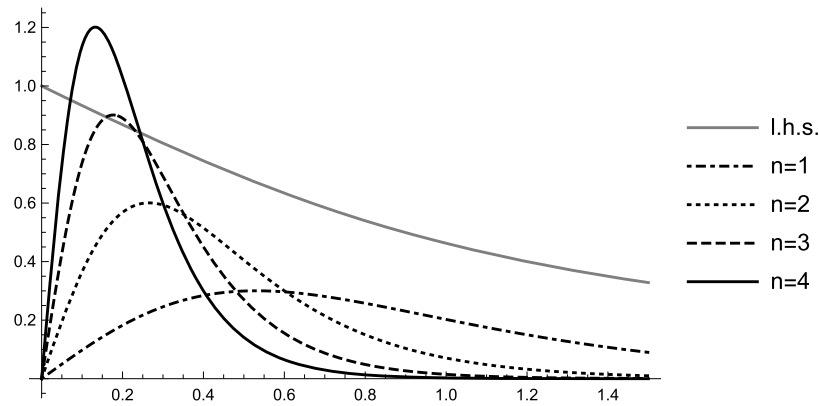


Fig. 6. Illustration of the two sides of equation (6.4) for fixed $a = 2$ and $n \in \{1, 2, 3, 4\}$. The illustration suggests that for any $n \geq 3$ we have two solutions in γ for equation (6.4).

We set $y = \gamma na$ and to imply the correctness of (6.4) we want to determine those n where

$$n \operatorname{sech}(y) \tanh\left(\frac{y}{2}\right) > 1.$$

Note that the solution depends on the product $\gamma na = y$ and so the optimal parameter γ depends on the lattice geometry (the parameter a) and the lattice density (the parameter n). We evaluate $\operatorname{sech}(y) \tanh(y)$ close to its maximum, i.e., close to $y = \operatorname{arccosh}\left(\frac{1+\sqrt{5}}{2}\right)$. To simplify computations, we set $y = 1$, which is sufficiently close to the point of interest. After an elementary manipulation of the last inequality, this leaves us with solving

$$n \tanh\left(\frac{1}{2}\right) > \cosh(1),$$

which could be done numerically, but also with comparable effort analytically. We want to show that, for n sufficiently large,

$$n \tanh\left(\frac{1}{2}\right) = n - \frac{2n}{e+1} > \frac{e+e^{-1}}{2} = \cosh(1).$$

It is not difficult to establish $e+1 > 3.7$ and $3.5 > e+e^{-1}$. So, it suffices to find n such that

$$2n - \frac{4n}{3.7} > 3.5,$$

which holds for all $n \geq 4$. Hence, for $n \geq 4$ the upper bound B (as a function of γ) has two critical points. By the behavior of the bound, the smaller value yields a local maximum and the larger value a minimum, which may be global, depending on n . For $n \in \{1, 2, 3\}$ we evaluate both sides of (6.4) with Mathematica at $\operatorname{arccosh}\left(\frac{1+\sqrt{5}}{2}\right)$ and see that (6.3) does not have a solution for $n \in \{1, 2\}$ and two solutions for $n = 3$. Again, γ may be exchanged for a . So, for $n \in \{1, 2\}$ we find that there is no optimal lattice minimizing B (the minimizing lattice degenerates) and for $n \geq 3$ there exist two critical lattices, one which is a local maximizer and the other a (local) minimizer, which is the only candidate for the global minimizer. As $B(0) = 2$, the minimizer is global if its value is less than 2 (compare Fig. 5).

6.2.3. The condition number

What is left to study is the condition number, which is simply given by the expression

$$\frac{B(\gamma)}{A(\gamma)} = e^{2\gamma a} \coth\left(\frac{\gamma na}{2}\right)^2. \quad (6.5)$$

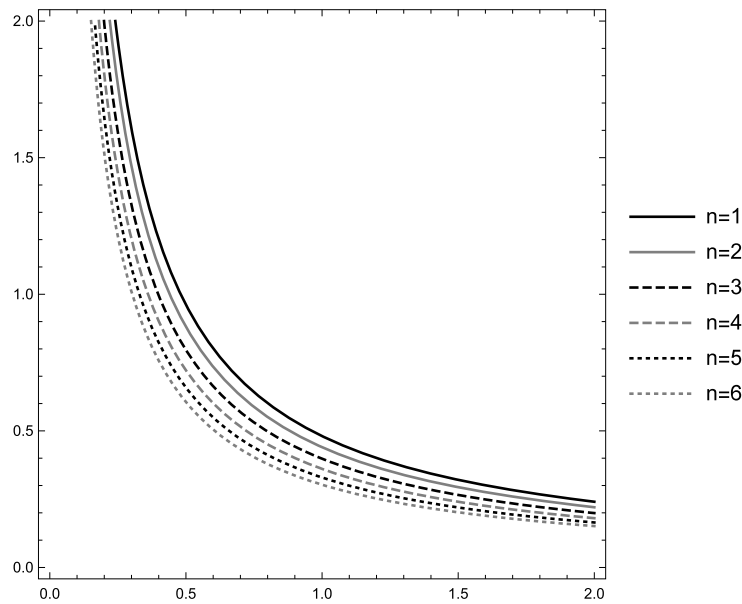


Fig. 7. Curves of optimal parameters (a, γ) for different lattice densities $n \in \{1, 2, 3, 4, 5, 6\}$.

Taking the derivative with respect to γ and looking for critical points leads to the equation

$$\frac{2a e^{2\gamma a} (e^{2n\gamma a} - n e^{n\gamma a} - 1)}{(e^{a\gamma n} - 1)^3} = 0 \iff e^{2n\gamma a} - n e^{n\gamma a} - 1 = 0.$$

Setting $x = e^{n\gamma a}$ leaves us with solving a simple quadratic equation: $x^2 - nx - 1 = 0$. There is only one positive solution to this equation, which is $x = (n + \sqrt{n^2 + 4})/2$. Therefore, the critical points of the condition number are attained for

$$\gamma = \frac{\log(n + \sqrt{n^2 + 4}) - \log(2)}{a n}. \quad (6.6)$$

Note that the dependence on the lattice parameter a is the same as for the decay parameter γ , which can easily be seen from (6.5). Moreover, we see that (6.6) describes a curve in the space of parameters (a, γ) yielding the optimal pair of decay and lattice parameters in dependence of the oversampling rate n (see Fig. 7). Lastly, we deduce from (6.5) that B/A is decreasing as a function of $n \in \mathbb{N}$ and

$$\frac{B}{A} \rightarrow e^{2\gamma a} \quad \text{for } n \rightarrow \infty.$$

So, for general $\gamma > 0$ and non-degenerate lattice ($a > 0$), the frame operator does not converge to a (multiple of) the identity operator, which however holds for any window $g \in S_0(\mathbb{R})$ [18].

This fact seems to be accompanied by the results in [17] (compare also [22, p. 132]). We remark that exchanging the parameters a and γ comes from the algebraic expressions in the calculations and is not a dilation result by applying a metaplectic operator.

7. The one-sided exponential

In this section we study frame bounds of one-sided exponentials. The lattice parameters satisfy $(ab)^{-1} = n \in \mathbb{N}$. In particular, we obtain a frame at the critical density as shown by Janssen [34, Sec. 4]. The function of interest is

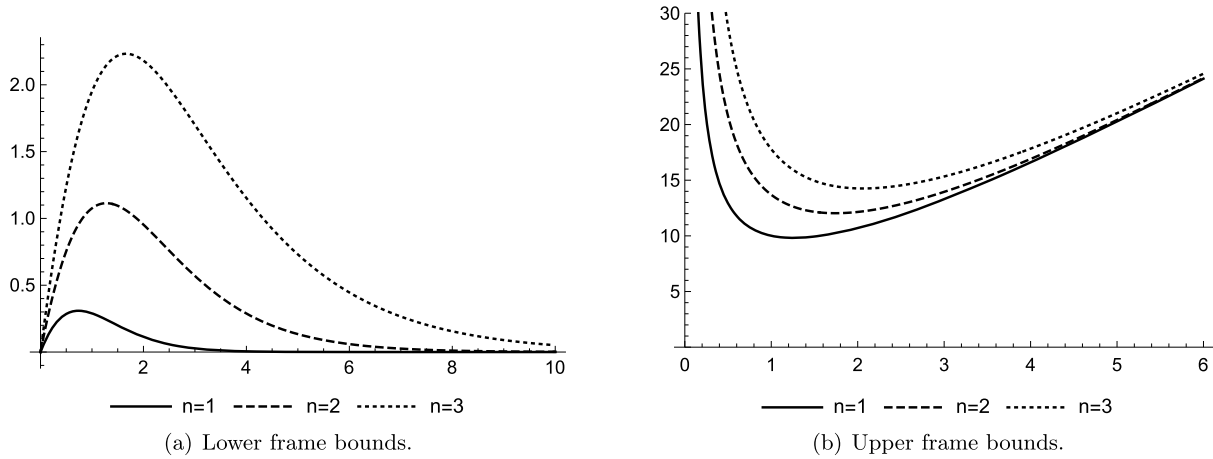


Fig. 8. Lower and upper bounds for the parameter $\gamma = 1$ and density $n \in \{1, 2, 3\}$.

$$g_\gamma(t) = \sqrt{2\gamma} e^{-\gamma t} \chi_{(0,\infty)}(t), \quad t \in \mathbb{R},$$

where $\gamma > 0$ is the dilation parameter. By Proposition 3.1, it suffices to consider the case $\gamma = 1$. Considering the dependencies as presented in [34], we will express the lattice pair again as $(\eta/n, 1/\eta)$, $\eta > 0$. Then the frame bounds are illustrated in Fig. 8 and given by

$$A\left(\frac{\eta}{n}, \frac{1}{\eta}\right) = 2\eta \tanh\left(\frac{\eta}{2}\right) \operatorname{csch}\left(\frac{\eta}{n}\right) e^{-\frac{\eta}{n}} \quad \text{and} \quad B\left(\frac{\eta}{n}, \frac{1}{\eta}\right) = 2\eta \coth\left(\frac{\eta}{2}\right) \operatorname{csch}\left(\frac{\eta}{n}\right) e^{\frac{\eta}{n}}.$$

7.1. The condition number

This time, we start with the condition number which has an obvious minimizing lattice. We easily compute

$$\kappa\left(\frac{\eta}{n}, \frac{1}{\eta}\right) = \left(\coth\left(\frac{\eta}{2}\right) e^{\frac{\eta}{n}}\right)^2,$$

which is minimal if and only if $\sqrt{\kappa}$ is minimal. We compute the derivative of $\sqrt{\kappa}$:

$$\frac{d}{d\eta} \left(\coth\left(\frac{\eta}{2}\right) e^{\frac{\eta}{n}} \right) = \frac{1}{n} e^{\frac{\eta}{n}} \operatorname{csch}^2\left(\frac{\eta}{2}\right) (\sinh(\eta) - n).$$

The sign of the derivative is determined by $\sinh$, which is a strictly monotonically increasing, unbounded, and attains 0 at 0. Therefore, the unique minimum of the condition number κ is attained at $\eta = \operatorname{arcsinh}(n)$. Next, we will show that this point lies between the maximizer of the lower and the minimizer of the upper bound.

7.2. The lower bound

For $\gamma = 1$, we start with computing the first derivative of the lower bound $A = A_1$ and show that it has a unique zero.

$$\frac{d}{d\eta} A\left(\frac{\eta}{n}, \frac{1}{\eta}\right) = e^{-\frac{\eta}{n}} \operatorname{csch}\left(\frac{\eta}{n}\right) \tanh\left(\frac{\eta}{2}\right) \left(1 + \eta \operatorname{csch}(\eta) - \frac{\eta}{n} - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right)\right),$$

where the last equality is due to the addition theorems⁸. Furthermore, the function

$$\eta \mapsto 1 + \eta \operatorname{csch}(\eta) - \frac{\eta}{n} - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right), \quad n \in \mathbb{N},$$

is a sum of strictly monotonically decreasing functions (f and $-x - x \coth(x)$). Taking into account the monotonicity in n , it suffices to verify there is a $\eta \in (0, \infty)$ with $A(\eta/n, \eta^{-1}) > 0$. One valid choice is $\eta = 0.5$. On the other hand, the function is unbounded from below:

$$\lim_{\eta \rightarrow \infty} \eta \operatorname{csch}(\eta) - \frac{\eta}{n} - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) = -\infty.$$

Therefore, there exists a unique $\eta_{A,n} > 0$ such that the lower bound $A(\frac{\eta}{n}, \frac{1}{\eta})$ has a unique maximum in $\eta_{A,n}$. We check whether it is $\eta = \operatorname{arcsinh}(n)$, which we know optimizes the condition number:

$$1 + \eta \operatorname{csch}(\eta) - \frac{\eta}{n} - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) = 1 + \frac{\eta}{n} - \frac{\eta}{n} - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) = 1 - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) < 0.$$

It follows that $\eta_{A,n} < \operatorname{arcsinh}(n)$. It is also simple to see that $(\eta_{A,n})_{n \in \mathbb{N}}$ is strictly monotonically increasing: $\eta_{A,n}$ is the unique solution to

$$1 + \eta \operatorname{csch}(\eta) = \frac{\eta}{n} + \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right).$$

The left-hand side is independent of n whereas the right-hand side is strictly decreasing in n .

7.3. The upper bound

We deal with the upper bound in the same manner and set $B = B_1$.

$$\frac{d}{d\eta} B\left(\frac{\eta}{n}, \frac{1}{\eta}\right) = e^{\frac{\eta}{n}} \operatorname{csch}\left(\frac{\eta}{n}\right) \coth\left(\frac{\eta}{2}\right) \left(1 + \frac{\eta}{n} - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) - \eta \operatorname{csch}(\eta)\right)$$

due to the addition theorems⁹. We again analyze the expression in the brackets. To begin with, $-x \operatorname{csch}(x)$ is a strictly increasing function. A curve discussion¹⁰ and employing the strict monotonicity of $x \coth(x)$, and $x \coth(x) \rightarrow 1$ for $x \rightarrow 0$, shows that $1 + x - x \coth(x)$ is a strictly concave, strictly monotonically increasing function. We now ensure the existence of a unique critical point of B :

$$\lim_{\eta \rightarrow 0^+} 1 + \frac{\eta}{n} - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) - \eta \operatorname{csch}(\eta) = 1 + 0 - 1 - 1 = -1 < 0.$$

If we evaluate the same expression at $2n$, we get

$$1 - 2n \operatorname{csch}(2n) + \frac{2n}{n} - \frac{2n}{n} \coth\left(\frac{2n}{n}\right) \geq 3 - 2 \coth(2) - 2 \operatorname{csch}(2) > 0.$$

We could again check whether $\eta = \operatorname{arcsinh}(n)$ is a critical point. From the general theory, though, we already know that it has to be between $\eta_{A,n}$ and $\eta_{B,n}$ and we have already established that it is greater than $\eta_{A,n}$.

⁸ The full computation is provided in the supplement.

⁹ The full computation is provided in the supplement.

¹⁰ The full computation is provided in the supplement.

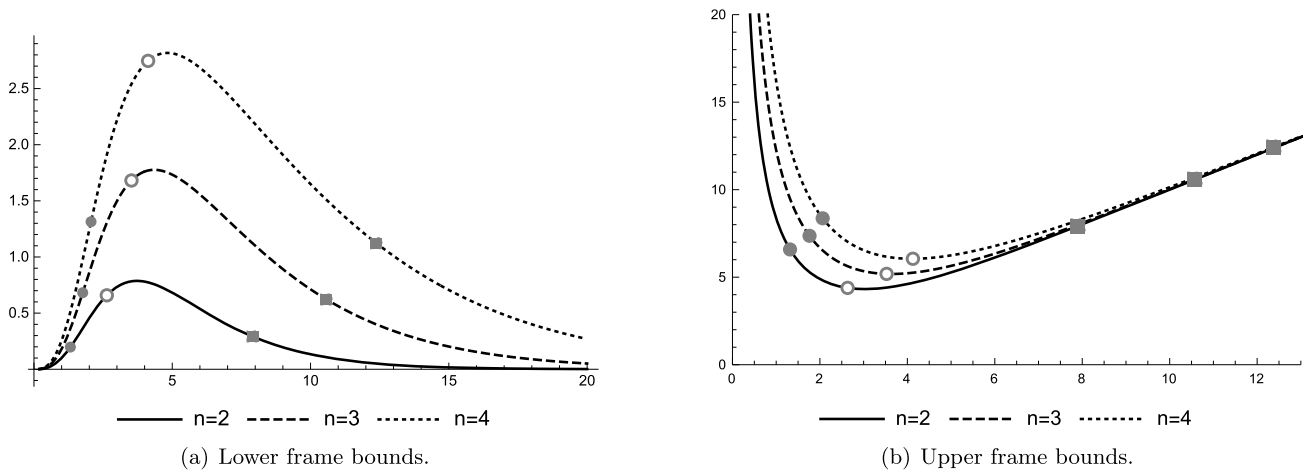


Fig. 9. The idea of the cut-offs for the lower and upper bound and parameter $\gamma = 1$ and densities $n \in \{2, 3, 4\}$. The $\bullet$ and $\blacksquare$ mark left and right cut-offs, respectively and $\circ$ marks η_n .

8. The two-sided exponential

The last section is dedicated to the two-sided exponential. At first sight, the bounds look quite similar to those already analyzed. However, this one has proven to be the most challenging to inspect. The current window function is

$$g_\gamma(t) = \sqrt{\gamma} e^{-\gamma|t|}, \quad t \in \mathbb{R},$$

where $\gamma > 0$ is the dilation parameter. By Proposition 3.1, it suffices to consider the case $\gamma = 1$. Unlike the one-sided exponential, the two-sided one is in Feichtinger's algebra, so we do not have a frame at the critical density. It has been shown in [34] that we obtain a frame for densities larger than 1. With the same parametrization $(\eta/n, 1/\eta)$ as in the previous sections, the frame bounds are given by

$$\begin{aligned} A\left(\frac{\eta}{n}, \frac{1}{\eta}\right) &= \tanh\left(\frac{\eta}{2}\right) \left(\frac{\eta}{n} \operatorname{csch}\left(\frac{\eta}{n}\right) - \eta \operatorname{csch}(\eta) \right), \\ B\left(\frac{\eta}{n}, \frac{1}{\eta}\right) &= \coth\left(\frac{\eta}{2}\right) \left(\frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) + \eta \operatorname{csch}(\eta) \right). \end{aligned} \quad (8.1)$$

When one looks at the graphs of the lower or the upper bound for a fixed density in Fig. 9, it seems obvious that there is a unique extremum. However, the first and second derivative of the concerned bound are somewhat difficult to analyze over $(0, \infty)$, as they are combinations of products and sums which exhibit different behaviors at different intervals. The visually “clear” situation is due to the interaction altogether, so dissecting each part separately leads to unsatisfactory estimates. Therefore, to determine the global extrema and to show their uniqueness, we will first exclude most of the positive half-axis through interval estimates (see Section 4.2 and Fig. 9) as yielding potential global extrema. Then for the lower bound, upper bound and condition number we will establish certain concavity, convexity and log-convexity results, respectively, on the small remaining interval. These will allow us to conclude that the extremum is unique in each case. For the lower bound, the cases for small n , which is $3 \leq n \leq 10$, and in particular $n = 3$, will need more attention at some points compared to large n . The case $n = 2$ needs to be handled completely separately, since it does not agree with the asymptotic positions of the extremal points. In the sequel, we will set

$$f(t) = t \operatorname{csch}(t), \quad \forall t > 0, \quad (8.2)$$

so as to have more compact expressions and we will repeatedly use the following: f is strictly monotonically decreasing and asymptotically satisfies¹¹

$$\lim_{x \rightarrow 0} x \operatorname{csch}(x) = 1 \quad \text{and} \quad \lim_{x \rightarrow \infty} x \operatorname{csch}(x) = 0.$$

8.1. The lower bound

The initial reference point will be $\eta_n = 2 \operatorname{arccosh}(n)$. We will first compare the value attained here with the function's values on $(0, \eta_n/2]$ and $[2\eta_n, \infty)$ in order to exclude global maxima in these intervals. Then we will show that there is a unique maximum in $(\eta_n/2, 2\eta_n)$. By using (8.2) and addition formulas, we write the lower frame bound evaluated at $\eta_n = 2 \operatorname{arccosh}(n)$ as

$$A\left(\frac{\eta_n}{n}, \frac{1}{\eta_n}\right) = \frac{(n^2 - 1)^{1/2}}{n} \left(f\left(\frac{\eta_n}{n}\right) - \frac{\eta_n}{2n(n^2 - 1)^{1/2}} \right) = \frac{(n^2 - 1)^{1/2}}{n} f\left(\frac{\eta_n}{n}\right) - \frac{\operatorname{arccosh}(n)}{n^2}. \quad (8.3)$$

8.1.1. The left cut-off

We will show that for any $3 \leq n \in \mathbb{N}$ the value of the lower bound achieved at $\eta_n = 2 \operatorname{arccosh}(n)$ is larger than any value attained in $(0, \eta_n/2]$. For small n we will later only be able to show a concavity result for slightly smaller intervals (depending on n). The needed intervals are given below and we use interval estimates, as presented in Section 4.2, to show that the maximum is still not attained on the mildly larger interval.

Before we begin with the comparison, we introduce a second function that will accompany us till the end:

$$\rho(x) := \frac{2 \operatorname{arccosh}(x)}{x}, \quad x > 0. \quad (8.4)$$

This is a strictly monotonically decreasing function on $[2, \infty)$, which can easily be seen when computing $\rho'(x)$ ¹². Using the expression from (8.1) for A evaluated at η_n in combination with the notation (8.2), we have for all $\eta \leq \eta_n/2$ that

$$A\left(\frac{\eta}{n}, \frac{1}{\eta}\right) < \tanh(\eta_n/4) (1 - f(\eta_n/2)) = \left(1 - \frac{2}{n+1}\right)^{1/2} \left(1 - \frac{\eta_n}{2(n^2-1)^{1/2}}\right).$$

Using (8.3), the desired inequality is

$$\begin{aligned} \left(1 - \frac{2}{n+1}\right)^{1/2} \left(1 - \frac{\eta_n}{2(n^2-1)^{1/2}}\right) &< \frac{(n^2-1)^{1/2}}{n} f\left(\frac{\eta_n}{n}\right) - \frac{\operatorname{arccosh}(n)}{n^2} \\ \iff \left(1 - \frac{2}{n+1}\right)^{1/2} - \frac{(n^2-1)^{1/2}}{n} f\left(\frac{\eta_n}{n}\right) &< \operatorname{arccosh}(n) \left(\left(1 - \frac{2}{n+1}\right)^{1/2} \frac{1}{(n^2-1)^{1/2}} - \frac{1}{n^2} \right). \end{aligned}$$

We claim that the left-hand side is negative and the right-hand side is positive. We start with the negativity of the left-hand side:

$$\left(1 - \frac{2}{n+1}\right)^{1/2} - \frac{(n^2-1)^{1/2}}{n} f\left(\frac{\eta_n}{n}\right) < 0 \quad \iff \quad \frac{n}{n+1} - f\left(\frac{\eta_n}{n}\right) < 0.$$

For all $n \geq 3$, it will suffice to show

¹¹ The full computation is provided in the supplement.

¹² The full computation is provided in the supplement.

$$1 - \frac{3}{4n} < f\left(\frac{\eta_n}{n}\right) \iff \left(1 - \frac{3}{4n}\right) \frac{n}{\eta_n} - \operatorname{csch}\left(\frac{\eta_n}{n}\right) < 0.$$

We claim that the left side of the last inequality is strictly monotonically increasing in n . Recall that $\rho(n) = \eta_n/n$ is strictly monotonically decreasing on $[2, \infty)$. We differentiate the left-hand side of the last inequality with respect to n :

$$\frac{d}{dn} \left[\left(1 - \frac{3}{4n}\right) \frac{n}{\eta_n} - \operatorname{csch}\left(\frac{\eta_n}{n}\right) \right] = \frac{3}{4n\eta_n} + \frac{\eta_n - 2n(n^2 - 1)^{-1/2}}{\eta_n^2} \left(\frac{\eta_n^2}{n^2} \operatorname{csch}'\left(\frac{\eta_n}{n}\right) + 1 - \frac{3}{4n} \right).$$

The only part whose positivity has not been proven stands in the brackets. Keeping in mind that $\eta_n/n = \rho(n) \leq \rho(3) = \eta_3/3 \leq 1.2$, we obtain

$$\frac{\eta_n^2}{n^2} \operatorname{csch}'\left(\frac{\eta_n}{n}\right) + 1 - \frac{3}{4n} = 1 - \frac{3}{4n} - \underbrace{\rho(n)^2 \operatorname{csch}(\rho(n))}_{\nearrow} \underbrace{(\rho(n) \coth(\rho(n)) - 1)}_{\searrow}.$$

The atomic parts of the expression are all positive, so in total, the expression is strictly monotonically increasing. Evaluating at $n = 3$ shows that the expression is always positive, therefore $\left(1 - \frac{3}{4n}\right) \frac{n}{\eta_n} - \operatorname{csch}\left(\frac{\eta_n}{n}\right)$ is strictly monotonically increasing in n . Due to the fact that $1/t - \operatorname{csch}(t) \rightarrow 0$, for $t \rightarrow 0$, we have

$$\lim_{n \rightarrow \infty} \left[\left(1 - \frac{3}{4n}\right) \frac{n}{\eta_n} - \operatorname{csch}\left(\frac{\eta_n}{n}\right) \right] = \lim_{n \rightarrow \infty} \left[\frac{n}{\eta_n} - \operatorname{csch}\left(\frac{\eta_n}{n}\right) - \frac{3}{4\eta_n} \right] = 0,$$

implying that the desired expression is negative. We now examine the right-hand side of the initial inequality. As $\operatorname{arccosh}(x) > 0$ for all $x > 1$, we only need to establish

$$\left(1 - \frac{2}{n+1}\right)^{1/2} \frac{1}{(n^2 - 1)^{1/2}} > \frac{1}{n^2} \iff \frac{1}{n+1} - \frac{1}{n^2} > 0, \quad n \geq 3.$$

All in all, we have shown that the maximum of A cannot be attained in the interval $(0, \eta_n/2]$.

Now, as mentioned above, we will need to improve our estimates for $n = 3, \dots, 10$ in order to show the necessary concavity/convexity later on.

- ($4 \leq n \leq 10$): we compare the value of $\tanh(\eta/2)(1 - f(\eta))$ (which majorizes to the lower bound) at $\eta = 3.08$ with that of $A(\eta_n/n, 1/\eta_n)/n$. We see that there is no maximum in $(0, 3.08] \supset (0, \eta_n/2]$.
- ($n = 3$): this case is a bit more subtle. Here, we perform an interval estimate on $(1.7, 3.3) \supset (\eta_3/2, 3.3)$ with sub-intervals of length 0.3 to see that the global maximum does not lie in the interval $(0, 3.3]$.

8.1.2. The right cut-off

Similarly to the left cut-off, we will show that for $3 \leq n \in \mathbb{N}$ the lower bound A does not assume its global maximum on $[2\eta_n, \infty)$. This is again achieved by comparing with the value at η_n and with extra estimates for small n . By using the properties of $f(t)$, defined by Equation (8.2), and the fact that $\tanh(\eta) < 1$, we have

$$A\left(\frac{\eta}{n}, \frac{1}{\eta}\right) < f\left(\frac{2\eta_n}{n}\right) = \frac{2\eta_n}{n} \operatorname{csch}\left(\frac{2\eta_n}{n}\right) = \frac{\eta_n}{n} \operatorname{csch}\left(\frac{\eta_n}{n}\right) \operatorname{sech}\left(\frac{\eta_n}{n}\right) = f\left(\frac{\eta_n}{n}\right) \operatorname{sech}\left(\frac{\eta_n}{n}\right).$$

As already introduced Equation (8.4), we use $\rho(n) = \frac{2 \operatorname{arccosh}(n)}{n}$. This leaves us with showing the following inequality¹³:

¹³ The full computation is provided in the supplement.

$$\begin{aligned} & \frac{(n^2 - 1)^{1/2}}{n} f(\rho(n)) - \frac{\rho(n)}{2n} > f(\rho(n)) \operatorname{sech}(\rho(n)) \\ \iff & (n^2 - 1)^{1/2} \cosh(\rho(n)) - n > \frac{\sinh(2\rho(n))}{4}. \end{aligned}$$

We want to show that

$$(n^2 - 1)^{1/2} \cosh(\rho(n)) - n - \frac{\sinh(2\rho(n))}{4} > 0.$$

We first assume that $n \geq 10$ and bound $\cosh$ from below and $\sinh$ from above. For $\sinh$, we note that $2\rho(n) \leq 2\rho(10) < 1.2$ (as $n \geq 10$), and use $\sinh(x) < x + x^3$ for $x \in [0, 1.2]$:

$$\begin{aligned} & (n^2 - 1)^{1/2} \cosh(\rho(n)) - n - \frac{\sinh(2\rho(n))}{4} > (n^2 - 1)^{1/2} \left(1 + \frac{\rho(n)^2}{2}\right) - n - \frac{2\rho(n) + (2\rho(n))^3}{4} \\ & = (n^2 - 1)^{1/2} - n + (n^2 - 1)^{1/2} \frac{\rho(n)^2}{2} - \rho(n) - 4\rho(n)^3. \end{aligned}$$

We claim that $(n^2 - 1)^{1/2} - n > -\frac{1}{n}$ and that the remaining part dominates $\frac{1}{n}$. The former can be verified with curve discussion¹⁴. We are left to prove

$$\begin{aligned} & \frac{1}{n} < \rho(n) \left((n^2 - 1)^{1/2} \frac{\rho(n)}{2} - 1 - 4\rho(n)^2 \right) \\ & = \frac{2 \operatorname{arccosh}(n)}{n} \left(\frac{(n^2 - 1)^{1/2}}{n} \operatorname{arccosh}(n) - 1 - 4\rho(n)^2 \right) \\ \iff & 1 < \underbrace{2 \operatorname{arccosh}(n)}_{\nearrow, >0} \left(\underbrace{\left(1 - \frac{1}{n^2}\right)^{1/2}}_{\nearrow, >0} \underbrace{\operatorname{arccosh}(n)}_{\nearrow, >0} - 1 - \underbrace{4\rho(n)^2}_{\searrow, >0} \right). \end{aligned}$$

In summary, for $n \geq 10$, the expression in the brackets is strictly monotonically increasing. Evaluating at 10 shows that it is also positive for all $n \geq 10$. Therefore, the right-hand side of the above inequality is strictly monotonically increasing. Finally, the evaluation at $n = 10$ also shows the claimed inequality. For the initial $3 \leq n \leq 9$ we evaluated the initial inequality to verify it.

Again, the intervals where we look for the maximum need to be updated for $n \leq 10$.

- ($4 \leq n \leq 10$): we ensure that there is no global maximum in $[4n/3, 2\eta_n)$. One can check this with a direct interval estimate for $5 \leq n \leq 10$. In the case $n = 4$ we split it in the intervals $[16/3, 5)$ and $[5, 2\eta_4)$.
- ($n = 3$): we divide $[4.8, 7.1) \supseteq [4.8, 2\eta_3)$ in sub-intervals of length 0.1 to show that there is no global maximum in this interval.

Summing up all the cases from the left cut-off and right cut-off, we are left to show that there is a unique (global) maximum in the interval (see also Fig. 10)

$$I_n = \left(\max\left\{3.08, \frac{\eta_n}{2}\right\}, \min\left\{\frac{4n}{3}, 2\eta_n\right\} \right), \quad 4 \leq n \in \mathbb{N} \quad \text{or} \quad I_n = (3.3, 4.8), \quad n = 3. \quad (8.5)$$

8.1.3. The concavity result

We fix $n \geq 3$ and set $\psi_n(\eta) = \psi(\eta) = f(\eta/n) - f(\eta)$, where $f(t) = t \operatorname{csch}(t)$ is still defined by (8.2). The lower frame bound can then be expressed as

$$A\left(\frac{\eta}{n}, \frac{1}{\eta}\right) = \tanh\left(\frac{\eta}{2}\right) \psi(\eta).$$

¹⁴ The full computation is provided in the supplement.

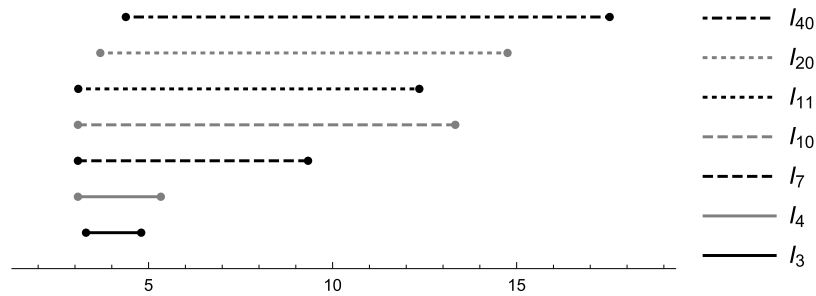


Fig. 10. Illustration of the intervals I_n for $n \in \{3, 4, 7, 10, 11, 20, 40\}$. We have the intervals $I_3 = (3.3, 4.8)$, $I_n = (3.08, 4n/3)$ for $4 \leq n \leq 10$ and $I_n = (\eta_n/2, 2\eta_n)$ for $n \geq 11$.

We want to show that A has a unique maximum in $(\eta_n/2, 2\eta_n)$. As A is strictly positive, we may perform the analysis for $\log(A)$, which has the same critical points. We compute

$$\frac{d^2}{d\eta^2} \left[\log \left(A\left(\frac{\eta}{n}, \frac{1}{\eta}\right) \right) \right] = \frac{\tanh''(\frac{\eta}{2}) \tanh(\frac{\eta}{2}) - \tanh'(\frac{\eta}{2})^2}{4 \tanh(\frac{\eta}{2})^2} + \frac{\psi''(\eta)\psi(\eta) - \psi'(\eta)^2}{\psi(\eta)^2}.$$

Since, $\tanh$ is positive, strictly increasing and strictly concave and ψ is positive (f is strictly monotonically decreasing), it suffices to show that ψ is concave in order to have the whole expression negative. We compute its second derivative, which we may compactly write as

$$\psi''(\eta) = \frac{1}{n^2} f''\left(\frac{\eta}{n}\right) - f''(\eta), \quad \eta \in (\eta_n/2, 2\eta_n).$$

Thus, we turn back to the properties of f , more specifically, $f''(t)$. It can be shown¹⁵ that f'' is strictly negative on $(0, 1.6]$ and strictly positive on $[1.61, \infty)$. Now, recall that

$$\eta \in (\eta_n/2, 2\eta_n) \quad \text{with} \quad \eta_n = 2 \operatorname{arccosh}(n).$$

We observe that $\eta_n/2 \geq \eta_3/2 > 1.76$. As a consequence, we have that $f''(\eta) > 0$ for all $\eta \geq \eta_n/2 > 1.61$. Next, recall that $\rho(x) = 2 \operatorname{arccosh}(x)/x$ is strictly decreasing for $x > 2$. Also, $\eta/n < 2\rho(n)$ and $2\rho(7) < 1.51$. In particular, $f''(\eta/n) < 0$ for all $\eta \leq 2\eta_n$ and $n \geq 7$. Therefore, $\psi''(\eta) < 0$ for all $n \geq 7$. We have shown, for $n \geq 7$, that:

- $A(\eta_n/n, 1/\eta_n) > A(\eta/n, 1/\eta)$, for all $\eta \in (0, \eta_n/2] \cup [2\eta_n, \infty)$.
- $\log(A(\eta/n, 1/\eta))$ is strictly concave in η on $(\eta_n/2, 2\eta_n)$ and therefore has a unique maximum on this interval. As $\log(A(\eta/n, 1/\eta))$ and $A(\eta/n, 1/\eta)$ have maxima at the same positions, it follows that $A(\eta/n, 1/\eta)$ has a unique maximum in $(\eta_n/2, 2\eta_n)$.

The fact that η/n needs to be less than 1.6 makes it necessary to shorten the intervals appropriately for $3 \leq n \leq 6$. This is achieved by using the intervals defined in (8.5). To that, due to the position of the maximum, this does not work for $n = 2$.

All in all, for $n \geq 3$ the lower frame bound has a unique global maximum in $\eta_{A,n} \in (\eta_n/2, 2\eta_n)$.

8.2. The upper bound

We carry on with the same tactic, i.e., we evaluate B given by (8.1) at $\eta_n/2$, η_n and $2\eta_n$ and obtain a convexity result to see that the minimum lies in $(\eta_n/2, 2\eta_n)$.

¹⁵ The full computation is provided in the supplement.

8.2.1. The convexity

This time we start with the convexity of the bound, which actually holds on $\mathbb{R}_+$. It then follows from the asymptotic behavior that we have a unique minimum.

$$\begin{aligned} B\left(\frac{\eta}{n}, \frac{1}{\eta}\right) &> \left(\frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) + 0\right) \rightarrow \infty, \quad \eta \rightarrow \infty, \\ B\left(\frac{\eta}{n}, \frac{1}{\eta}\right) &> \coth\left(\frac{\eta}{2}\right) (1 + 0) \rightarrow \infty, \quad \eta \rightarrow 0^+. \end{aligned}$$

We may already include the case $n = 2$ this time. For a fixed $n \geq 2$, we set

$$h(\eta) = \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) + \eta \operatorname{csch}(\eta). \quad (8.6)$$

The first two derivatives of h are given by¹⁶

$$\begin{aligned} h'(\eta) &= \frac{h(\eta)}{\eta} - \eta \left(\frac{1}{n^2} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 + \operatorname{csch}(\eta) \coth(\eta) \right), \\ h''(\eta) &= -2 \left(\frac{1}{n^2} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 + \operatorname{csch}(\eta) \coth(\eta) \right) + \eta \left(\frac{2}{n^3} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 \coth\left(\frac{\eta}{n}\right) + \operatorname{csch}(\eta) + 2 \operatorname{csch}(\eta)^3 \right). \end{aligned}$$

We apply the Leibniz rule and obtain¹⁷

$$\begin{aligned} \frac{d^2}{d\eta^2} \left[B\left(\frac{\eta}{n}, \frac{1}{\eta}\right) \right] &= \frac{1}{2} \coth\left(\frac{\eta}{2}\right) \operatorname{csch}\left(\frac{\eta}{2}\right)^2 \left(\frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) + \eta \operatorname{csch}(\eta) \right) \\ &\quad - \operatorname{csch}\left(\frac{\eta}{2}\right)^2 \left(\frac{1}{n} \coth\left(\frac{\eta}{n}\right) + \operatorname{csch}(\eta) - \eta \left(\frac{1}{n^2} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 + \operatorname{csch}(\eta) \coth(\eta) \right) \right) \\ &\quad + \coth\left(\frac{\eta}{2}\right) \left(-2 \left(\frac{1}{n^2} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 + \operatorname{csch}(\eta) \coth(\eta) \right) + \right. \\ &\quad \left. \eta \left(\frac{2}{n^3} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 \coth\left(\frac{\eta}{n}\right) + \operatorname{csch}(\eta) + 2 \operatorname{csch}(\eta)^3 \right) \right). \end{aligned}$$

First, we deal with the parts independent of n . For uniformity, all hyperbolic functions will be transformed to an argument $\eta/2$ by using addition formulas. Set

$$\begin{aligned} T(\eta) &= \operatorname{csch}\left(\frac{\eta}{2}\right)^2 \cdot \left(\frac{1}{2} \coth\left(\frac{\eta}{2}\right) \eta \operatorname{csch}(\eta) - \operatorname{csch}(\eta) + \eta \operatorname{csch}(\eta) \coth(\eta) \right) \\ &\quad + \coth\left(\frac{\eta}{2}\right) \left(-2 \operatorname{csch}(\eta) \coth(\eta) + \eta \operatorname{csch}(\eta) + 2 \eta \operatorname{csch}(\eta)^3 \right). \end{aligned}$$

With addition formulas for hyperbolic functions, this can be written as¹⁸

$$T(\eta) = \frac{1}{4} \operatorname{csch}\left(\frac{\eta}{2}\right)^2 \cdot \eta \operatorname{csch}\left(\frac{\eta}{2}\right)^2 \cdot \sum_{k=1}^{\infty} \eta^{2k} \frac{2k-1}{(2k+1)!} > 0.$$

So, the part independent of n is positive. If we can show the same for the remaining terms of the second derivative, then we have proven the strict convexity of the bound on its entire domain. We collect the terms dependent on n and set

$$\begin{aligned} T_n(\eta) &= \frac{1}{2} \cdot \frac{\eta}{n} \coth\left(\frac{\eta}{2}\right) \operatorname{csch}\left(\frac{\eta}{2}\right)^2 \coth\left(\frac{\eta}{n}\right) - \frac{1}{n} \operatorname{csch}\left(\frac{\eta}{2}\right)^2 \coth\left(\frac{\eta}{n}\right) + \frac{\eta}{n^2} \operatorname{csch}\left(\frac{\eta}{2}\right)^2 \operatorname{csch}\left(\frac{\eta}{n}\right)^2 \\ &\quad - \frac{2}{n^2} \coth\left(\frac{\eta}{2}\right) \operatorname{csch}\left(\frac{\eta}{n}\right)^2 + \frac{2\eta}{n^3} \coth\left(\frac{\eta}{2}\right) \operatorname{csch}\left(\frac{\eta}{n}\right)^2 \coth\left(\frac{\eta}{n}\right). \end{aligned}$$

¹⁶ The full computation is provided in the supplement.

¹⁷ The full computation is provided in the supplement.

¹⁸ The full computation is provided in the supplement.

This can be written as

$$T_n(\eta) = \underbrace{\frac{1}{n} \operatorname{csch}\left(\frac{\eta}{2}\right)^2 \coth\left(\frac{\eta}{n}\right)}_{>0} \underbrace{\left(\frac{\eta}{2} \coth\left(\frac{\eta}{2}\right) - 1\right)}_{>0} + \underbrace{\frac{\eta}{n^2} \operatorname{csch}\left(\frac{\eta}{2}\right)^2 \operatorname{csch}\left(\frac{\eta}{n}\right)^2}_{>0} \\ + \underbrace{\frac{2}{n^2} \coth\left(\frac{\eta}{2}\right) \operatorname{csch}\left(\frac{\eta}{n}\right)^2}_{>0} \underbrace{\left(\frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) - 1\right)}_{>0} > 0.$$

All in all, the upper frame bound is strictly convex and tends to infinity when approaching the boundary of its domain, which is $\mathbb{R}_+$. Therefore, it has a unique minimizer $\eta_{B,n}$, as claimed. By comparing the values of B for $\eta \in \{\eta_n/2, \eta_n, 2\eta_n\}$, one sees that $\eta_{B,n}$ has to lie in I_n . Since the points are chosen to allow for the addition formulas, the desired inequalities

$$B\left(\frac{\eta_n}{n}, \frac{1}{\eta_n}\right) < B\left(\frac{\eta_n}{2n}, \frac{2}{\eta_n}\right) \quad \text{and} \quad B\left(\frac{\eta_n}{n}, \frac{1}{\eta_n}\right) < B\left(\frac{2\eta_n}{n}, \frac{1}{2\eta_n}\right)$$

are, after a few equivalence transformations, reduced to easily verifiable inequalities and we omitted them from the manuscript¹⁹. We remark already that for the first $3 \leq n \leq 10$ we will use the intervals I_n defined in (8.5) in order to obtain the needed concavity results for the condition number. It will suffice to evaluate the first derivative on the boundaries of I_n .

8.3. The condition number

Unlike for all other windows, things seem to get only more complicated with the condition number. It is given by

$$\kappa\left(\frac{\eta}{n}, \frac{1}{\eta}\right) = \frac{B\left(\frac{\eta}{n}\right)}{A\left(\frac{\eta}{n}\right)} = \coth\left(\frac{\eta}{2}\right)^2 \frac{\frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) + \eta \operatorname{csch}(\eta)}{\frac{\eta}{n} \operatorname{csch}\left(\frac{\eta}{n}\right) - \eta \operatorname{csch}(\eta)}.$$

We try to use maximally what we already know and to simplify the problem of determining an optimal lattice as much as possible. By the asymptotic behavior of A and B , we know that κ is unbounded towards 0 and ∞ . Showing strict convexity would therefore ensure that we have a unique minimum. This turns out to be a rather difficult problem due to the algebraic nature of the condition number. We simplify the problem by showing the strict log-convexity on I_n , defined by (8.5), and that there is no minimum outside. We compute

$$\log(\kappa(\frac{\eta}{n}, \frac{1}{\eta})) = -\log(A(\frac{\eta}{n}, \frac{1}{\eta})) + \log(B(\frac{\eta}{n}, \frac{1}{\eta})) = -\log(A(\frac{\eta}{n}, \frac{1}{\eta})) - 2\log(\tanh(\frac{\eta}{2})) + \log(h(\eta)),$$

where, for $3 \leq n \in \mathbb{N}$ fixed,

$$h(\eta) = \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) + \eta \operatorname{csch}(\eta),$$

is still defined by Equation (8.6), just as in the proof of the convexity of the upper bound. We know that $\tanh(x)$ is strictly log-concave for $x > 0$ and that the lower bound is strictly log-concave on I_n . If we can show that h is log-convex on I_n , then we have proven that κ is strictly log-convex, hence strictly convex, on I_n . As A does not have a maximum outside I_n and B does not possess a minimum there, the minimum in I_n has to be global and, by convexity, unique. So, we want to show

¹⁹ The full computation is provided in the supplement.

$$\frac{d^2}{d\eta^2} \log(h(\eta)) > 0, \quad \eta \in I_n \iff h(\eta)h''(\eta) - h'(\eta)^2 > 0, \quad \eta \in I_n.$$

We continue by using the explicit formulas for h , h' and h'' and, then further simplifying the expression by grouping terms and using the addition theorems for hyperbolic functions. The computations need very good intuition of how to group the terms efficiently, making this part rather laborious and quite lengthy. However, as each step only needs elementary manipulations of hyperbolic functions, we omit the details at this point²⁰. The result to further analyze is

$$\begin{aligned} h(\eta)h''(\eta) - h'(\eta)^2 &= \frac{\operatorname{csch}(\frac{\eta}{n})^2}{n^2} \underbrace{\left((\frac{\eta}{n})^2 \coth(\frac{\eta}{n})^2 - \cosh(\frac{\eta}{n})^2 + (\frac{\eta}{n})^2 \right)}_{(1)} \\ &\quad + \frac{1}{n} \coth(\frac{\eta}{n}) \operatorname{csch}(\eta) \left(\underbrace{\eta \left(\eta - \frac{2\eta}{n} \coth(\frac{\eta}{n}) \coth(\eta) \right)}_{(2)} + \underbrace{2 \left((\frac{\eta}{n})^2 \coth(\frac{\eta}{n})^2 - 1 \right)}_{>0} \right) \\ &\quad + \frac{2\eta}{n} \operatorname{csch}(\eta) \left(\underbrace{\frac{\eta \operatorname{csch}(\eta)}{2} \left(\cosh(\eta) - \frac{n}{\eta^2} \right)}_{(3)} + \underbrace{\frac{\eta}{2} \coth(\eta) - \frac{\eta}{n} \coth(\frac{\eta}{n})}_{(4)} \right) \\ &\quad + \underbrace{\operatorname{csch}(\eta)^2 \left(\frac{2\eta^2}{n} \coth(\frac{\eta}{n}) \operatorname{csch}(\eta) + \eta^2 \operatorname{csch}(\eta)^2 \right)}_{>0}. \end{aligned}$$

Assuming $n \geq 4$, one can check in a straightforward manner²¹ that all of the expressions (1) – (4) in the above calculation are positive on I_n .

8.3.1. The case $n = 3$

For the interval $I_3 = (3.3, 4.8)$ we make the following three claims:

$$h \geq 1.6 \quad \text{and} \quad |h'| \leq 0.22 \quad \text{and} \quad h'' > 0.$$

They can easily be verified²² with interval estimates on subintervals of I_3 of length 0.1. With these properties of h , we can estimate as follows:

$$\begin{aligned} h(\eta)h''(\eta) - h'(\eta)^2 &> 1.6h''(\eta) - 0.22^2 \\ &= \frac{3.2}{9} \operatorname{csch}(\frac{\eta}{3})^2 \left(\frac{\eta}{3} \coth(\frac{\eta}{3}) - 1 \right) \\ &\quad + 3.2 \operatorname{csch}(\eta) \left(\eta \operatorname{csch}(\eta)^2 + \frac{\eta}{2} - \coth(\eta) \right) - 0.0484 \\ &> \frac{3.2}{9} \underbrace{\frac{\eta}{3} \operatorname{csch}(\frac{\eta}{3}) \cdot \operatorname{csch}(\frac{\eta}{3}) \coth(\frac{\eta}{3})}_{\searrow} - \frac{3.2}{9} \cdot 0.64 \underbrace{\operatorname{csch}(\eta)}_{\searrow} - 0.0484 > 0. \end{aligned}$$

The final estimate can be confirmed by an interval estimate on subintervals of length 0.1. In conclusion, the condition number is strictly log-convex, hence strictly convex, on $(3.3, 4.8)$, where the minimum lies. Therefore, the minimum is unique on in this interval and by what we know about the bounds A and B outside this interval, the minimum has to be global.

²⁰ The full computation is provided in the supplement.

²¹ The full computation is provided in the supplement.

²² The full computation is provided in the supplement.

8.4. The case $n = 2$

This case needs to be treated separately as the necessary adjustments would go beyond simple interval shrinkage. On the other hand, this case has an additional symmetry, as the arguments $\eta/2$ and η/n now coincide. This allows us to manipulate the relevant functions more easily with the addition theorems for hyperbolic functions. Since the expressions to analyze are simplifications of those already considered above, we omit the estimates²³ and summarize the important observations. The lower bound attains its unique maximum $\eta_{A,2}$ in (3.7, 3.74), the upper bound attains its minimum $\eta_{B,2}$ in (3, 3.1) and the condition number is uniquely minimizes between these extremal points: $\eta_{B,2} < \eta_2 < \eta_{A,2}$.

8.5. Three optimal points

Lastly, we want to show that the locations of the optimum for the lower bound, upper bound and condition number differ from each other. As we already know that the optimum is unique in each case, it suffices to show that A and B do not assume their optima at the same point. By recalling (1.1), it is immediate that also κ cannot have its optimum at one of these positions.

8.5.1. The lower bound

We are now in a position to determine the location of the maximum a bit better. We will show that $\eta_{A,n} > \eta_n$. It suffices to show that the first derivative of A is positive at η_n . Since f is strictly monotonically decreasing, we have

$$\begin{aligned} \left. \frac{d}{d\eta} A\left(\frac{\eta}{n}, \frac{1}{\eta}\right) \right|_{\eta=\eta_n} &= \frac{1}{2} \operatorname{sech}\left(\frac{\eta_n}{2}\right)^2 \underbrace{\left(f\left(\frac{\eta_n}{n}\right) - f(\eta_n)\right)}_{>0} + \tanh\left(\frac{\eta_n}{2}\right) \left(\frac{1}{n} f'\left(\frac{\eta_n}{n}\right) - f'(\eta_n)\right) \\ &> \tanh\left(\frac{\eta_n}{2}\right) \left(\frac{1}{n} f'\left(\frac{\eta_n}{n}\right) - f'(\eta_n)\right). \end{aligned}$$

By the definition of f (Equation (8.2)), this is strictly positive if and only if

$$0 < n \operatorname{csch}\left(\frac{\eta_n}{n}\right) \left(1 - \frac{\eta_n}{n} \coth\left(\frac{\eta_n}{n}\right)\right) - n^2 \operatorname{csch}(\eta_n) (1 - \eta_n \coth(\eta_n)).$$

We make several observations. First of all, the addition formulas can be applied wherever the term $\eta_n = 2 \operatorname{arccosh}(n)$ appears. Secondly, the parts involving η_n/n are not as straightforward to analyze, so we transform the parameter to $t = \rho(n) = \eta_n/n$ and use that it is strictly monotonically decreasing in $n \geq 2$. We further manipulate the right-hand side of the above inequality with the mentioned addition formulas and substitutions to an equivalent inequality;

$$0 < nt \left(\frac{\operatorname{csch}(t)}{t} - \operatorname{csch}(t) \coth(t) + \frac{1}{2} \right) - \frac{1}{2} \left(1 + \frac{1}{n^2-1} \right)^{1/2} + \frac{nt}{4(n^2-1)} = (\otimes).$$

As usual, we decompose the expression $(\otimes)$ into parts;

$$\frac{\operatorname{csch}(t)}{t} - \operatorname{csch}(t) \coth(t) + \frac{1}{2} > \operatorname{csch}(t)^2 \left(\frac{\sinh(t)}{t} - 1 \right) = \frac{\operatorname{csch}(t)}{t} - \operatorname{csch}(t)^2.$$

We will now show that the last expression is strictly monotonically decreasing for all $t > 0$;

$$\frac{d}{dt} \left[\frac{\operatorname{csch}(t)}{t} - \operatorname{csch}(t)^2 \right] = -\frac{\operatorname{csch}(t)^4}{t^2} \sum_{k=2}^{\infty} \frac{4t^{2k+2}}{(2k+2)!} \left((k+2)(2^{2k-1} - 2k - 1) + (2k+1) \right).$$

²³ The full computation is provided in the supplement.

It is not hard to convince oneself of the positivity of

$$(k+2)(2^{2k-1} - 2k - 1) + (2k+1), \quad k \geq 2,$$

through the behavior of $2^{x-1} - x$ for $x \geq 4$. Therefore,

$$\frac{\operatorname{csch}(\rho(n))}{\rho(n)} - \operatorname{csch}(\rho(n))^2 \geq \frac{\operatorname{csch}(\rho(2))}{\rho(2)} - \operatorname{csch}(\rho(2))^2 > 0, \quad \forall n \geq 2.$$

So, $\operatorname{csch}(t)/t - \operatorname{csch}(t)^2$ is decreasing for $t > 0$, as claimed. We go back to proving positivity of the expression $(\otimes)$.

$$\begin{aligned} (\otimes) &> nt \operatorname{csch}(t) \left(\frac{1}{t} - \operatorname{csch}(t) \right) - \frac{1}{2} \left(1 + \frac{1}{n^2-1} \right)^{1/2} \\ &= \underbrace{2 \operatorname{arccosh}(n)}_{\nearrow} \underbrace{\left(\frac{\operatorname{csch}(\rho(n))}{\rho(n)} - \operatorname{csch}(\rho(n))^2 \right)}_{>0, \nearrow} \underbrace{- \frac{1}{2} \left(1 + \frac{1}{n^2-1} \right)^{1/2}}_{\nearrow}. \end{aligned}$$

We evaluate at $n = 5$ to verify the positivity of $(\otimes)$ for all $n \geq 5$ by the monotonicity of all terms as indicated above. For the initial $n = 2, 3, 4$, we actually evaluate the derivative of A at η_n and conclude that

$$\left. \frac{d}{d\eta} A\left(\frac{\eta}{n}, \frac{1}{\eta}\right) \right|_{\eta=\eta_n} > 0$$

for all $n \geq 2$. This allows us to conclude $\eta_n < \eta_{A,n}$ for all $n \geq 2$.

8.5.2. The upper bound

We finalize by showing that $\eta_{B,n} \neq \eta_{A,n}$. In fact we show a stronger statement: $\eta_{B,n} < \eta_n$ for $n \geq 4$. Using the calculations for h as defined in (8.6), and $\eta_n = n\rho(n) = 2 \operatorname{arccosh}(n)$, we compute²⁴

$$\begin{aligned} \left. \frac{d}{d\eta} B\left(\frac{\eta}{n}, \frac{1}{n}\right) \right|_{\eta=\eta_n} &= \frac{1}{2(n^2-1)} \left(1 - \frac{\eta_n}{n} \frac{1}{2(n^2-1)^{1/2}} - \frac{\eta_n}{n} \coth\left(\frac{\eta_n}{n}\right) \right) \\ &\quad + \frac{1}{(n^2-1)^{1/2}} \left(\underbrace{\coth\left(\frac{\eta_n}{n}\right) - \frac{\eta_n}{n} \operatorname{csch}\left(\frac{\eta_n}{n}\right)^2 - \frac{\eta_n}{2n}}_{=(I)} + \underbrace{\frac{\eta_n}{n} \frac{1}{4(n^2-1)}}_{\geq 0} \right). \end{aligned} \quad (8.7)$$

If the last expression, and hence the derivative of the upper frame bound at η_n , is positive, then we can conclude that $\eta_{B,n} < \eta_n$. We will first prove this for $n \geq 4$.

We start with two substitutions in (I), namely $\operatorname{csch}^2 = \coth^2 - 1$ and $t = \eta_n/n$;

$$\coth(t) - t \operatorname{csch}(t)^2 - \frac{t}{2} + \frac{t}{4(n^2-1)} > \coth(t) - t \coth(t)^2 + \frac{t}{2} - \frac{t}{14}, \quad t > 0, n \geq 2.$$

We will now study the properties of this new lower bound. We compute²⁵

$$\frac{d^2}{dt^2} \left[\coth(t) - t \coth(t)^2 + \frac{t}{2} - \frac{t}{14} \right] = -\operatorname{csch}(t)^4 \sum_{k=1}^{\infty} 2k \frac{(2t)^{2k+1}}{(2k+1)!} < 0, \quad \forall t > 0.$$

This shows that the expression $\coth(t) - t \coth(t)^2 + \frac{t}{2} - \frac{t}{14}$ is strictly concave on $\mathbb{R}_+$. We observe that, $\coth(1.05) - 1.05 \coth(1.05)^2 + \frac{3 \cdot 1.05}{7} > 0$. Additionally,²⁶

²⁴ The full computation is provided in the supplement.

²⁵ The full computation is provided in the supplement.

²⁶ The full computation is provided in the supplement.

$$\lim_{t \rightarrow 0^+} \coth(t) - t \coth(t)^2 + \frac{t}{2} - \frac{t}{14} = \lim_{t \rightarrow 0^+} \sum_{k=1}^{\infty} \frac{2^{2k+1} (1-k)}{(2k+1)!} t^{2k-1} = 0,$$

due to the convergence of the series. So, on $(0, 1.05) \supseteq (0, \rho(n))$ and for $n \geq 4$, this implies

$$\coth(t) - t \coth(t)^2 + \frac{t}{2} > \frac{t}{14}.$$

We establish²⁷ a lower bound for (8.7) and $n \geq 4$.

$$\left. \frac{d}{d\eta} B\left(\frac{\eta}{n}, \frac{1}{n}\right) \right|_{\eta=\eta_n} > \frac{1}{2(n^2-1)} \left(1 - \underbrace{\frac{\rho(n)}{2(n^2-1)^{1/2}}}_{\searrow} - \underbrace{\rho(n) \coth(\rho(n))}_{\searrow} + \underbrace{\left(1 - \frac{1}{n^2}\right)^{1/2}}_{\nearrow} \underbrace{\frac{2 \operatorname{arccosh}(n)}{7}}_{\nearrow} \right).$$

By evaluating at $n = 4$, we obtain

$$\left. \frac{d}{d\eta} B\left(\frac{\eta}{n}, \frac{1}{n}\right) \right|_{\eta=\eta_n} > \frac{1}{2(n^2-1)} \left(1 - \frac{\rho(4)}{2(4^2-1)^{1/2}} - \rho(4) \coth(\rho(4)) + \left(1 - \frac{1}{4^2}\right)^{1/2} \frac{2 \operatorname{arccosh}(4)}{7} \right) > 0.$$

Hence, $\eta_{B,n} < \eta_n < \eta_{A,n}$ for all $n \geq 4$. For $n = 2, 3$ we have $\eta_{B,n} > \eta_n$. For $n = 2$, we have already established that $\eta_{B,2} < 3.2 < \eta_{A,2}$. As far as $n = 3$ is concerned,

$$\left. \frac{d}{d\eta} A\left(\frac{\eta}{3}, \frac{1}{\eta}\right) \right|_{\eta=3.9} > 0 \quad \text{and} \quad \left. \frac{d}{d\eta} B\left(\frac{\eta}{3}, \frac{1}{\eta}\right) \right|_{\eta=3.9} > 0,$$

finally confirming that

$$\eta_{B,n} < \eta_{A,n}, \quad \forall n \geq 2, n \in \mathbb{N}.$$

Data availability

No data was used for the research described in the article.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.acha.2023.101574>.

References

- [1] R. Balian, Un principe d'incertitude fort en théorie du signal ou en mécanique quantique, C. R. Acad. Sci. 292 (2) (1981) 1357–1361.
- [2] P.S. Bruckman, On the evaluation of certain infinite series by elliptic functions, Fibonacci Q. 15 (3) (1977) 293–310.
- [3] Á. Bényi, K.A. Okoudjou, Modulation Spaces – With Applications to Pseudodifferential Operators and Nonlinear Schrödinger Equations, Applied and Numerical Harmonic Analysis, Birkhäuser, 2020.
- [4] L. Bétermin, M. Faulhuber, S. Steinerberger, A variational principle for Gaussian lattice sums, arXiv preprint, arXiv: 2110.06008, 2021.
- [5] O. Christensen, An Introduction to Frames and Riesz Bases, 2nd edition, Applied and Numerical Harmonic Analysis, Birkhäuser, 2016.
- [6] X.-R. Dai, Q. Sun, The abc-problem for Gabor systems and uniform sampling in shift-invariant spaces, in: R. Balan, J.M. Begué, J.J. Benedetto, W. Czaja, A.K. Okoudjou (Eds.), Excursions in Harmonic Analysis, Volume 3: The February Fourier Talks at the Norbert Wiener Center, Springer, 2015, pp. 177–194.
- [7] X.-R. Dai, Q. Sun, The abc-Problem for Gabor Systems, Memoirs of the American Mathematical Society, vol. 244, American Mathematical Society, 2016.

²⁷ The full computation is provided in the supplement.

- [8] I. Daubechies, Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, 1992.
- [9] I. Daubechies, H.J. Landau, Z. Landau, Gabor time-frequency lattices and the Wexler–Raz identity, *J. Fourier Anal. Appl.* 1 (4) (1995) 437–478.
- [10] M. Faulhuber, Minimal frame operator norms via minimal theta functions, *J. Fourier Anal. Appl.* 24 (2) (2018) 545–559.
- [11] M. Faulhuber, On the parity under metaplectic operators and an extension of a result of Lyubarskii and Nes, *Results Math.* 75 (8) (2020).
- [12] M. Faulhuber, Extremal determinants of Laplace–Beltrami operators for rectangular tori, *Math. Z.* (2021).
- [13] M. Faulhuber, An application of hypergeometric functions to heat kernels on rectangular tori and a “Weltkonstante” - or - how Ramanujan split temperatures, *Ramanujan J.* (2021).
- [14] M. Faulhuber, S. Steinerberger, Optimal Gabor frame bounds for separable lattices and estimates for Jacobi theta functions, *J. Math. Anal. Appl.* 445 (1) (2017) 407–422.
- [15] H.G. Feichtinger, On a new Segal algebra, *Monatshefte Math.* 92 (4) (1981) 269–289.
- [16] H.G. Feichtinger, A new family of functional spaces on the Euclidean n -space, in: *Proceedings of Theory of Approximation of Functions*, 1983.
- [17] H.G. Feichtinger, A.J.E.M. Janssen, Validity of WH-frame bound conditions depends on lattice parameters, *Appl. Comput. Harmon. Anal.* 8 (1) (2000) 104–112.
- [18] H.G. Feichtinger, W. Kozek, Quantization of TF lattice-invariant operators on elementary LCA groups, in: H.G. Feichtinger, T. Strohmer (Eds.), *Gabor Analysis and Algorithms*, Springer, 1998, pp. 233–266.
- [19] G.B. Folland, *Harmonic Analysis in Phase Space*, Annals of Mathematics Studies, vol. 122, Princeton University Press, 1989.
- [20] M.A. de Gosson, *Symplectic Methods in Harmonic Analysis and in Mathematical Physics*, Pseudo-Differential Operators. Theory and Applications, vol. 7, Birkhäuser/Springer Basel AG, Basel, 2011.
- [21] I.S. Gradshteyn, I.M. Ryzhik, *Table of Integrals, Series and Products*, 8th edition, Academic Press, Oxford, 2014.
- [22] K. Gröchenig, *Foundations of Time-Frequency Analysis*, Applied and Numerical Harmonic Analysis, Birkhäuser, Boston, MA, 2001.
- [23] K. Gröchenig, The mystery of Gabor frames, *J. Fourier Anal. Appl.* 20 (4) (2014) 865–895.
- [24] K. Gröchenig, P. Jaming, E. Malinnikova, Zeros of the Wigner distribution and the short-time Fourier transform, *Rev. Mat. Complut.* 33 (2020) 723–744.
- [25] K. Gröchenig, S. Koppensteiner, Gabor frames: characterizations and coarse structure, in: A. Aldroubi, C. Cabrelli, S. Jaffard, U. Molter (Eds.), *New Trends in Applied Harmonic Analysis*, in: *Applied and Numerical Harmonic Analysis*, vol. 2, Springer, 2019, pp. 93–120.
- [26] K. Gröchenig, J.-L. Romero, J. Stöckler, Sampling theorems for shift-invariant spaces, Gabor frames, and totally positive functions, *Invent. Math.* 211 (3) (2018) 1119–1148.
- [27] K. Gröchenig, J. Stöckler, Gabor frames and totally positive functions, *Duke Math. J.* 162 (6) (2013) 1003–1031.
- [28] C. Heil, History and evolution of the density theorem for Gabor frames, *J. Fourier Anal. Appl.* 13 (2) (2007) 113–166.
- [29] M.S. Jakobsen, On a (no longer) new Segal algebra: a review of the Feichtinger algebra, *J. Fourier Anal. Appl.* 24 (6) (2018) 1579–1660.
- [30] M.S. Jakobsen, J. Lemvig, Density and duality theorems for regular Gabor frames, *J. Funct. Anal.* 270 (1) (2016) 229–263.
- [31] A.J.E.M. Janssen, Gabor representation of generalized functions, *J. Math. Anal. Appl.* 83 (2) (1981) 377–394.
- [32] A.J.E.M. Janssen, The Zak transform: a signal transform for sampled time-continuous signals, *Philips J. Res.* 43 (1988) 23–69.
- [33] A.J.E.M. Janssen, Duality and biorthogonality for Weyl–Heisenberg frames, *J. Fourier Anal. Appl.* 1 (4) (1995) 403–436.
- [34] A.J.E.M. Janssen, Some Weyl–Heisenberg frame bound calculations, *Indag. Math.* 7 (2) (1996) 165–183.
- [35] A.J.E.M. Janssen, On generating tight Gabor frames at critical density, *J. Fourier Anal. Appl.* 9 (2) (2003) 175–214.
- [36] A.J.E.M. Janssen, Zak transforms with few zeros and the tie, in: Hans G. Feichtinger, Thomas Strohmer (Eds.), *Advances in Gabor Analysis*, Birkhäuser Boston, Boston, MA, 2003, pp. 31–70.
- [37] A.J.E.M. Janssen, T. Strohmer, Hyperbolic secants yield Gabor frames, *Appl. Comput. Harmon. Anal.* 12 (2) (2002) 259–267.
- [38] F. Low, Complete sets of wave packets, in: *A Passion for Physics*, World Scientific, 1985, pp. 17–22.
- [39] Y. Lyubarskii, Frames in the Bargmann space of entire functions, in: *Entire and Subharmonic Functions*, American Mathematical Society, Providence, RI, 1992, pp. 167–180.
- [40] Y. Lyubarskii, K. Seip, Convergence and summability of Gabor expansions at the Nyquist density, *J. Fourier Anal. Appl.* 5 (2–3) (1999) 127–157.
- [41] H.L. Montgomery, Minimal theta functions, *Glasg. Math. J.* 30 (1) (1988) 75–85.
- [42] A. Ron, Z. Shen, Weyl–Heisenberg frames and Riesz bases in $L_2(\mathbb{R}^d)$, *Duke Math. J.* 89 (2) (1997) 237–282.
- [43] W. Rudin, *Principles of Mathematical Analysis*, third edition, International Series in Pure and Applied Mathematics, McGraw-Hill Inc., 1976.
- [44] K. Seip, Density theorems for sampling and interpolation in the Bargmann–Fock space I, *J. Reine Angew. Math.* 429 (1992) 91–106.
- [45] K. Seip, R. Wallstén, Density theorems for sampling and interpolation in the Bargmann–Fock space II, *J. Reine Angew. Math.* 429 (1992) 107–114.
- [46] T. Strohmer, S. Beaver, Optimal OFDM design for time–frequency dispersive channels, *IEEE Trans. Commun.* 51 (7) (2003) 1111–1122.
- [47] J. Wexler, S. Raz, Discrete Gabor expansions, *Signal Process.* 21 (3) (1990) 207–220.
- [48] Research Wolfram, Inc., *Mathematica*, Version 12–13, Champaign, IL, 2019–2021.

SUPPLEMENT TO: GABOR FRAME BOUND OPTIMIZATIONS

MARKUS FAULHUBER AND IRINA SHAFKULOVSKA

In this supplement, we provide technical details which were omitted in the manuscript for the sake of readability. This supplement, however, should not be considered on its own and it should be used as follows: each calculation has a number. This number appears in the original manuscript as a superscript in the mentioned section and at the spot where we omitted the respective calculation/explanation. The context is given there.

2. The hyperbolic functions are the analogues of the ordinary trigonometric functions and fulfill similar identities. For the reader's convenience, we not only recall them, but also provide some easy bounds which we used in the manuscript. Although we only observe them with strictly positive arguments, their maximal domain is the complex plane. We write for $x > 0$

$$\begin{aligned}\cosh(x) &= \frac{e^x + e^{-x}}{2} = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!}, & \operatorname{sech}(x) &= \frac{1}{\cosh(x)}, \\ \sinh(x) &= \frac{e^x - e^{-x}}{2} = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!}, & \operatorname{csch}(x) &= \frac{1}{\sinh(x)} \\ \tanh(x) &= \frac{\sinh(x)}{\cosh(x)} = 1 - 2(e^{2x} + 1)^{-1} & \coth(x) &= \frac{\cosh(x)}{\sinh(x)} = 1 + 2(e^{2x} - 1)^{-1}.\end{aligned}$$

A wealth of identities analogous to the trigonometric functions hold. We will frequently use

$$\begin{aligned}\cosh(x)^2 - \sinh(x)^2 &= 1, & \cosh(2x) &= 2\cosh(x)^2 - 1, \\ \tanh(x)^2 + \operatorname{sech}(x)^2 &= 1, & \sinh(2x) &= 2\sinh(x)\cosh(x) \\ \coth(x)^2 - \operatorname{csch}(x)^2 &= 1.\end{aligned}$$

The series expansions of $\exp$, $\cosh$, $\sinh$ consist of strictly positive terms and converge quite fast. The monotonicity properties of hyperbolic functions can be read easily from their Taylor expansions or the relation to the exponential function. Truncation of the Taylor series also provide good upper and lower bounds. It is not difficult (for instance, from the Taylor expansions) to see that $\cosh$, $\sinh$ and $\tanh$ are strictly monotonically increasing, whereas the others are strictly monotonically decreasing.

For further bounds, estimates of the following type are usually good enough:

$$\frac{e^x}{2} < \cosh(x) < \frac{e^x + 1}{2} < e^x, \quad x > 0.$$

NUHAG, FACULTY OF MATHEMATICS, UNIVERSITY OF VIENNA

OSKAR-MORGENSTERN-PLATZ 1, 1090 VIENNA, AUSTRIA

E-mail addresses: `markus.faulhuber@univie.ac.at`, `irina.shafkulovska@univie.ac.at`.

1991 *Mathematics Subject Classification.* primary: 42C15; secondary: 26A06, 33B10.

Key words and phrases. exponential functions, frame bounds, Gabor frame, hyperbolic functions, lattice.

In some of our calculations we will use the symbols $\nearrow$ and $\searrow$ to denote that an expression is increasing or decreasing, respectively.

3. As $h_A(t) = t f'_A(t)$, we start with computing derivatives of f_A :

$$\begin{aligned} f'_A(t) &= \sum_{k=0}^{\infty} \operatorname{sech}(\alpha_k t)^2 (1 - 2\alpha_k t \tanh(\alpha_k t)), \\ f''_A(t) &= -4 \sum_{k=0}^{\infty} \alpha_k \operatorname{sech}(\alpha_k t)^2 \tanh(\alpha_k t) \\ &\quad + 4t \sum_{k=0}^{\infty} \alpha_k^2 \operatorname{sech}(\alpha_k t)^2 \tanh(\alpha_k t)^2 - 2t \sum_{k=0}^{\infty} \alpha_k^2 \operatorname{sech}(\alpha_k t)^4, \\ f'''_A(t) &= 6 \sum_{k=0}^{\infty} \alpha_k^2 \operatorname{sech}(\alpha_k t)^2 (2 \tanh^2(\alpha_k t) - \operatorname{sech}(\alpha_k t)^2) \\ &\quad + 8t \sum_{k=0}^{\infty} \alpha_k^3 \operatorname{sech}(\alpha_k t)^2 \tanh(\alpha_k t) (2 \operatorname{sech}(\alpha_k t)^2 - \tanh(\alpha_k t)^2). \end{aligned}$$

4. For $x > 2$, the polynomial $p(x)$ is clearly negative and decreasing. However, for t close to 1 we have that $x = \alpha_0 t \tanh(\alpha_0 t) \in (1/2, 2)$, where p is positive. More precisely, for all $t \in (1, 1.04)$ it holds that

$$\alpha_0 t \tanh(\alpha_0 t) \in \left(\frac{\pi}{2} \tanh\left(\frac{\pi}{2}\right), \frac{1.04\pi}{2} \tanh\left(\frac{1.04\pi}{2}\right) \right) \subseteq (1.44, 1.52).$$

The polynomial p is strictly positive and strictly monotonically decreasing on $(1.44, 1.52)$. All other terms land in the negative part of p as

$$\alpha_k t \tanh(\alpha_k t) \geq \alpha_1 \tanh(\alpha_1) > 4 > 2.$$

We recall that $0 < \tanh < 1$, as well as that p is strictly monotonically decreasing and strictly negative on $(2, \infty)$. With a standard curve discussion, we see that $p(x) > -3x^2$ for all $x \geq 1$.

5.

$$\begin{aligned} \frac{1}{2} h''_a(t) &\geq \sum_{k=0}^{\infty} 6\alpha_k \operatorname{sech}^2(\alpha_k t) - 16.14 \sum_{k=1}^{\infty} \operatorname{sech}(1.04 \alpha_k) \\ &\quad + \alpha_0 t \tanh(\alpha_0 t) \operatorname{sech}(\alpha_0 t)^2 p(\alpha_0 t \tanh(\alpha_0 t)) \\ &\geq 6 \sum_{k=0}^{\infty} e^{-2\pi k t} e^{-\pi t} - 16.14 \sum_{k=1}^{\infty} 2 e^{-1.04 \pi k} e^{-0.52 \pi} \\ &\quad + 1.44 \operatorname{sech}\left(\frac{1.04\pi}{2}\right)^2 p(1.52) \\ &> 3 \frac{1}{\sinh(\pi t)} - 16.14 \frac{e^{-1.04 \pi}}{\sinh(0.52 \pi)} + 1.44 \cdot 0.14 \cdot 0.9792 \\ &> 3 \cdot 0.07 - 16.14 \cdot \frac{0.04}{2.46} + 0.19 > 0.13 > 0. \end{aligned}$$

6. We compute the first and second derivative of f_B :

$$\begin{aligned} f'_B(t) &= \sum_{k=0}^{\infty} \operatorname{csch}(\alpha_k t)^2 (1 - 2\alpha_k t \coth(\alpha_k t)), \\ f''_B(t) &= -4 \sum_{k=0}^{\infty} \alpha_k \operatorname{csch}(\alpha_k t)^2 \coth(\alpha_k t) + 4t \sum_{k=0}^{\infty} \alpha_k^2 \operatorname{csch}(\alpha_k t)^2 \coth(\alpha_k t)^2 \\ &\quad + 2t \sum_{k=0}^{\infty} \alpha_k^2 \operatorname{csch}(\alpha_k t)^4. \end{aligned}$$

Using the identity $\operatorname{csch}(x)^2 = \coth(x)^2 - 1$ and $h_B(t) = t f_B(t)$, we compute h'_B .

7. We are presenting lower bounds for $\psi(x) = 1 - 2x^2 + 6x^2 \coth(x)^2 - 6x \coth(x)$,

$$\psi(x) = q_1(x) + 6 q_2(x \coth(x)), \quad q_1(x) = 1 - 2x^2 \quad \text{and} \quad q_2(x) = x(x - 1)$$

with interval estimates on $(0, 1.6)$.

a): $x \in (0, \sqrt{0.5})$.

$$\psi(x) > q_1(\sqrt{0.5}) + 0 = 0.$$

b): $x \in (0.7, 1.01)$.

The choice of the lower interval bound is due to $0.7 = \sqrt{0.49} < \sqrt{0.5}$. To that, the upper bound is determined by $0.7 \cdot \coth(0.7) > 1.158$. We have

$$\psi(x) > q_1(1.01) + 6 q_2(1.158) = 0.057584 > 0.$$

c): $x \in (1, 1.3)$.

$$\psi(x) > q_1(1.3) + 6 q_2(1.31) = 0.0566 > 0.$$

d): $x \in (1.29, 1.6)$.

$$\psi(x) > q_1(1.6) + 6 q_2(1.5) = 0.38 > 0.$$

This proves that ψ , hence h'_B is strictly positive on $(0, \infty)$.

8. The first derivative of the upper frame bound is given by

$$\begin{aligned} \frac{d}{d\eta} A\left(\frac{\eta}{n}, \frac{1}{\eta}\right) &= \tanh\left(\frac{\eta}{2}\right) e^{-\frac{\eta}{n}} \operatorname{csch}\left(\frac{\eta}{n}\right) + \eta \frac{1}{2} \operatorname{sech}^2\left(\frac{\eta}{2}\right) e^{-\frac{\eta}{n}} \operatorname{csch}\left(\frac{\eta}{n}\right) \\ &\quad - \frac{\eta}{n} \tanh\left(\frac{\eta}{2}\right) e^{-\frac{\eta}{n}} \operatorname{csch}\left(\frac{\eta}{n}\right) - \frac{\eta}{n} \tanh\left(\frac{\eta}{2}\right) e^{-\frac{\eta}{n}} \operatorname{csch}\left(\frac{\eta}{n}\right) \coth\left(\frac{\eta}{n}\right) \\ &= e^{-\frac{\eta}{n}} \operatorname{csch}\left(\frac{\eta}{n}\right) \tanh\left(\frac{\eta}{2}\right) \left(1 + \eta \frac{\operatorname{sech}^2(\frac{\eta}{2})}{2 \tanh(\frac{\eta}{2})} - \frac{\eta}{n} - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right)\right) \\ &= e^{-\frac{\eta}{n}} \operatorname{csch}\left(\frac{\eta}{n}\right) \tanh\left(\frac{\eta}{2}\right) \left(1 + \eta \operatorname{csch}(\eta) - \frac{\eta}{n} - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right)\right), \end{aligned}$$

9. The first derivative of the upper frame bound is given by

$$\begin{aligned}
\frac{d}{d\eta} B\left(\frac{\eta}{n}, \frac{1}{\eta}\right) &= \coth\left(\frac{\eta}{2}\right) e^{\frac{\eta}{n}} \operatorname{csch}\left(\frac{\eta}{n}\right) - \eta \frac{1}{2} \operatorname{csch}\left(\frac{\eta}{2}\right)^2 e^{\frac{\eta}{n}} \operatorname{csch}\left(\frac{\eta}{n}\right) \\
&\quad + \frac{\eta}{n} \coth\left(\frac{\eta}{2}\right) e^{\frac{\eta}{n}} \operatorname{csch}\left(\frac{\eta}{n}\right) - \frac{\eta}{n} \coth\left(\frac{\eta}{2}\right) e^{\frac{\eta}{n}} \operatorname{csch}\left(\frac{\eta}{n}\right) \coth\left(\frac{\eta}{n}\right) \\
&= e^{\frac{\eta}{n}} \operatorname{csch}\left(\frac{\eta}{n}\right) \coth\left(\frac{\eta}{2}\right) \left(1 - \eta \frac{\operatorname{csch}\left(\frac{\eta}{2}\right)^2}{2 \coth\left(\frac{\eta}{2}\right)} + \frac{\eta}{n} - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right)\right) \\
&= e^{\frac{\eta}{n}} \operatorname{csch}\left(\frac{\eta}{n}\right) \coth\left(\frac{\eta}{2}\right) \left(1 + \frac{\eta}{n} - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) - \eta \operatorname{csch}(\eta)\right)
\end{aligned}$$

10. We want to show that $x \mapsto 1 + x - x \coth(x)$ is a strictly concave, strictly monotonically increasing function. It holds

$$\begin{aligned}
\frac{d}{dx} [1 + x - x \coth(x)] &= 1 - \coth(x) + x \operatorname{csch}(x)^2, \\
\frac{d^2}{dx^2} [1 + x - x \coth(x)] &= \operatorname{csch}(x)^2 + \operatorname{csch}(x)^2 - 2x \operatorname{csch}(x)^2 \coth(x) \\
&= 2 \operatorname{csch}(x)^2 (1 - x \coth(x)) < 0,
\end{aligned}$$

due to the following claim.

Claim 1. *The function $x \coth(x)$ is unbounded, strictly monotonically increasing on $(0, \infty)$ and can be continuously extended at 0 by setting the value to 1.*

Proof. First of all, we observe

$$\lim_{x \rightarrow 0^+} x \coth(x) = \lim_{x \rightarrow 0^+} x + \lim_{x \rightarrow 0^+} \left(\frac{e^{2x} - e^0}{2x - 0} \right)^{-1} = 0 + \frac{1}{\exp'(0)} = 1.$$

For the monotonicity, we compute the first derivative. For all $x > 0$ holds

$$\begin{aligned}
\frac{d}{dx} [x \coth(x)] &= \coth(x) - x \operatorname{csch}(x)^2 = \operatorname{csch}(x)^2 (\sinh(x) \cosh(x) - x) \\
&= \operatorname{csch}(x)^2 \left(\frac{1}{2} \sinh(2x) - x \right) = \operatorname{csch}(x)^2 \frac{1}{2} \sum_{k=1}^{\infty} \frac{(2x)^{2k+1}}{(2k+1)!} > 0
\end{aligned}$$

□

The claim above shows that the first derivative of $1 + x - x \coth(x)$ is strictly decreasing. It is strictly positive, because

$$\lim_{x \rightarrow \infty} \frac{d}{dx} [1 + x - x \coth(x)] = 1 - 1 + 0 = 0.$$

11. We set

$$f(x) = t \operatorname{csch}(x) = \frac{x}{\sinh(x)}, \quad \forall t > 0.$$

One can easily verify the asymptotics by the series expansion of $\sinh$:

$$\frac{\sinh(x)}{x} = \frac{\sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!}}{x} = \sum_{k=0}^{\infty} \underbrace{\frac{x^{2k}}{(2k+1)!}}_{\nearrow}.$$

This shows that f is strictly monotonically decreasing, tends to 1 as x tends to 0 and tends to 0 as x increases.

12. We will need the following property only for natural numbers, but the result holds for positive real numbers. For $x > 0$, we set

$$\rho(x) = \frac{2 \operatorname{arccosh}(x)}{x}.$$

It holds

$$\rho'(x) = -\frac{2 \operatorname{arccosh}(x)}{x^2} + \frac{2}{x} \cdot \frac{1}{(x^2-1)^{1/2}} = \frac{1}{x} \left(\frac{2}{(x^2-1)^{1/2}} - \rho(x) \right).$$

This derivative is negative if and only if

$$\left(1 + \frac{1}{x^2-1}\right) < \operatorname{arccosh}(x)^2.$$

This can be easily verified for $x = 2$, and holds for all other $x \in (2, \infty)$:

$$\left(1 + \frac{1}{x^2-1}\right) < \left(1 + \frac{1}{2^2-1}\right) < \operatorname{arccosh}(2)^2 < \operatorname{arccosh}(x)^2.$$

13.

$$\begin{aligned} & \frac{(n^2-1)^{1/2}}{n} f(\rho(n)) - \frac{\rho(n)}{2n} > f(\rho(n)) \operatorname{sech}(\rho(n)) \\ \iff & \left(\frac{(n^2-1)^{1/2}}{n} - \operatorname{sech}(\rho(n)) \right) \rho(n) \operatorname{csch}(\rho(n)) > \frac{\rho(n)}{2n} \\ \iff & \left((n^2-1)^{1/2} - n \operatorname{sech}(\rho(n)) \right) \operatorname{csch}(\rho(n)) > \frac{1}{2} \\ \iff & (n^2-1)^{1/2} \cosh(\rho(n)) - n > \frac{2 \sinh(\rho(n)) \cosh(\rho(n))}{4} \\ \iff & (n^2-1)^{1/2} \cosh(\rho(n)) - n > \frac{\sinh(2\rho(n))}{4}. \end{aligned}$$

14. We compute

$$\frac{d}{dn} \left[\underbrace{(n^2-1)^{1/2} - n + \frac{1}{n}}_{= (*)} \right] = \left(1 + \frac{1}{(n^2-1)}\right)^{1/2} - 1 - \frac{1}{n^2} < 0,$$

where the inequality holds if and only if

$$1 + \frac{1}{(n^2-1)} < \left(1 + \frac{1}{n^2}\right)^2 = 1 + \frac{2}{n^2} + \frac{1}{n^4} \iff n^2 < (2 + \frac{1}{n^2})(n^2-1).$$

Since $n^2 < 2(n^2-1) < (2 + 1/n^2)(n^2-1)$ holds for all $2 \leq n \in \mathbb{N}$, $(*)$ is strictly monotonically decreasing. To that, considering it tends to 0 as n tends to infinity, it is strictly positive.

15. We compute

$$(1) \quad f''(t) = -\frac{2t \operatorname{csch}(t)}{\coth(t)^2} \left(\frac{1}{2} \tanh(t)^2 + \frac{\tanh(t)}{t} - 1 \right).$$

Clearly, $-(2t \operatorname{csch}(t))/\coth(t)^2 < 0$ on $\mathbb{R}_+$. So, in order to determine where $f''(t) < 0$, we only need to know where

$$\frac{1}{2} \tanh(t)^2 + \frac{\tanh(t)}{t} - 1 > 0.$$

We compute the derivative of this expression:

$$\begin{aligned} \frac{d}{dt} \left(\frac{1}{2} \tanh(t)^2 + \frac{\tanh(t)}{t} - 1 \right) &= \tanh(t) \operatorname{sech}(t)^2 + \frac{\operatorname{sech}(t)^2}{t} - \frac{\tanh(t)}{t^2} \\ &= \tanh(t) \left(\operatorname{sech}(t)^2 + \frac{2}{t \sinh(2t)} - \frac{1}{t^2} \right). \end{aligned}$$

An issue which appears is that we have a difference of two unbounded functions (close to $t = 0$). Yet, it is not as uncontrolled as it may seem at a first glance:

$$\begin{aligned} \frac{2}{t \sinh(2t)} - \frac{1}{t^2} &= \frac{1}{t^2 \sum_{k=0}^{\infty} \frac{(2t)^{2k}}{(2k+1)!}} - \frac{1}{t^2} = \frac{t^{-2} \left(1 - \frac{(2t)^{2 \cdot 0}}{1!} \right) - 4(2t)^{-2} \sum_{k=1}^{\infty} \frac{(2t)^{2k}}{(2k+1)!}}{\frac{(2t)^{2 \cdot 0}}{1!} + \sum_{k=1}^{\infty} \frac{(2t)^{2k}}{(2k+1)!}} \\ &= - \frac{4 \sum_{k=0}^{\infty} \frac{(2t)^{2k}}{(2k+3)!}}{1 + \sum_{k=1}^{\infty} \frac{(2t)^{2k}}{(2k+1)!}} = - \frac{\frac{4}{3!} + 4 \sum_{k=1}^{\infty} \frac{(2t)^{2k}}{(2k+1)!}}{1 + \sum_{k=1}^{\infty} \frac{(2t)^{2k}}{(2k+1)!}}. \end{aligned}$$

We now take the limit for $t \rightarrow 0^+$:

$$\lim_{t \rightarrow 0^+} \frac{2}{t \sinh(2t)} - \frac{1}{t^2} = \lim_{t \rightarrow 0^+} - \frac{\frac{4}{3!} + 4 \sum_{k=1}^{\infty} \frac{(2t)^{2k}}{(2k+1)!}}{1 + \sum_{k=1}^{\infty} \frac{(2t)^{2k}}{(2k+1)!}} = -\frac{2}{3}.$$

From the above calculation, we also get

$$\operatorname{sech}(t)^2 - 4 \sum_{k=0}^{\infty} \frac{(2t)^{2k}}{(2k+3)!} < \operatorname{sech}(t)^2 + \frac{2}{t \sinh(t)} - \frac{1}{t^2}.$$

The minor $\operatorname{sech}(t)^2 - 4 \sum_{k=0}^{\infty} \frac{(2t)^{2k}}{(2k+3)!}$ is strictly monotonically decreasing on $(0, \infty)$ and positive at $t = 0.5$. On the other hand, due to the monotonicity of f , we have for all $t > 1.1$

$$\frac{2}{t \sinh(2t)} - \frac{1}{t^2} < -\frac{1}{2t^2} \iff \frac{2}{t \sinh(2t)} < \frac{1}{2t^2} \iff f(2t) < \frac{1}{2}.$$

To that, for all $t > 1.2$, we have

$$\operatorname{sech}(t)^2 + \frac{2}{t \sinh(2t)} - \frac{1}{t^2} < \operatorname{sech}(t)^2 - \frac{1}{4t^2} < 0,$$

or equivalently, $2 < \cosh(t)/t$. With a standard curve discussion, one sees that $\cosh(t)/t$ is strictly increasing on $(1.2, \infty)$. In summary, the expression in brackets in (2), i.e.,

$$(2) \quad \underbrace{\frac{1}{2} \tanh(t)^2}_{\nearrow} + \underbrace{\tanh(t)}_{\nearrow} \underbrace{\frac{1}{t}}_{\searrow} - 1$$

is strictly monotonically increasing on the interval $(0, 0.5)$ and strictly monotonically decreasing on the interval $(1.2, \infty)$. By l'Hôpital's rule, the value at $t = 0$, is given by

$$\lim_{t \rightarrow 0^+} \left(\frac{1}{2} \tanh(t)^2 + \frac{\tanh(t)}{t} - 1 \right) = 0 + \left(\lim_{t \rightarrow 0^+} \frac{\operatorname{sech}(t)^2}{1} \right) - 1 = \frac{1}{1} - 1 = 0.$$

This shows that the expression (2) is strictly positive on $(0, 0.5]$. We claim that this is preserved on the interval $(0.5, 1.2)$. For a useful interval estimate (with the indicated monotonicities

in ()), one should split it (0.5,1.2) in sub-intervals of length 0.01. Finally, we check that the expression is still strictly positive at 1.6 and less than 0 at 1.61 by evaluating it at these points. Strict negativity on (1.61, ∞) follows from the monotonicity.

16. The first two derivatives of $h(\eta) = \frac{\eta}{n} \coth(\frac{\eta}{n}) + \eta \operatorname{csch}(\eta)$ are given by

$$\begin{aligned} h'(\eta) &= \frac{1}{n} \coth(\frac{\eta}{n}) + \operatorname{csch}(\eta) - \eta \left(\frac{1}{n^2} \operatorname{csch}(\frac{\eta}{n})^2 + \operatorname{csch}(\eta) \coth(\eta) \right) \\ &= \frac{h(\eta)}{\eta} - \eta \left(\frac{1}{n^2} \operatorname{csch}(\frac{\eta}{n})^2 + \operatorname{csch}(\eta) \coth(\eta) \right), \end{aligned}$$

$$\begin{aligned} h''(\eta) &= -2 \left(\frac{1}{n^2} \operatorname{csch}(\frac{\eta}{n})^2 + \operatorname{csch}(\eta) \coth(\eta) \right) \\ &\quad + \eta \left(\frac{2}{n^3} \operatorname{csch}(\frac{\eta}{n})^2 \coth(\frac{\eta}{n}) + \operatorname{csch}(\eta) \coth(\eta)^2 + \operatorname{csch}(\eta)^3 \right) \\ &= -2 \left(\frac{1}{n^2} \operatorname{csch}(\frac{\eta}{n})^2 + \operatorname{csch}(\eta) \coth(\eta) \right) + \eta \left(\frac{2}{n^3} \operatorname{csch}(\frac{\eta}{n})^2 \coth(\frac{\eta}{n}) + \operatorname{csch}(\eta) + 2 \operatorname{csch}(\eta)^3 \right). \end{aligned}$$

17.

$$\begin{aligned} \frac{d^2}{d\eta^2} \left[B_1\left(\frac{\eta}{n}, \frac{1}{n}\right) \right] &= \frac{2}{4} \coth(\frac{\eta}{2}) \operatorname{csch}(\frac{\eta}{2})^2 h(\eta) - \frac{2}{2} \operatorname{csch}(\frac{\eta}{2})^2 h'(\eta) + \coth(\frac{\eta}{2}) h''(\eta) \\ &= \frac{1}{2} \coth(\frac{\eta}{2}) \operatorname{csch}(\frac{\eta}{2})^2 \left(\frac{\eta}{n} \coth(\frac{\eta}{n}) + \eta \operatorname{csch}(\eta) \right) \\ &\quad - \operatorname{csch}(\frac{\eta}{2})^2 \left(\frac{1}{n} \coth(\frac{\eta}{n}) + \operatorname{csch}(\eta) - \eta \left(\frac{1}{n^2} \operatorname{csch}(\frac{\eta}{n})^2 + \operatorname{csch}(\eta) \coth(\eta) \right) \right) \\ &\quad + \coth(\frac{\eta}{2}) \left(-2 \left(\frac{1}{n^2} \operatorname{csch}(\frac{\eta}{n})^2 + \operatorname{csch}(\eta) \coth(\eta) \right) + \right. \\ &\quad \left. \eta \left(\frac{2}{n^3} \operatorname{csch}(\frac{\eta}{n})^2 \coth(\frac{\eta}{n}) + \operatorname{csch}(\eta) + 2 \operatorname{csch}(\eta)^3 \right) \right). \end{aligned}$$

18.

$$\begin{aligned} T(\eta) &= \frac{1}{4} \operatorname{csch}(\frac{\eta}{2})^2 \cdot \left(\eta \coth(\frac{\eta}{2}) \operatorname{csch}(\frac{\eta}{2}) \operatorname{sech}(\frac{\eta}{2}) - 2 \operatorname{csch}(\frac{\eta}{2}) \operatorname{sech}(\frac{\eta}{2}) \right. \\ &\quad + \eta \operatorname{csch}(\frac{\eta}{2})^2 \operatorname{sech}(\frac{\eta}{2})^2 (2 \cosh(\frac{\eta}{2})^2 - 1) \\ &\quad - 2 \sinh(\frac{\eta}{2}) \cosh(\frac{\eta}{2}) \operatorname{csch}(\frac{\eta}{2})^2 \operatorname{sech}(\frac{\eta}{2})^2 (2 \cosh(\frac{\eta}{2})^2 - 1) \\ &\quad + 2\eta \sinh(\frac{\eta}{2}) \cosh(\frac{\eta}{2}) \operatorname{csch}(\frac{\eta}{2}) \operatorname{sech}(\frac{\eta}{2}) \\ &\quad \left. + \eta \sinh(\frac{\eta}{2}) \cosh(\frac{\eta}{2}) \operatorname{csch}(\frac{\eta}{2})^3 \operatorname{sech}(\frac{\eta}{2})^3 \right) \\ &= \frac{1}{4} \operatorname{csch}(\frac{\eta}{2})^2 \cdot (\circ). \end{aligned}$$

We aim to show positivity of the expression $(\circ)$, implying positivity of the entire part independent of n . Using addition formulas for squares of hyperbolic functions we compute

$$\begin{aligned} (\circ) &= \left(\eta \operatorname{csch}(\frac{\eta}{2})^2 - 2 \operatorname{csch}(\frac{\eta}{2}) \operatorname{sech}(\frac{\eta}{2}) + 2\eta \operatorname{csch}(\frac{\eta}{2})^2 \operatorname{sech}(\frac{\eta}{2})^2 \cosh(\frac{\eta}{2})^2 - \eta \operatorname{csch}(\frac{\eta}{2})^2 \operatorname{sech}(\frac{\eta}{2})^2 \right. \\ &\quad \left. - 4 \operatorname{csch}(\frac{\eta}{2}) \operatorname{sech}(\frac{\eta}{2}) \cosh(\frac{\eta}{2})^2 + 2 \operatorname{csch}(\frac{\eta}{2}) \operatorname{sech}(\frac{\eta}{2}) + 2\eta + \operatorname{csch}(\frac{\eta}{2})^2 \operatorname{sech}(\frac{\eta}{2})^2 \right) \\ &= \operatorname{csch}(\frac{\eta}{2})^2 \left(\eta \cosh(\eta) + 2\eta - 2 \sinh(\eta) \right) \\ &= \eta \operatorname{csch}(\frac{\eta}{2})^2 \left(\sum_{k=0}^{\infty} \frac{\eta^{2k}}{(2k)!} + 2 - \sum_{k=0}^{\infty} \frac{2\eta^{2k}}{(2k+1)!} \right) = \eta \operatorname{csch}(\frac{\eta}{2})^2 \cdot \sum_{k=1}^{\infty} \eta^{2k} \frac{2k-1}{(2k+1)!} > 0. \end{aligned}$$

19. We show that $\eta_{B,n}$ is between $\eta_n/2$ and $2\eta_n$, where $\eta_n = 2 \operatorname{arccosh}(n)$, by comparing the values at $\eta_n/2$, η_n and $2\eta_n$:

$$\begin{aligned}
B\left(\frac{\eta_n}{2n}, \frac{2}{\eta_n}\right) &= \frac{n+1}{(n^2-1)^{1/2}} \frac{\eta_n}{2} \left(\frac{1}{n} \coth\left(\frac{\eta_n}{2n}\right) + \frac{1}{(n^2-1)^{1/2}} \right) \\
&= \frac{n+1}{(n^2-1)^{1/2}} \frac{\eta_n}{2} \left(\frac{1}{n} \left(\coth\left(\frac{\eta_n}{n}\right) + \operatorname{csch}\left(\frac{\eta_n}{n}\right) \right) + \frac{1}{(n^2-1)^{1/2}} \right), \\
B\left(\frac{\eta_n}{n}, \frac{1}{\eta_n}\right) &= \frac{n}{(n^2-1)^{1/2}} \left(\frac{\eta_n}{n} \coth\left(\frac{\eta_n}{n}\right) + \frac{\eta_n}{2n(n^2-1)^{1/2}} \right) = \frac{\eta_n}{(n^2-1)^{1/2}} \left(\coth\left(\frac{\eta_n}{n}\right) + \frac{1}{2(n^2-1)^{1/2}} \right), \\
B\left(\frac{2\eta_n}{n}, \frac{1}{2\eta_n}\right) &= \frac{2n^2-1}{2n(n^2-1)^{1/2}} \cdot 2\eta_n \left(\frac{1}{n} \coth\left(\frac{2\eta_n}{n}\right) + \operatorname{csch}(4 \operatorname{arccosh}(n)) \right) \\
&= \frac{2n^2-1}{n(n^2-1)^{1/2}} \cdot \eta_n \left(\frac{1}{n} \coth\left(\frac{\eta_n}{n}\right) - \frac{1}{n} \operatorname{csch}\left(\frac{2\eta_n}{n}\right) + \frac{1}{4n(n^2-1)^{1/2}(2n^2-1)} \right) \\
&= \frac{2n^2-1}{n^2} \cdot \frac{\eta_n}{(n^2-1)^{1/2}} \left(\coth\left(\frac{\eta_n}{n}\right) \left(1 - \frac{1}{2} \operatorname{sech}\left(\frac{\eta_n}{n}\right)^2\right) + \frac{1}{4(n^2-1)^{1/2}(2n^2-1)} \right).
\end{aligned}$$

The left cut-off. Because of the convexity of $B(\eta/n, 1/\eta)$ in the entire domain, we only need to compare at $\eta_n/2$ and η_n and do not need any interval estimates this time. We perform standard equivalence transformations to arrive at an equivalent, easy to handle, inequality:

$$\begin{aligned}
&B\left(\frac{\eta_n}{n}, \frac{1}{\eta_n}\right) < B\left(\frac{\eta_n}{2n}, \frac{2}{\eta_n}\right) \\
\iff &2 \coth\left(\frac{\eta_n}{n}\right) + \frac{1}{(n^2-1)^{1/2}} < (n + \frac{1}{n}) \coth\left(\frac{\eta_n}{n}\right) + (n + \frac{1}{n}) \operatorname{csch}\left(\frac{\eta_n}{n}\right) + \frac{n+1}{(n^2-1)^{1/2}} \\
\iff &(1 - \frac{1}{n}) \coth\left(\frac{\eta_n}{n}\right) < (1 + \frac{1}{n}) \operatorname{csch}\left(\frac{\eta_n}{n}\right) + \frac{n}{(n^2-1)^{1/2}} \\
\iff &(1 - \frac{1}{n})^2 \coth\left(\frac{\eta_n}{n}\right)^2 < (1 + \frac{1}{n})^2 (\coth\left(\frac{\eta_n}{n}\right)^2 - 1) + \frac{n^2}{n^2-1} + \frac{2}{n-1} \operatorname{csch}\left(\frac{\eta_n}{n}\right) \\
\iff &-\frac{4}{n} \coth\left(\frac{\eta_n}{n}\right)^2 < -\frac{(n+1)^2}{n^2} + \frac{n^2}{n^2-1} + \frac{2}{n-1} \operatorname{csch}\left(\frac{\eta_n}{n}\right) \\
\iff &4 \coth\left(\frac{\eta_n}{n}\right)^2 > \frac{(n+1)^2}{n} - \frac{n^3}{n^2-1} - \frac{2n}{n-1} \operatorname{csch}\left(\frac{\eta_n}{n}\right).
\end{aligned}$$

The left-hand side is strictly greater than 4. For the right-hand side, we estimate

$$\frac{(n+1)^2}{n} - \frac{n^3}{n^2-1} - \frac{2n}{n-1} \operatorname{csch}\left(\frac{\eta_n}{n}\right) < \frac{(n^2+2n+1)(n^2-1)-n^4}{n(n^2-1)} = \frac{2n^3-2n-1}{n(n^2-1)} = 2 - \frac{1}{n(n^2-1)}.$$

This is clearly less than 2 and so the value at $\eta_n/2$ is less than the value at η_n .

The right cut-off. We now compare the values at η_n and $2\eta_n$. We aim to show that

$$\begin{aligned}
&B\left(\frac{\eta_n}{n}, \frac{1}{\eta_n}\right) < B\left(\frac{2\eta_n}{n}, \frac{1}{2\eta_n}\right) \\
\iff &\coth\left(\frac{\eta_n}{n}\right) + \frac{1}{2(n^2-1)^{1/2}} < \frac{2n^2-1}{n^2} \coth\left(\frac{\eta_n}{n}\right) \left(1 - \frac{1}{2} \operatorname{sech}\left(\frac{\eta_n}{n}\right)^2\right) + \frac{1}{4n^2(n^2-1)^{1/2}} \\
\iff &\coth\left(\frac{\eta_n}{n}\right) \left(1 - 2 + \frac{1}{n^2} - (2 - \frac{1}{n^2}) \left(1 - \frac{1}{2} \operatorname{sech}\left(\frac{\eta_n}{n}\right)^2\right)\right) < (-1 + \frac{1}{2n^2}) \frac{1}{2(n^2-1)^{1/2}}.
\end{aligned}$$

The right-hand side is clearly negative and greater than -1 . We show that the left-hand side is also negative, but less than -1 . This will then confirm the initial inequality. It suffices to observe the following:

$$-1 + \frac{1}{n^2} - (2 - \frac{1}{n^2}) \left(1 - \frac{1}{2} \operatorname{sech}\left(\frac{\eta_n}{n}\right)^2\right) < -1 + \frac{1}{n^2} - \frac{1}{2} (2 - \frac{1}{n^2}) = -2 + \frac{1}{2n^2} < -1.$$

This confirms the claim for all $n \geq 2$.

20. We set

$$H(\eta) = h(\eta)h''(\eta) - h'(\eta)^2,$$

where $h(\eta) = \eta/n \coth(\eta/n) + \eta \operatorname{csch}(\eta)$, $n \geq 4$. We compute

$$\begin{aligned} H(\eta) &= h(\eta) \left(-2 \underbrace{\left(\frac{1}{n^2} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 + \operatorname{csch}(\eta) \coth(\eta) \right)}_{(X)} \right. \\ &\quad \left. + \eta \left(\frac{2}{n^3} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 \coth\left(\frac{\eta}{n}\right) + \operatorname{csch}(\eta) + 2 \operatorname{csch}(\eta)^3 \right) \right) \\ &\quad - \frac{h(\eta)^2}{\eta^2} + 2h(\eta) \underbrace{\left(\frac{1}{n^2} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 + \operatorname{csch}(\eta) \coth(\eta) \right)}_{(X)} \\ &\quad - \eta^2 \left(\frac{1}{n^2} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 + \operatorname{csch}(\eta) \coth(\eta) \right)^2. \end{aligned}$$

Note that the expressions (X) have a factor $-2h(\eta)$ and $2h(\eta)$, respectively, so they cancel out. We further have

$$\begin{aligned} H(\eta) &= \eta^2 \left(\frac{1}{n} \coth\left(\frac{\eta}{n}\right) + \operatorname{csch}(\eta) \right) \left(\frac{2}{n^3} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 \coth\left(\frac{\eta}{n}\right) + \operatorname{csch}(\eta) + 2 \operatorname{csch}(\eta)^3 \right) - \frac{1}{n^2} \coth\left(\frac{\eta}{n}\right)^2 \\ &\quad - \operatorname{csch}(\eta)^2 - \frac{2}{n} \coth\left(\frac{\eta}{n}\right) \operatorname{csch}(\eta) - \frac{\eta^2}{n^4} \operatorname{csch}\left(\frac{\eta}{n}\right)^4 - \eta^2 \operatorname{csch}(\eta)^2 \coth(\eta)^2 \\ &\quad - 2 \frac{\eta^2}{n^2} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 \operatorname{csch}(\eta) \coth(\eta). \end{aligned}$$

We multiply out the remaining brackets and order by positive and negative sign to obtain

$$\begin{aligned} H(\eta) &= \frac{2\eta^2}{n^4} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 \coth\left(\frac{\eta}{n}\right)^2 + \frac{2\eta^2}{n} \coth\left(\frac{\eta}{n}\right) \operatorname{csch}(\eta)^3 + \frac{\eta^2}{n} \coth\left(\frac{\eta}{n}\right) \operatorname{csch}(\eta) \\ &\quad + \frac{2\eta^2}{n^3} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 \coth\left(\frac{\eta}{n}\right) \operatorname{csch}(\eta) + 2\eta^2 \operatorname{csch}(\eta)^4 + \eta^2 \operatorname{csch}(\eta)^2 \\ &\quad - \frac{1}{n^2} \coth\left(\frac{\eta}{n}\right)^2 - \operatorname{csch}(\eta)^2 - \frac{2}{n} \coth\left(\frac{\eta}{n}\right) \operatorname{csch}(\eta) - \frac{\eta^2}{n^4} \operatorname{csch}\left(\frac{\eta}{n}\right)^4 \\ &\quad - \eta^2 \operatorname{csch}(\eta)^2 \coth(\eta)^2 - \frac{2\eta^2}{n^2} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 \operatorname{csch}(\eta) \coth(\eta). \end{aligned}$$

We now group together terms where we find a common factor $1/n^2$, or $\operatorname{csch}(\eta)^2$, or where we find $2\eta/n^2 \operatorname{csch}(\eta/n)^2 \operatorname{csch}(\eta)$. We get

$$\begin{aligned} H(\eta) &= \frac{1}{n^2} \left(\frac{2\eta^2}{n^2} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 \coth\left(\frac{\eta}{n}\right)^2 - \frac{\eta^2}{n^2} \operatorname{csch}\left(\frac{\eta}{n}\right)^4 - \coth\left(\frac{\eta}{n}\right)^2 \right) + \frac{1}{n} \coth\left(\frac{\eta}{n}\right) \operatorname{csch}(\eta) (\eta^2 - 2) \\ &\quad + \operatorname{csch}(\eta)^2 \left(\frac{2\eta^2}{n} \coth\left(\frac{\eta}{n}\right) \operatorname{csch}(\eta) + 2\eta^2 \operatorname{csch}(\eta)^2 + \eta^2 - 1 - \eta^2 \coth(\eta)^2 \right) \\ &\quad + \frac{2\eta}{n^2} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 \operatorname{csch}(\eta) \left(\frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) - \eta \coth(\eta) \right). \end{aligned}$$

We use the formula $\coth(x)^2 - \operatorname{csch}(x)^2 = 1$ from above. This gives us

$$\begin{aligned} H(\eta) &= \frac{1}{n^2} \left(\frac{\eta^2}{n^2} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 \coth\left(\frac{\eta}{n}\right)^2 - \frac{\eta^2}{n^2} \operatorname{csch}\left(\frac{\eta}{n}\right)^2 - \operatorname{csch}\left(\frac{\eta}{n}\right)^2 - 1 \right) + \frac{1}{n} \coth\left(\frac{\eta}{n}\right) \operatorname{csch}(\eta) (\eta^2 - 2) \\ &\quad + \operatorname{csch}(\eta)^2 \left(\frac{2\eta^2}{n} \coth\left(\frac{\eta}{n}\right) \operatorname{csch}(\eta) + 2\eta^2 \operatorname{csch}(\eta)^2 + \eta^2 - 1 - \eta^2 \operatorname{csch}(\eta)^2 - \eta^2 \right) \\ &\quad + \frac{2\eta}{n^2} \coth\left(\frac{\eta}{n}\right)^2 \operatorname{csch}(\eta) \left(\frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) - \eta \coth(\eta) \right) - \frac{2\eta}{n^2} \operatorname{csch}(\eta) \left(\frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) - \eta \coth(\eta) \right). \end{aligned}$$

We re-group the terms in the first bracket, collect terms with factor $1/n$ and pull out $\coth(\eta/n) \operatorname{csch}(\eta)$ for this collection, and collect terms with factor $2\eta/n$ and pull out $\operatorname{csch}(\eta)$, and lastly, collect terms with factor $\operatorname{csch}(\eta)^2$. We obtain

$$\begin{aligned}
H(\eta) &= \frac{1}{n^2} \left(\operatorname{csch}\left(\frac{\eta}{n}\right)^2 \left(\left(\frac{\eta}{n}\right)^2 \coth\left(\frac{\eta}{n}\right)^2 - 1 \right) + \left(\frac{\eta}{n}\right)^2 \operatorname{csch}\left(\frac{\eta}{n}\right)^2 - 1 \right) \\
&\quad + \frac{1}{n} \coth\left(\frac{\eta}{n}\right) \operatorname{csch}(\eta) \left(\eta^2 - 2 - \frac{2\eta}{n} \coth\left(\frac{\eta}{n}\right) (\eta \coth(\eta) - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right)) \right) \\
&\quad + \frac{2\eta}{n} \operatorname{csch}(\eta) \left(\eta \coth(\eta) - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) \right) \\
&\quad + \operatorname{csch}(\eta)^2 \left(\frac{2\eta^2}{n} \coth\left(\frac{\eta}{n}\right) \operatorname{csch}(\eta) + \eta^2 \operatorname{csch}(\eta)^2 - 1 \right) \\
&= \frac{1}{n^2} \left(\operatorname{csch}\left(\frac{\eta}{n}\right)^2 \left(\left(\frac{\eta}{n}\right)^2 \coth\left(\frac{\eta}{n}\right)^2 - 1 \right) + \left(\frac{\eta}{n}\right)^2 \operatorname{csch}\left(\frac{\eta}{n}\right)^2 - 1 \right) \\
&\quad + \frac{1}{n} \coth\left(\frac{\eta}{n}\right) \operatorname{csch}(\eta) \left(\eta^2 - 2 - \frac{2\eta^2}{n} \coth\left(\frac{\eta}{n}\right) \coth(\eta) + 2\left(\frac{\eta}{n}\right)^2 \coth\left(\frac{\eta}{n}\right)^2 \right) \\
&\quad + \frac{2\eta}{n} \operatorname{csch}(\eta) \left(2 \cdot \frac{1}{2} \eta \coth(\eta) - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) - \frac{\eta}{2\eta} \operatorname{csch}(\eta) \right) \\
&\quad + \operatorname{csch}(\eta)^2 \left(\frac{2\eta^2}{n} \coth\left(\frac{\eta}{n}\right) \operatorname{csch}(\eta) + \eta^2 \operatorname{csch}(\eta)^2 \right) \\
&= \frac{\operatorname{csch}\left(\frac{\eta}{n}\right)^2}{n^2} \underbrace{\left(\left(\frac{\eta}{n}\right)^2 \coth\left(\frac{\eta}{n}\right)^2 - 1 \right) + \left(\frac{\eta}{n}\right)^2 - \sinh\left(\frac{\eta}{n}\right)^2 \right)}_{(1)} \\
&\quad + \frac{1}{n} \coth\left(\frac{\eta}{n}\right) \operatorname{csch}(\eta) \left(\eta \underbrace{\left(\eta - \frac{2\eta}{n} \coth\left(\frac{\eta}{n}\right) \coth(\eta) \right)}_{(2)} + 2 \underbrace{\left(\left(\frac{\eta}{n}\right)^2 \coth\left(\frac{\eta}{n}\right)^2 - 1 \right)}_{>0} \right) \\
&\quad + \frac{2\eta}{n} \operatorname{csch}(\eta) \left(\underbrace{\frac{\eta \operatorname{csch}(\eta)}{2} \left(\cosh(\eta) - \frac{n}{\eta^2} \right)}_{(3)} + \underbrace{\frac{\eta}{2} \coth(\eta) - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right)}_{(4)} \right) \\
&\quad + \underbrace{\operatorname{csch}(\eta)^2 \left(\frac{2\eta^2}{n} \coth\left(\frac{\eta}{n}\right) \operatorname{csch}(\eta) + \eta^2 \operatorname{csch}(\eta)^2 \right)}_{>0},
\end{aligned}$$

which is the final expression we use.

21. We will now go through the expressions (1) – (4) separately. We first analyze everything for $n \geq 4$ and then give the details for the case $n = 3$.

Expression (1): we use the change of variables $t = \eta/n$ and subsequently the duplication formulas for hyperbolic functions. For $t \in \frac{1}{n}I_n$ we want to show that

$$\begin{aligned}
&0 < t^2 \coth(t)^2 - \cosh(t)^2 + t^2 \\
\iff &\cosh(t)^2 < t^2 \frac{\cosh(t)^2 + \sinh(t)^2}{\sinh(t)^2} \\
\iff &1 < 4t^2 \frac{\cosh(t)^2 + \sinh(t)^2}{4 \sinh(t)^2 \cosh(t)^2} = 4t^2 \cosh(2t) \operatorname{csch}(2t)^2 = 4t^2 \coth(2t) \operatorname{csch}(2t) \\
\stackrel{(x=2t)}{\iff} &0 > \sinh(x) - x^2 \coth(x), \quad x \in \frac{2}{n}I_n.
\end{aligned}$$

We observe that the asymptotic behavior of the last expression close to 0 is

$$\lim_{x \rightarrow 0^+} (\sinh(x) - x^2 \coth(x)) = 0.$$

To see when our inequality holds, we compute the derivative of the above expression:

$$\begin{aligned}
\frac{d}{dx} [\sinh(x) - x^2 \coth(x)] &= \cosh(x) - 2x \coth(x) + x^2 \operatorname{csch}(x)^2 \\
&< \cosh(x) - x \coth(x) - 1 + x^2 \operatorname{csch}(x)^2 \\
&= \cosh(x) (1 - x \operatorname{csch}(x)) - (1 + x \operatorname{csch}(x)) (1 - x \operatorname{csch}(x)) \\
&= (1 - x \operatorname{csch}(x)) (\cosh(x) - 1 - x \operatorname{csch}(x)).
\end{aligned}$$

The expression $\cosh(x) - 1 - x \operatorname{csch}(x)$ is strictly monotonically increasing as $\cosh(x)$ is increasing and $x \operatorname{csch}(x)$ is decreasing (see computation 11) and its value at 0 is -1. Therefore, it has a unique zero x_0 and we can check that it is in the interval $(1.19, 1.2)$. To sum it up, this tells us that $\sinh(x) - x^2 \coth(x)$ is strictly monotonically decreasing on $(0, x_0]$ and strictly monotonically increasing on $[x_0, \infty)$. As a consequence, $\sinh(2t) - (2t)^2 \coth(2t)$ has its minimum at $t_0 = x_0/2$. In particular, for all $t \in (0, 4/3)$ the function's maximum is only attained at one of the boundary points;

$$\sinh(2t) - 4t^2 \coth(2t) < \max \left\{ \sinh\left(2 \cdot \frac{4}{3}\right) - 4 \cdot \frac{16}{9} \coth\left(2 \cdot \frac{4}{3}\right), 0 \right\} = 0,$$

proving the initial inequality and that expression (1) is positive. Here we used the behavior of $x \coth(x)$ close to zero (see Claim 1 in computation 10):

$$\lim_{x \rightarrow 0^+} \sinh(x) - x^2 \coth(x) = 0 - \lim_{x \rightarrow 0^+} x \cdot x \coth(x) = -0 \cdot 1 = 0.$$

This first expression, as well as the following one, are the reason why we had to improve the intervals $(\eta_n/2, 2\eta_n)$ to I_n for the first few $4 \leq n \leq 10$ and why the case $n = 3$ needs then to be treated separately.

Expression (2): Recall that $t \coth(t)$ is strictly increasing for all $t > 0$ (see Claim 1 in computation 10). Hence, for $\eta/n = t \in (0, 4/3)$ we obtain the following estimate:

$$\eta - \frac{2\eta}{n} \coth\left(\frac{\eta}{n}\right) \coth(\eta) \geq \eta - 2 \cdot \frac{4}{3} \coth\left(\frac{4}{3}\right) \coth(\eta).$$

What we are left with is strictly monotonically increasing. We evaluate the right-hand side at $\operatorname{arccosh}(11) > 3.08$ to see that it is strictly positive for all $n \geq 11$ and $\eta \in I_n$. For $4 \leq n \leq 10$, we evaluate the right-hand side of the above inequality at $\eta = 3.08 > \operatorname{arccosh}(n)$, which suffices for the positivity and explains the choice of the bounds of the interval I_n for $n \geq 4$.

Expression (3): for $n \geq 4$, $\eta \in I_n \subset (\eta_n/2, 2\eta_n)$ and $\eta_n = 2 \operatorname{arccosh}(n)$, we have $\eta > 1$ and

$$\cosh(\eta) - \frac{n}{\eta^2} > n - \frac{n}{1} = 0.$$

Expression (4): we want to show that

$$\frac{1}{2}\eta \coth(\eta) - \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) > 0 \iff \frac{1}{2}\eta \coth(\eta) > \frac{\eta}{n} \coth\left(\frac{\eta}{n}\right).$$

The left-hand side of the second inequality is bounded from below by

$$\frac{1}{2}\eta \coth(\eta) > \frac{1}{2} \cdot 3.08 \coth(3.08) > 1.54,$$

whereas its right-hand side is bounded from above:

$$\frac{\eta}{n} \coth\left(\frac{\eta}{n}\right) < \frac{4}{3} \coth\left(\frac{4}{3}\right) < 1.53.$$

Therefore, the original inequality holds. This concludes the estimate, i.e., for all $n \geq 4$,

$$h(\eta)h''(\eta) - h'(\eta)^2 > 0, \quad \eta \in I_n.$$

22. Towards the first two claims, we use interval estimates for subintervals of I_3 of length 0.01. This already shows that $h \geq 1.6$ on I_3 . For $h'(\eta)$, $\eta \in I_3$, we observe that

$$h'(\eta) = \underbrace{\frac{1}{3} \coth(\frac{\eta}{3})}_{\searrow} + \underbrace{\operatorname{csch}(\eta)}_{\searrow} - \underbrace{\frac{\eta}{9} \operatorname{csch}(\frac{\eta}{3})^2}_{\searrow} - \underbrace{\eta \operatorname{csch}(\eta)}_{\searrow} \underbrace{\coth(\eta)}_{\searrow}.$$

To show $-0.22 < h' < 0.22$, we use subintervals of length 0.01, again. For the third claim,

$$\begin{aligned} \frac{9}{2} h''(\eta) &= -\operatorname{csch}(\frac{\eta}{3})^3 - 9 \operatorname{csch}(\eta) \coth(\eta) + \eta \left(\frac{1}{3} \operatorname{csch}(\frac{\eta}{3})^2 \coth(\frac{\eta}{3}) + 9 \operatorname{csch}(\eta)^3 + \frac{9}{2} \operatorname{csch}(\eta) \right) \\ &= \operatorname{csch}(\frac{\eta}{3})^2 \underbrace{\left(\frac{\eta}{3} \coth(\frac{\eta}{3}) - 1 \right)}_{>0} + 9 \operatorname{csch}(\eta) \underbrace{\left(\eta \operatorname{csch}(\eta)^2 + \frac{\eta}{2} - \coth(\eta) \right)}_{>0}, \quad \eta \in I_3. \end{aligned}$$

The last estimate has not been established before. Yet, it is easy to see that on I_3

$$\eta \operatorname{csch}(\eta)^2 + \frac{\eta}{2} - \coth(\eta) \geq 0 + \frac{3.3}{2} - \coth(3.3) > 0.64.$$

23. We provide the proofs for $n = 2$ for all cases.

The lower bound. We start again with the lower bound.

$$\begin{aligned} A(\frac{\eta}{2}, \frac{1}{\eta}) &= \frac{\eta}{2} \tanh(\frac{\eta}{2}) (\operatorname{csch}(\frac{\eta}{2}) - 2 \operatorname{csch}(\eta)) = \frac{\eta}{2} \tanh(\frac{\eta}{2}) (\operatorname{csch}(\frac{\eta}{2}) - \operatorname{csch}(\frac{\eta}{2}) \operatorname{sech}(\frac{\eta}{2})) \\ &= \frac{\eta}{2} \frac{\sinh(\frac{\eta}{2})}{\cosh(\frac{\eta}{2})} \operatorname{csch}(\frac{\eta}{2}) (1 - \operatorname{sech}(\frac{\eta}{2})) = \frac{\eta}{2} \operatorname{sech}(\frac{\eta}{2}) (1 - \operatorname{sech}(\frac{\eta}{2})). \end{aligned}$$

To avoid all the factors of $1/2$ which arise from derivatives, we optimize $t \operatorname{sech}(t) (1 - \operatorname{sech}(t))$.

$$\frac{d}{dt} [t \operatorname{sech}(t) (1 - \operatorname{sech}(t))] = \operatorname{sech}(t) (1 - \operatorname{sech}(t) - t \tanh(t) (1 - 2 \operatorname{sech}(t))).$$

Since the fixed point of $\coth$ is in $[1.19, 1.2]$, the first derivative is strictly positive on $(0, 1.19]$. On the other hand, for all $t > 1.4$, $2 \operatorname{sech}(t) < 1$. Therefore,

$$\cosh(t) \cdot \frac{d}{dt} [t \operatorname{sech}(t) (1 - \operatorname{sech}(t))] < 1 - 0 - t \tanh(t) (1 - 2 \operatorname{sech}(t)), \quad t \geq 1.4,$$

and the left-hand side of the above inequality is strictly monotonically decreasing. Now, by evaluating at $t = 2.1$, we see that the right-hand side is negative there and, hence, we conclude that the left-hand side is strictly negative for all $t \geq 2.1$. Taking into account that $t = \eta/2$, this means that A has its maximum in $(2.38, 4.2)$. Notice that with the methods for $n \geq 3$, we would have reduced it to only $(1.31, 7.91)$ at best (see Figure 1).

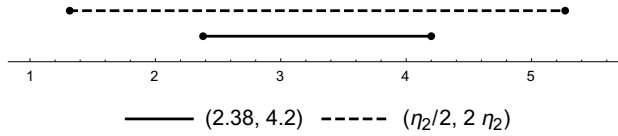


Figure 1. Reduction of the interval in question from $(\eta_2/2, 2\eta_2)$ to $(2.38, 4.2)$.

Finally, we show that $A'(t)$ has only one zero in $(2.38, 4.2)$. We compute (keeping $t = \eta/2$)

$$\begin{aligned} \frac{d}{dt} [1 - \operatorname{sech}(t) - t \tanh(t) + 2t \operatorname{sech}(t) \tanh(t)] \\ &= t \operatorname{sech}(t) (2 - 3t \tanh(t)^2) + \operatorname{sech}(t) \tanh(t) \left(\frac{3}{2} - \cosh(t) \right) \\ &\quad + \operatorname{sech}(t) \tanh(t) \left(\frac{3}{2} - t \operatorname{csch}(t) - t \tanh(t) \right). \end{aligned}$$

Considering the monotonicity of the hyperbolic functions, as well as their combinations with monomials, which we have already exploited numerous times, we get

$$2 - 3t \tanh(t)^2 < 0, \quad \frac{3}{2} - \cosh(t) < 0, \quad \frac{3}{2} - t \operatorname{csch}(t) - t \tanh(t) < 0$$

with a simple interval estimate. All in all, the expression

$$1 - \operatorname{sech}(t) - t \tanh(t) + 2t \operatorname{sech}(t) \tanh(t)$$

is strictly monotonically decreasing on (1.19, 2.1), so $t \operatorname{sech}(t) (1 - \operatorname{sech}(t))$ has exactly one critical point in this interval. By evaluating, one can now tell that it lies in the interval (1.85, 1.87). Equivalently, we have a unique maximum for $A(\frac{\eta}{2}, \frac{1}{\eta})$, which is assumed in the interval $(2 \cdot 1.85, 2 \cdot 1.87) = (3.7, 3.74)$.

The upper bound. The convexity of the upper bound still stands as already shown. The minimum is therefore also unique. Since we have fixed $n = 2$, we can evaluate and also improve the interval for $\eta_{B,2}$. Indeed, by evaluating the first derivative, we can tell that the unique minimizer $\eta_{B,n}$ lies in (3, 3.1).

The condition number. What remains to be determined is the minimum of the condition number κ . The only possible interval where the derivative of A and B have the same sign is (3, 3.74). We compute

$$\begin{aligned} \kappa(\frac{\eta}{2}, \frac{1}{\eta}) &= \coth(\frac{\eta}{2})^2 \frac{\frac{\eta}{2} \coth(\frac{\eta}{2}) + \eta \operatorname{csch}(\eta)}{\frac{\eta}{2} \operatorname{csch}(\frac{\eta}{2}) - \eta \operatorname{csch}(\eta)} = \coth(\frac{\eta}{2})^2 \frac{\coth(\frac{\eta}{2}) + \operatorname{csch}(\frac{\eta}{2}) \operatorname{sech}(\frac{\eta}{2})}{\operatorname{csch}(\frac{\eta}{2}) - \operatorname{csch}(\frac{\eta}{2}) \operatorname{sech}(\frac{\eta}{2})} \\ &= \coth(\frac{\eta}{2})^2 (\cosh(\frac{\eta}{2})^2 + 1) \frac{1}{\cosh(\frac{\eta}{2}) - 1}. \end{aligned}$$

We change the variable to $t = \eta/2$ and show the strict log-convexity of each of the terms on the interval (1.5, 1.87).

$$\frac{d^2}{dt^2} [\log(\coth(t)^2)] = -2 \frac{d}{dt} [\operatorname{csch}(t) \operatorname{sech}(t)] = 2 (\operatorname{csch}(t)^2 + \operatorname{sech}(t)^2) > 0.$$

Secondly,

$$\frac{d^2}{dt^2} [\log(\cosh(t)^2 + 1)] = \frac{4 + 12 \cosh(2t)}{(\cosh(2t) + 3)^2} > 0.$$

Finally,

$$-\frac{d^2}{dt^2} [\log(\cosh(t) - 1)] = \frac{\cosh(t)}{(\cosh(t) + 1)^2} > 0.$$

So, the condition number is strictly log-convex on (3, 3.74) and assumes its unique minimum at $\eta_{\kappa,2}$ in this interval.

24. We compute

$$\begin{aligned} \frac{d}{d\eta} B(\frac{\eta}{n}, \frac{1}{n}) \Big|_{\eta=\eta_n} &= -\frac{1}{2} \operatorname{csch}(\frac{\eta_n}{2})^2 \left(\frac{\eta_n}{n} \coth(\frac{\eta_n}{n}) + \eta_n \operatorname{csch}(\eta_n) \right) \\ &\quad + \coth(\frac{\eta_n}{2}) \left(\frac{1}{n} \coth(\frac{\eta_n}{n}) + \operatorname{csch}(\eta_n) \right) \\ &\quad - \coth(\frac{\eta_n}{2}) \left(\eta_n \left(\frac{1}{n^2} \operatorname{csch}(\frac{\eta_n}{n})^2 + \operatorname{csch}(\eta_n) \coth(\eta_n) \right) \right) \\ &= -\frac{1}{2(n^2-1)} \left(\frac{\eta_n}{n} \coth(\frac{\eta_n}{n}) + \frac{\eta_n}{2n(n^2-1)^{1/2}} \right) \\ &\quad + \frac{n}{(n^2-1)^{1/2}} \left(\frac{1}{n} \coth(\frac{\eta_n}{n}) + \frac{1}{2n(n^2-1)^{1/2}} \right) \\ &\quad - \frac{n}{(n^2-1)^{1/2}} \left(\eta_n \left(\frac{1}{n^2} \operatorname{csch}(\frac{\eta_n}{n})^2 + \frac{1}{2n(n^2-1)^{1/2}} \frac{2n^2-1}{2n(n^2-1)^{1/2}} \right) \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2n(n^2-1)^{1/2}} \left(-\frac{\eta_n}{2(n^2-1)} + \frac{n}{(n^2-1)^{1/2}} \right) - \frac{1}{2(n^2-1)} \frac{\eta_n}{n} \coth\left(\frac{\eta_n}{n}\right) \\
&\quad + \frac{1}{(n^2-1)^{1/2}} \coth\left(\frac{\eta_n}{n}\right) - \frac{1}{(n^2-1)^{1/2}} \frac{\eta_n}{n} \operatorname{csch}\left(\frac{\eta_n}{n}\right)^2 - \frac{\eta_n}{n} \frac{1}{(n^2-1)^{1/2}} \frac{2n^2-1}{4(n^2-1)} \\
&= \frac{1}{2(n^2-1)} \left(1 - \frac{\eta_n}{n} \frac{1}{2(n^2-1)^{1/2}} - \frac{\eta_n}{n} \coth\left(\frac{\eta_n}{n}\right) \right) \\
&\quad + \frac{1}{(n^2-1)^{1/2}} \left(\underbrace{\coth\left(\frac{\eta_n}{n}\right) - \frac{\eta_n}{n} \operatorname{csch}\left(\frac{\eta_n}{n}\right)^2 - \frac{\eta_n}{2n}}_{=(I)} + \underbrace{\frac{\eta_n}{n} \frac{1}{4(n^2-1)}}_{\geq 0} \right).
\end{aligned}$$

25. We have for all $t > 0$

$$\begin{aligned}
\frac{d^2}{dt^2} \left[\coth(t) - t \coth(t)^2 + \frac{t}{2} - \frac{t}{14} \right] &= \frac{d}{dt} \left[-\operatorname{csch}(t)^2 - \coth(t)^2 + 2t \operatorname{csch}(t)^2 \coth(t) + \frac{3}{7} \right] \\
&= \frac{d}{dt} \left[-2 \operatorname{csch}(t)^2 - 1 + 2t \operatorname{csch}(t)^2 \coth(t) \right] \\
&= 4 \operatorname{csch}(t)^2 \coth(t) + 2 \coth(t) \operatorname{csch}(t)^2 \\
&\quad - 2t \operatorname{csch}(t)^4 - 4t \operatorname{csch}(t)^2 \coth(t)^2 \\
&= 2 \operatorname{csch}(t)^2 (3 \coth(t) - 3t \coth(t)^2 + t) \\
&= \operatorname{csch}(t)^4 (6 \cosh(t) \sinh(t) - 6t \coth(t)^2 + t \sinh(t)^2) \\
&= \operatorname{csch}(t)^4 (3 \sinh(2t) - 3t(\cosh(2t) + 1) + t(\cosh(2t) - 1)) \\
&= \operatorname{csch}(t)^4 \left(6t + \sum_{k=1}^{\infty} \frac{(2t)^{2k+1}}{(2k+1)!} - 2t - \sum_{k=1}^{\infty} \frac{(2t)^{2k+1}}{(2k)!} - 4t \right) \\
&= -\operatorname{csch}(t)^4 \sum_{k=1}^{\infty} 2k \frac{(2t)^{2k+1}}{(2k+1)!} < 0.
\end{aligned}$$

26. We let $t \rightarrow 0$ from above and compute

$$\begin{aligned}
\lim_{t \rightarrow 0^+} \coth(t) - t \coth(t)^2 + \frac{t}{2} - \frac{t}{14} &= \lim_{t \rightarrow 0^+} t^2 \operatorname{csch}(t)^2 \frac{\cosh(t) \sinh(t) - t \cosh(t)^2}{t^2} \\
&= \lim_{t \rightarrow 0^+} 1 \frac{\sum_{k=0}^{\infty} \frac{2^{2k+1}}{(2k+1)!} t^{2k+1} - 2t - t \sum_{k=1}^{\infty} \frac{2^{2k}}{(2k)!} t^{2k}}{2t^2} \\
&= \lim_{t \rightarrow 0^+} \sum_{k=1}^{\infty} \frac{2^{2k+1} (1-k)}{(2k+1)!} t^{2k-1} = 0,
\end{aligned}$$

due to the convergence of the series.

27.

$$\begin{aligned}
\left. \frac{d}{d\eta} B\left(\frac{\eta}{n}, \frac{1}{n}\right) \right|_{\eta=\eta_n} &= \frac{1}{2(n^2-1)} \left(1 - \frac{\rho(n)}{2(n^2-1)^{1/2}} - \rho(n) \coth(\rho(n)) \right) \\
&\quad + \frac{1}{(n^2-1)^{1/2}} \left(\coth(\rho(n)) - \rho(n) \operatorname{csch}(\rho(n))^2 - \frac{\rho(n)}{2} + \frac{\rho(n)}{4(n^2-1)} \right) \\
&> \frac{1}{2(n^2-1)} \left(1 - \frac{\rho(n)}{2(n^2-1)^{1/2}} - \rho(n) \coth(\rho(n)) + (n^2-1)^{1/2} \frac{\rho(n)}{7} \right) \\
&> \frac{1}{2(n^2-1)} \left(\underbrace{1 - \frac{\rho(n)}{2(n^2-1)^{1/2}}}_{\searrow} - \underbrace{\rho(n) \coth(\rho(n))}_{\searrow} + \underbrace{\left(1 - \frac{1}{n^2}\right)^{1/2}}_{\nearrow} \underbrace{\frac{2 \operatorname{arccosh}(n)}{7}}_{\nearrow} \right).
\end{aligned}$$

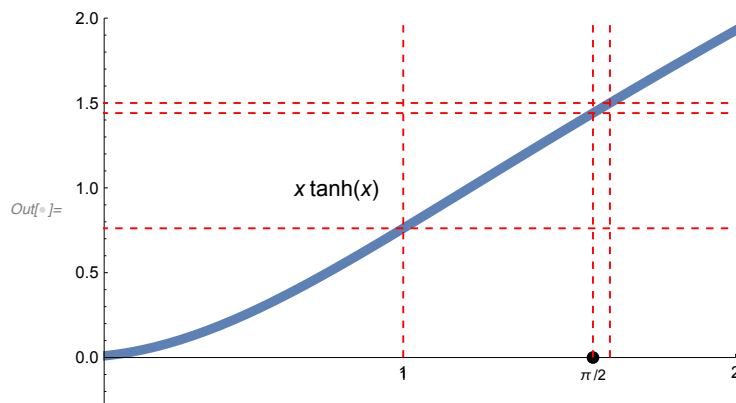
5. Often used functions

$x \tanh(x)$

```

In[ ]:= lineStyle = {Thickness[0.003], Red, Dashed};
line1 = Line[{{1, 0}, {1, 2}}];
line2 = Line[{{1.6219, 0}, {1.6219, 2}}];
line3 = Line[{{0, 1.5}, {2, 1.5}}];
line4 = Line[{{0, Tanh[1.]}}, {2, Tanh[1.]}];
line5 = Line[{{Pi / 2, 0}, {Pi / 2, 2}}];
line6 = Line[{{0, Pi / 2 Tanh[Pi / 2]}}, {2, Pi / 2 Tanh[Pi / 2]}];
pA1 = Plot[{x Tanh[x]}, {x, 0.1, 2}, PlotRange -> {{0.1, 2}, {-0.3, 2}},
  PlotStyle -> {Thickness[0.015]}, Ticks -> {{1, 2}, Automatic},
  Epilog -> {PointSize[0.02], Black, Point[{Pi / 2, 0}], Black,
    FontSize -> 8, Text[Pi "/2", {Pi / 2, -0.1}], Text[1, {1, -0.1}],
    Directive[lineStyle], line1, line2, line3, line4, line5,
    line6, Black, FontSize -> 12, Text[x Tanh[x], {0.8, 1}]}]

```



```
NumberForm[1.19 Tanh[1.19], 24]
```

```
NumberForm[1.2 Tanh[1.2], 24]
```

$x \coth(x)$

```

p52 := Plot[Style[x Coth[x], Black], {x, 0, 5},
  PlotLegends -> x Coth[x], PlotRange -> {Automatic, {0, Automatic}}, Epilog ->
    {Gray, Dashed, Line[{{0, 1}, {5, 1}}], Text[Style["1", Black], {5, 1.18}]}]
p52

```

$x \operatorname{csch}(x)$ and $x^2 \operatorname{csch}(x)$

```

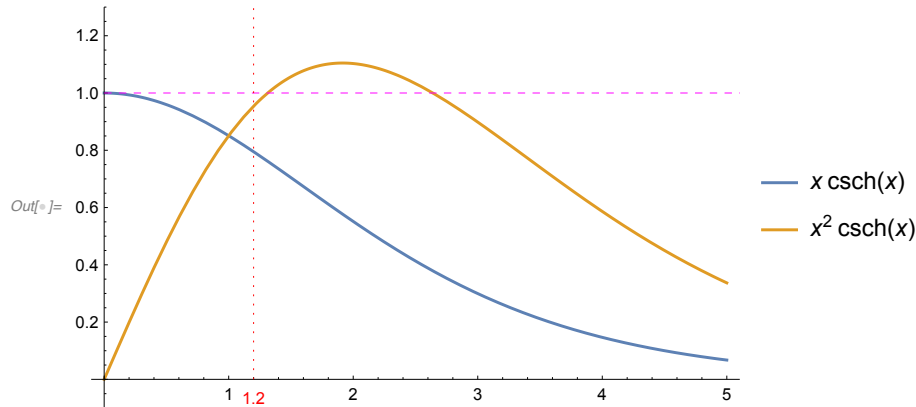
p54 := Plot[Style[x Csch[x], Black], {x, 0, 5},
  PlotRange -> {Automatic, {-0.1, 1.3}}, PlotLegends -> "Expressions",
  Epilog -> {Gray, Dashed, Line[{{0, 1}, {10, 1}}], Gray, Dotted,
    Line[{{1.2, 0}, {1.2, 1.3}}], Text[Style["1.2", Red], {1.2, -0.065}]}]

```

```

In[ ]:= Plot[{x Csch[x], x^2 Csch[x]}, {x, 0, 5},
  PlotRange -> {Automatic, {-0.1, 1.3}}, PlotLegends -> "Expressions",
  Epilog -> {Magenta, Dashed, Line[{{0, 1}, {10, 1}}], Red, Dotted,
    Line[{{1.2, 0}, {1.2, 1.3}}], Text[Style["1.2", Red], {1.2, -0.065}]}]

```



$$4 \sinh(x) < x + x^3 \text{ if } 0 \leq x \leq 1.2$$

```

In[ ]:= Plot[Sinh[x] - x - x^3, {x, 0, 1.2}]
NumberForm[Sinh[1.2] / 1.2, 24]
ρ(n) = 2 arccosh(n)/n

```

```

In[ ]:= rho[n_] := 2 ArcCosh[n] / n
p57 := Plot[Style[rho[n], Black], {n, 2, 100}]
Show[p57]

```

```

In[ ]:= 1. + 1 / (2^2 - 1)
ArcCosh[2.]^2

```

6. The hyperbolicsecant

```

In[ ]:= a[k_] := Pi * (k + 1 / 2)
fA6[t_] := t * Sum[Sech[a[k] t]^2, {k, 0, 20}]
fB6[t_] := t * Sum[Csch[a[k] t]^2, {k, 0, 20}]
hA6[t_] := t * fA6'[t]
hB6[t_] := t * fB6'[t]
p1 = Plot[{fA6[t], fB6[t], hA6[t], hB6[t]},
  {t, 0.1, 4}, PlotRange -> {{0.1, 3}, {-2, 2}}, PlotLegends ->
  Placed[{Subscript[f, A], Subscript[f, B], Subscript[h, A], Subscript[h, B]},
    Below], PlotLabel -> "Comparison of the functions' behaviour",
  Epilog -> {Directive[Dashed, Thickness[0.001]], Line[{{0, 1/Pi}, {4, 1/Pi}}],
    Line[{{0, hA[1.]}, {4, hA[1.]}], Line[{{1, 0}, {1, -1}}]}]

```

```

p2 = Plot[{2 Pi fA6[t / 2] + 2 Pi fA6[1 / t] - 2, 2 Pi fB6[2 / t] + 2 Pi fB6[t] + 2},
  {t, 0.1, 5}, PlotRange → {{0.1, 3}, {0, 6}},
  PlotLegends → Placed[{ "A( $\eta/2, 1/\eta$ )", "B( $\eta/2, 1/\eta$ )" }, Below],
  PlotLabel → "The optimal bounds for density 2"]

p3 = Plot[{5 Pi fA6[t / 5] + 5 Pi fA6[1 / t] - 5, 5 Pi fB6[5 / t] + 5 Pi fB6[t] + 5},
  {t, 0.2, 5}, PlotRange → {{0.1, 3}, {0, 6}},
  PlotLegends → Placed[{ "A( $\eta/5, 1/\eta$ )", "B( $\eta/5, 1/\eta$ )" }, Below],
  PlotLabel → "The optimal bounds for density 5"]

p4 = Plot[{Pi fA6[t / 2] + Pi fA6[1 / t] - 1,
  Pi fA6[t / 5] + 1 Pi fA6[1 / t] - 1, 1 Pi fA6[t / 8] + 1 Pi fA6[1 / t] - 1},
  {t, 0.2, 5}, PlotRange → {{0.1, 4}, {0, 2}}, PlotLegends →
  Placed[LineLegend[{ "A( $\eta/2, 1/\eta$ )" / 2, "A( $\eta/5, 1/\eta$ )" / 5, "A( $\eta/8, 1/\eta$ ) / 8" / 8 },
  LegendLayout → {"Row", 1}], Below],
  PlotLabel → "The optimal lower bounds for densities 2, 5, 8"]

p5 = Plot[{Pi fB6[2 / t] + Pi fB6[t] + 1,
  Pi fB6[t] + 1 Pi fB6[5 / t] + 1, 1 Pi fB6[t] + 1 Pi fB6[8 / t] + 1},
  {t, 0.2, 5}, PlotRange → {{0.1, 4}, {0, 4}}, PlotLegends →
  Placed[LineLegend[{ "B( $\eta/2, 1/\eta$ )" / 2, "B( $\eta/5, 1/\eta$ )" / 5, "B( $\eta/8, 1/\eta$ ) / 8" / 8 },
  LegendLayout → {"Row", 1}], Below],
  PlotLabel → "The optimal upper bounds for densities 2, 5, 8"]

```

Lower bound

Monotonicity of h_A

```

NumberForm[1.6218 Tanh[1.6218], 24]
NumberForm[1.6219 Tanh[1.6219], 24]
NumberForm[1 - 2  $\times$  1.22 Sech[1.19]2, 24]
NumberForm[3.2438 / Pi, 24]

```

Strict convexity close to 1

```

In[ ]:= NumberForm[Pi / 2 Tanh[Pi / 2.], 24]
NumberForm[1.04 Pi / 2 Tanh[1.04 Pi / 2], 24]
NumberForm[3. Pi / 2 Tanh[3 Pi / 2], 24]

In[ ]:= Plot[{-3 x2, -2 x2 + 5 x - 2}, {x, 1, 20}, PlotLegends → "Expressions"]

In[ ]:= NumberForm[-12.  $\times$  27 Exp[-3], 24]

```

```

In[ ]:= NumberForm[Csch[1.04 Pi], 24]
NumberForm[Sech[1.04 Pi / 2]^2, 24]
NumberForm[(-2 × 1.52^2 + 5 × 1.52 - 2), 24]
NumberForm[Exp[-1.04 Pi], 24]
NumberForm[Sinh[0.52 Pi], 24]
NumberForm[1.44 × 0.14 × 0.9792, 24]
NumberForm[3 × 0.07 - 16.14 × 0.04 / 2.46 + 0.19, 24]

```

Upper bound

```

In[ ]:= NumberForm[4 (1.5 - 0.825)^2 - 1.7225, 24]
q1[x_] := 1 - 2 x^2
q2[x_] := x^2 - x
NumberForm[q1[1 / Sqrt[2]] + 6 q2[1], 24]
NumberForm[q1[1.01] + 6 q2[0.7 Coth[0.7]], 24]
NumberForm[q1[1.3] + 6 q2[Coth[1]], 24]
NumberForm[q1[1.6] + 6 q2[1.29 Coth[1.29]], 24]

```

7. Cut-off exponentials

```

In[ ]:= Plot[{Piecewise[{{0, t ≤ 0},
  {Sqrt[2 / (1 - Exp[-2 / 2])] Exp[-t], 0 < t ≤ 1 / 2}, {0, t > 1 / 2}}], Piecewise[
  {{0, t ≤ 0}, {Sqrt[2 / (1 - Exp[-2 / 3])] Exp[-t], 0 < t ≤ 1 / 3}, {0, t > 1 / 3}}],
  Piecewise[{{0, t ≤ 0}, {Sqrt[2 / (1 - Exp[-2 / 5])] Exp[-t], 0 < t ≤ 1 / 5},
    {0, t > 1 / 5}}}], {t, -.1, 0.1}], Exclusions → None,
  PlotStyle → {Black, {Black, Dashed}, {Black, Dotted}},
  Axes → True, AxesOrigin → {0, 0},
  PlotLegends → {"b=2", "b=3", "b=5"}]

```

```

In[ ]:= Plot[{Piecewise[{{0, t ≤ 0},
  {Sqrt[2 / (1 - Exp[-4 / 2])] Exp[-t], 0 < t ≤ 2 / 2}, {0, t > 2 / 2}}], Piecewise[
  {{0, t ≤ 0}, {Sqrt[2 / (1 - Exp[-4 / 3])] Exp[-t], 0 < t ≤ 2 / 3}, {0, t > 2 / 3}}],
  Piecewise[{{0, t ≤ 0}, {Sqrt[2 / (1 - Exp[-4 / 5])] Exp[-t], 0 < t ≤ 2 / 5},
    {0, t > 2 / 5}}}], {t, -.1, 1.1}], Exclusions → None,
  PlotStyle → {Black, {Black, Dashed}, {Black, Dotted}},
  Axes → True, AxesOrigin → {0, 0},
  PlotLegends → {"b=2", "b=3", "b=5"}]

```

```

In[ ]:= NumberForm[ArcCosh[0.5 + Sqrt[5] / 2], 24]

```

```

In[ ]:= Plot[{n, 3.5 / (2 - 4 / 3.7)}, {n, 4, 20},
  PlotRange → {Automatic, {0, 20}}, PlotLegends → "Expressions"]

```

8. The one-sided exponential

```

A8[η_, n_] := 2 η Tanh[η / 2] Csch[η / n] Exp[-η / n]
B8[η_, n_] := 2 η Coth[η / 2] Csch[η / n] Exp[η / n]
Plot[Evaluate@Table[A8[η, n], {n, 1, 7}],
  {η, 0.2, 36}, PlotRange → {Automatic, {0, 10}},
  PlotLabel → "The lower bounds", PlotLegends → Automatic, PlotLegends →
    LineLegend[Automatic, LegendFunction → (Framed[#1, FrameMargins → 0] &)]
Plot[Evaluate@Table[B8[η, n], {n, 1, 7}],
  {η, 0.2, 20}, PlotRange → {Automatic, {0, 100}},
  PlotLabel → "The upper bounds", PlotLegends → Automatic, PlotLegends →
    LineLegend[Automatic, LegendFunction → (Framed[#1, FrameMargins → 0] &)]
Plot[Evaluate@Table[B8[η, n] / A8[η, n], {n, 1, 7}],
  {η, 0.2, 20}, PlotRange → {Automatic, {0, 100}},
  PlotLabel → "The condition numbers", PlotLegends → Automatic, PlotLegends →
    LineLegend[Automatic, LegendFunction → (Framed[#1, FrameMargins → 0] &)]

```

The lower bound

```
NumberForm[1 + 0.5 Csch[0.5] - 0.5 - 0.5 Coth[0.5], 24]
```

The upper bound

```
NumberForm[3 - 2 Coth[2.] - 2 Csch[2], 24]
```

9. The two-sided exponential

```

f[x_] := x Csch[x]
A9[η_, n_] := n Tanh[η / 2] (f[η / n] - f[η])
B9[η_, n_] := n Coth[η / 2] (f[η / n] Cosh[η / n] + f[η])
Kappa9[η_, n_] := B9[η, n] / A9[η, n]
Plot[Evaluate@Table[A9[η, n], {n, 1, 7}],
  {η, 0.2, 36}, PlotRange → {Automatic, {0, 10}},
  PlotLabel → "The lower bounds", PlotLegends → Automatic, PlotLegends →
    LineLegend[Automatic, LegendFunction → (Framed[#1, FrameMargins → 0] &)]
Plot[Evaluate@Table[B9[η, n], {n, 1, 7}],
  {η, 0.2, 20}, PlotRange → {Automatic, {0, 100}},
  PlotLabel → "The upper bounds", PlotLegends → Automatic, PlotLegends →
    LineLegend[Automatic, LegendFunction → (Framed[#1, FrameMargins → 0] &)]
Plot[Evaluate@Table[Kappa9[η, n], {n, 1, 7}],
  {η, 0.2, 20}, PlotRange → {Automatic, {0, 100}},
  PlotLabel → "The condition numbers", PlotLegends → Automatic, PlotLegends →
    LineLegend[Automatic, LegendFunction → (Framed[#1, FrameMargins → 0] &)]

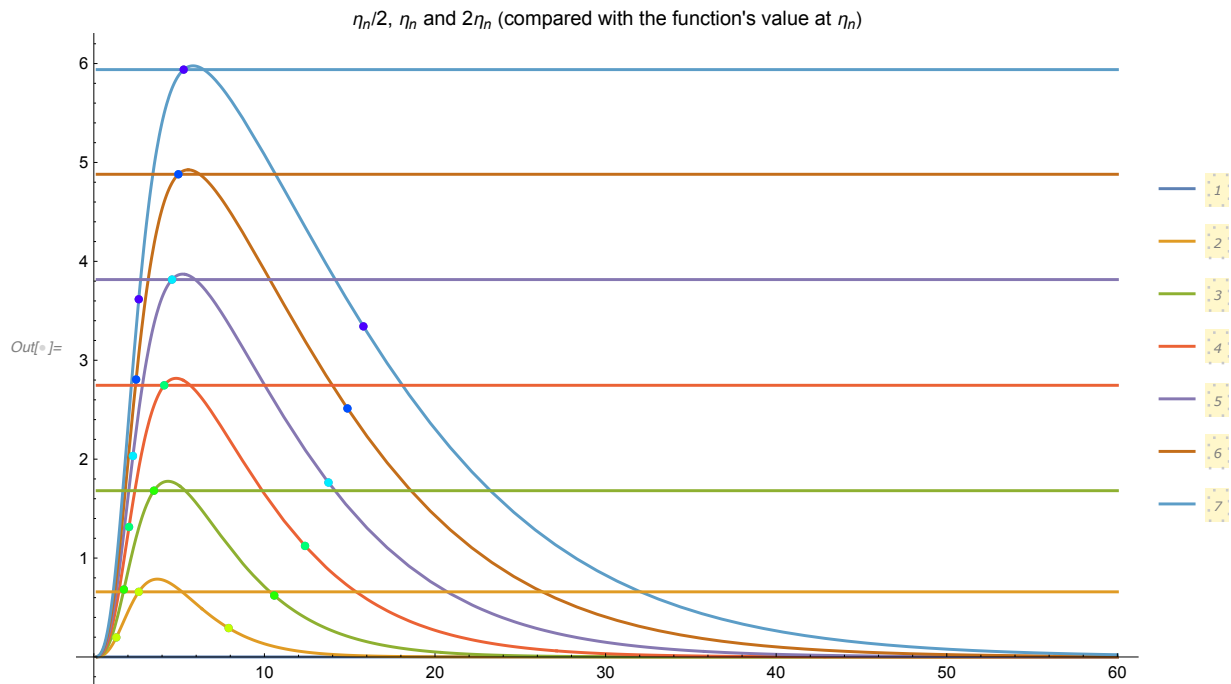
```

The idea of the cut-off

```

In[ ]:= K = 7;
p1 :=
  Plot[{Evaluate@Table[A9[t, n], {n, 1, K}]}], {t, 0.2, 60}, PlotLegends → Automatic,
  PlotLabel → " $\eta_n/2$ ,  $\eta_n$  and  $2\eta_n$  (compared with the function's value at  $\eta_n$ )",
  PlotRange → All, PlotLegends →
    LineLegend[Automatic, LegendFunction → (Framed[#1, FrameMargins → 0] &)]
p2 := Plot[{Evaluate@Table[A9[2 ArcCosh[n], n], {n, 1, K}]}],
  {t, 0.2, 60}, PlotRange → All]
p3 := ListPlot[Table[Style[{2 ArcCosh[n], A9[2 ArcCosh[n], n]}, Hue[-n (2 Pi) / K]],
  {n, 2, K, 1}], PlotStyle → PointSize[Medium]]
p4 := ListPlot[Table[Style[{ArcCosh[n], A9[ArcCosh[n], n]}, Hue[-n (2 Pi) / K]],
  {n, 2, K, 1}], PlotStyle → PointSize[Medium]]
p5 := ListPlot[
  Table[Style[{(2 + 1) 2 ArcCosh[n], A9[(2 + 1) 2 ArcCosh[n], n]}, Hue[-n (2 Pi) / K]],
  {n, 2, K, 1}], PlotStyle → PointSize[Medium]]
Show[p1, p2, p3, p4, p5, ImageSize → Large]
Clear[K];

```



The lower bound

The left cut-off

```
ln[*]:= NumberForm[2. ArcCosh[3] / 3, 24]
```

Interval update for small n

```
ln[*]:= A[4.34, 3] / 3. (*Point of comparison*)
ArcCosh[3.] (*Already proven cut-off*)
delta = 0.3;
list3 := Table[
  NumberForm[Tanh[(k + delta) / 2] (f[k / 3] - f[k + delta]) - A[4.34, 3] / 3., 24],
  {k, 1.7, 3.3, delta}]
list3

n = 4; (* Go through n = 4, .., 10 *)
Tanh[3.08 / 2] (1 - f[3.08])
A[2. ArcCosh[n], n] / n
```

The right cut-off

```
ln[*]:= NumberForm[2. rho[10], 24]
NumberForm[2. ArcCosh[10] ((1 - 1 / 10^2)^{1/2} ArcCosh[10] - 1 - 4 rho[10]^2), 24]
Table[
  NumberForm[Sqrt[n^2 - 1] Cosh[rho[n]] - n - Sinh[2 rho[n]] / 4., 24], {n, 3, 9}]
```

Interval update for small n

```
ln[*]:= NumberForm[A[4.34, 3] / 3., 24] (*Point of comparison*)
4 ArcCosh[3.] (*Already proven cut-off*)
delta = 0.1;
list4 :=
  Table[NumberForm[Tanh[(k + delta) / 2] (f[k / 3] - f[k + delta]) - A[4.347534, 3] / 3.,
    24], {k, 4.8, 7.1, delta}]
list4

4. ArcCosh[4]
NumberForm[f[4 / 3.] - f[5], 24]
NumberForm[f[5.] - f[4 ArcCosh[4]], 24]
NumberForm[A[4.82, 4] / 4, 24] (*Reference point for n = 4*)
```

```

In[ ]:=
n = 5; (* Go through n = 5, ..,10 *)
NumberForm[f[4 / 3.] - f[4 ArcCosh[n]], 24]
NumberForm[A[2. ArcCosh[n], n] / n, 24] (*Reference Value for n>4*)

```

Concavity result

```

In[ ]:=
Plot[1 / 2 Tanh[t]^2 + Tanh[t] / t - 1, {t, 0.4, 4 / 3}]

In[ ]:=
NumberForm[Sech[0.5]^2 - 4 (Sinh[2 × 0.5] - 2 × 0.5) / (2 × 0.5)^3, 24]

In[ ]:= delta = 0.01;
ls := Table[
  NumberForm[1 / 2 Tanh[k]^2 + Tanh[k] / (k + delta) - 1, 24], {k, 0.5, 1.2, delta}];
MemberQ[ls, Negative]
NumberForm[1 / 2 Tanh[1.606]^2 + Tanh[1.606] / 1.606 - 1, 24]
NumberForm[1 / 2 Tanh[1.61]^2 + Tanh[1.61] / 1.61 - 1, 24]
NumberForm[2 ArcCosh[3] / 2., 24]

```

The Upper Bound

Interval update for small n

```

Clear[n];
D[B[η, 3], η] /. η → 3.3
Table[D[B[η, n], η] /. η → 3.08, {n, 4, 10}]
D[B[η, 3], η] /. η → 4.8
Table[D[B[η, n], η] /. η → 4. / 3 n, {n, 4, 10}]

```

The condition number

```

NumberForm[Cosh[1.19] - 1 - 1.19 Csch[1.19], 24]
NumberForm[Cosh[1.2] - 1 - 1.2 Csch[1.2], 24]

NumberForm[Sinh[2 × 4 / 3] - 4. × 4^2 / 3^2 Coth[2 × 4 / 3], 24]

In[ ]:= NumberForm[ArcCosh[11.], 24]
NumberForm[3.08 - 8 / 3. Coth[4 / 3] Coth[3.08], 24]

NumberForm[1 / 2 × 3.08 Coth[3.08], 24]
NumberForm[4 / 3. Coth[4 / 3], 24]

```

The case n=3

```

H[η_] := η / 3 Coth[η / 3] + f[η]

```

```
Plot[{h[η], h'[η], h''[η]}, {η, 3.25, 4.85}, PlotLegends → {h, h', h''},
PlotRange → {Automatic, {-0.3, 2.5}}, Epilog → {Dotted, Red,
Line[{{3.3, -0.3}, {3.3, 2.5}}], Line[{{4.8, -0.3}, {4.8, 2.5}}], Dashed,
Line[{{3.25, -0.22}, {4.85, -0.22}}], Line[{{3.25, 0.22}, {4.85, 0.22}}]}]
```

Claim 1:

```
delta = 0.01;
ls1 :=
Table[NumberForm[1/3 Coth[k/3] + Csch[k] - (k + delta)/9 Csch[(k + delta)/3]^2 -
(k + delta) Csch[k + delta] Coth[k + delta] - 0.22, 24], {k, 3.3, 4.8, delta}]
MemberQ[ls1, Positive] (*Test whether the upper bound is correct*)
ls2 := Table[NumberForm[1/3 Coth[(k + delta)/3] + Csch[k + delta] -
k/9 Csch[k]^3 - (k) Csch[k] Coth[k] + 0.22, 24], {k, 3.3, 4.8, delta}];
MemberQ[ls2, Negative] (*Test whether the lower bound is correct*)
```

Claim 2:

```
ls3 :=
Table[NumberForm[k/3 Coth[k/3] + f[k + delta] - 1.6, 24], {k, 3.3, 4.8, delta}]
MemberQ[ls3, Negative] (*Test whether the lower bound is correct*)
```

Claim 3:

```
ln[ε] := NumberForm[3.3/2 - Coth[3.3], 24]
```

Final estimate:

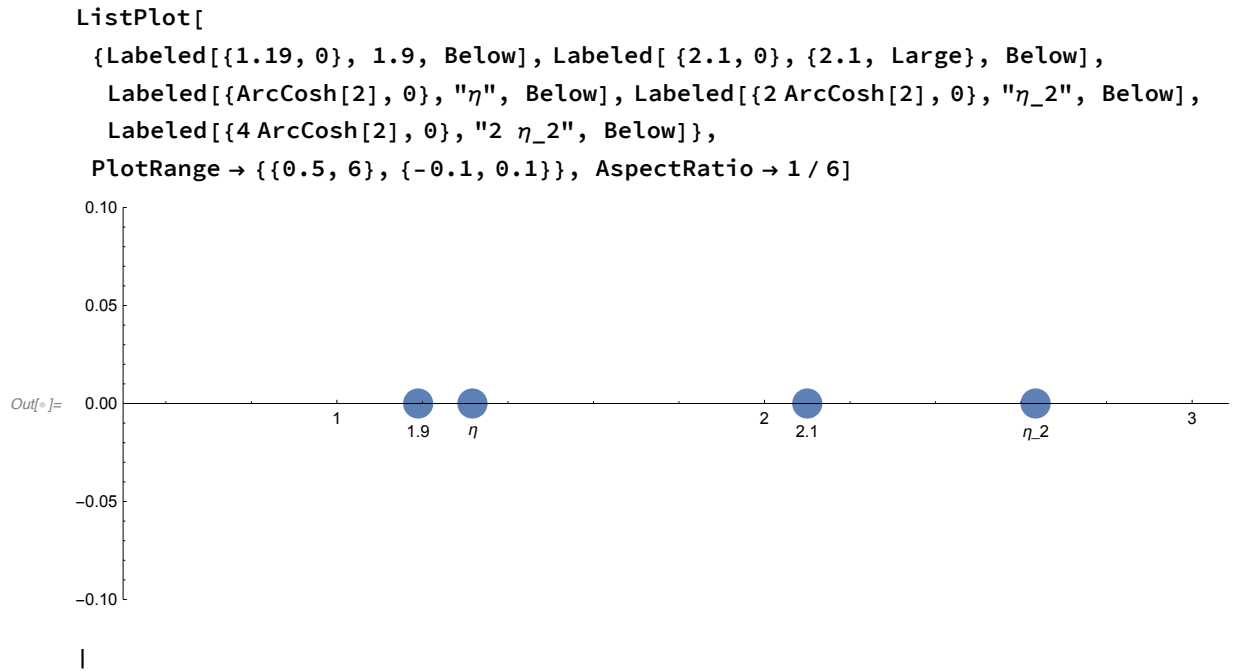
```
Plot[{3.2/9 Csch[η/3]^2 (η/3 Coth[η/3] - 1) + 3.2 Csch[η] * 0.64 - 0.22^2},
{η, 3.25, 4.85}, PlotRange → {Automatic, {-0.5, 0.5}},
Epilog → {Dotted, Red, Line[{{3.3, -0.3}, {3.3, 2.5}}],
Line[{{4.8, -0.3}, {4.8, 2.5}}], Dashed,
Line[{{3.25, -0.3}, {4.85, -0.3}}], Line[{{3.25, 0.3}, {4.85, 0.3}}]}]
delta = 0.1;
ls4 := Table[NumberForm[3.2/9 Csch[η/3]^2 (η/3 Coth[η/3] - 1) +
3.2 Csch[η - delta] * 0.64 - 0.22^2, 24], {η, 3.4, 4.9, delta}]
MemberQ[ls4, Negative]
```

```
ln[ε] := ls4
```

The case n = 2

The lower bound

```
ln[ε] := Plot[{Coth[x], x}, {x, 1.19, 1.2}]
```



```
In[ ]:= NumberForm[2 Sech[1.4] - 1., 24]
NumberForm[1 - 2.1 Tanh[2.1] (1 - 2 Sech[2.1]), 24]

In[ ]:= NumberForm[2 - 3 × 1.19 Tanh[1.19]^2, 24]
NumberForm[3 / 2 - Cosh[1.19], 24]
NumberForm[3 / 2 - 2.1 Csch[2.1] - 1.19 Tanh[1.19], 24]

In[ ]:= NumberForm[1 - Sech[1.85] - 1.85 Tanh[1.85] + 2 × 1.85 Sech[1.85] Tanh[1.85], 24]
NumberForm[1 - Sech[1.87] - 1.87 Tanh[1.87] + 2 × 1.87 Sech[1.87] Tanh[1.87], 24]
Plot[1 - Sech[t] - t Tanh[t] + 2 t Sech[t] Tanh[t], {t, 1.84, 1.88}]
```

The upper bound

```
D[B[η, 2], η] /. η → 3.
D[B[η, 2], η] /. η → 3.1
```

Three optimal points

The lower bound

```
NumberForm[rho[4.], 24]
Plot[(k + 2) (2^2 k - 1) + 2 k + 1, {k, 2, 30}]
NumberForm[Csch[rho[2]] / rho[2] - Csch[rho[2.]]^2, 24]
In[ ]:= NumberForm[2. ArcCosh[5] (Csch[rho[5]] / rho[5] - Csch[rho[5]]^2) -
  1 / 2 Sqrt[1 - 1 / (5^2 - 1)], 24]
```

```
ln[*]:= NumberForm[D[A[t, 2], t] /. t -> 2 ArcCosh[2.], 24]
NumberForm[D[A[t, 3], t] /. t -> 2 ArcCosh[3.], 24]
NumberForm[D[A[t, 4], t] /. t -> 2 ArcCosh[4.], 24]
```

The upper bound

```
ln[*]:= NumberForm[Coth[1.05] - 1.05 Coth[1.05]^2 + 3 / 7 * 1.05, 24]

ln[*]:= NumberForm[
  1. - rho[4] / 2 / Sqrt[4^2 - 1] - rho[4] Coth[4] + Sqrt[1 + 4^-2] 2 / 7 ArcCosh[4], 24]

ln[*]:= NumberForm[D[A[t, 3], t] /. t -> 3.9, 24]
NumberForm[D[B[t, 3], t] /. t -> 3.9, 24]
```

3.4 A note on energy minimization in dimension 2

Bibliographic Information

Faulhuber, M., Shafkulovska, I., Zlotnikov, I., A note on energy minimization in dimension 2, *submitted* (2023)

Submission date

29 June 2023

Contribution

Markus Faulhuber, Irina Shafkulovska, and Ilia Zlotnikov contributed equally to the idea, concept, development, and writing of the paper.

E-Print

arXiv:2306.16266

A NOTE ON ENERGY MINIMIZATION IN DIMENSION 2

MARKUS FAULHUBER, IRINA SHAFKULOVSKA, AND ILIA ZLOTNIKOV

ABSTRACT. Proving the *universal optimality of the hexagonal lattice* is one of the big open challenges of nowadays mathematics. We show that the hexagonal lattice outperforms certain “natural” classes of periodic configurations. Also, we rule out the option that the canonical non-lattice rival – the honeycomb – has lower energy than the hexagonal lattice at any scale.

1. PROBLEMS AND RESULTS

1.1. Energy minimization and universal optimality. For a periodic set Γ in $\mathbb{R}^d$ and a potential function $p : \mathbb{R}_+ \rightarrow \mathbb{R}$ we consider the quantity

$$E_p(\Gamma) = \lim_{R \rightarrow \infty} \frac{1}{|\Gamma \cap B_R^d(0)|} \sum_{\substack{\gamma, \gamma' \in \Gamma \cap B_R^d(0) \\ \gamma \neq \gamma'}} p(|\gamma - \gamma'|),$$

where $B_R^d(0)$ is the ball of radius R centered at 0 in $\mathbb{R}^d$. The above quantity is the energy of Γ . The density of Γ gives the average number of points per unit volume and is given by

$$\rho(\Gamma) = \lim_{R \rightarrow \infty} \frac{|\Gamma \cap B_R^d(0)|}{\text{vol}(B_R^d(0))}.$$

Throughout this work, we will assume that $\rho(\Gamma)$ is fixed. As we consider periodic configurations, we can write Γ as a finite union of translates of a lattice $\Lambda \subseteq \mathbb{R}^d$;

$$\Gamma = \bigcup_{j=1}^J \Lambda + x_j \quad \text{with} \quad x_j - x_k \notin \Lambda \quad \text{for all } j \neq k.$$

A lattice is always assumed to be full-rank. The energy is invariant under shifts;

$$E_p(\Gamma) = E_p(\Gamma + x), \quad x \in \mathbb{R}^d.$$

Due to the special structure of Γ we can re-write the energy in the following way

$$(1.1) \quad E_p(\Gamma) = \frac{1}{J} \sum_{j,k=1}^J \sum_{\lambda \in \Lambda \setminus \{x_j - x_k\}} p(|\lambda + x_j - x_k|).$$

The goal is to minimize (1.1) subject to Γ (density fixed). In accordance with Cohn and Kumar [7] (see also [8]) we introduce *universal optimality* in the following way.

Definition (Universal optimality). *Among all periodic configurations $\Gamma \subset \mathbb{R}^d$ of fixed density, we say Γ_α is a ground state of (1.1) if for a fixed $\alpha > 0$*

$$E_{\phi_\alpha}(\Gamma_\alpha) \leq E_{\phi_\alpha}(\Gamma), \quad \text{where} \quad \phi_\alpha(r) = e^{-\alpha r^2}, \quad r \in \mathbb{R}_+.$$

A periodic configuration Γ_0 is universally optimal if it is a ground state for all $\alpha > 0$;

$$E_{\phi_\alpha}(\Gamma_0) \leq E_{\phi_\alpha}(\Gamma), \quad \forall \alpha > 0.$$

NUHAG, FACULTY OF MATHEMATICS, UNIVERSITY OF VIENNA

OSKAR-MORGENSTERN-PLATZ 1, 1090 VIENNA, AUSTRIA

2020 *Mathematics Subject Classification.* 52C25, 74G65.

Key words and phrases. energy minimization, periodic configuration, universal optimality.

This work was supported by the Austrian Science Fund (FWF) grant P33217.

1.2. Results. In this work, we will prove that at all scales the hexagonal lattice Λ_2 has lower energy than the honeycomb structure. Given the fact that we already know that no other lattice can beat Λ_2 [17], the honeycomb structure was the most natural rival. We will also exclude the class of tensor periodic structures having lower energy than Λ_2 by reducing the dimension of the problem and exploiting universal optimality of the scaled integers [7]. Also, we pose the problem of distributing 3 points in the unit square (considered periodically) so that they minimize energy.

Theorem 1.1. *For any fixed $\alpha > 0$ and fixed density 1 (and hence any fixed density) the hexagonal lattice Λ_2 has a smaller energy than the honeycomb structure $\Gamma_\circ$;*

$$E_{\phi_\alpha}(\Lambda_2) < E_{\phi_\alpha}(\Gamma_\circ), \quad \forall \alpha > 0.$$

Theorem 1.2. *We denote the Cartesian tensor of 1-d periodic configurations by*

$$(1.2) \quad \Gamma_\times = \bigotimes_{l=1}^d \Gamma_l, \quad \Gamma_l \text{ a periodic configuration in } \mathbb{R}.$$

Denote by $\Gamma_\otimes$ the minimizing configuration of the energy among periodic configurations of type (1.2). Then, for any fixed $\alpha > 0$, we have $E_{\phi_\alpha}(\Gamma_\otimes) \leq E_{\phi_\alpha}(\Gamma_\times)$ with equality if and only if Γ_l is equispaced for all $l = 1, \dots, d$. In other words, $\Gamma_\otimes$ is a (cuboid-shaped) lattice. Moreover, for any fixed density $\rho > 0$, the cubic lattice $\rho^{-1/d}\mathbb{Z}^d$ is universally optimal among all periodic configurations of type (1.2).

Corollary 1.3. *In dimension 2, consider the set of lattices, the set of Cartesian tensor configurations and the set of unions of two shifted rectangular or hexagonal lattices. We denote these sets, respectively, by*

$$\begin{aligned} \Lambda &= \{\Lambda \subset \mathbb{R}^2 \mid \Lambda \text{ a lattice}\}, & \Gamma_\times &= \{\Gamma \subset \mathbb{R}^2 \mid \Gamma \text{ as in (1.2)}\} \\ \Gamma_x &= \{\Gamma \subset \mathbb{R}^2 \mid \Gamma = \Lambda \cup (\Lambda + x), \Lambda \in \{(a\mathbb{Z} \times a^{-1}\mathbb{Z}), \Lambda_2\}, a > 0, x \in \mathbb{R}^2\}, \end{aligned}$$

We denote the union of these sets by

$$\Xi = \Lambda \cup \Gamma_\times \cup \Gamma_x.$$

For any fixed $\alpha > 0$ and for any fixed density $\rho > 0$, the hexagonal lattice $\rho^{-1/2}\Lambda_2$ is the unique minimizer among periodic configurations from the set Ξ ;

$$E_{\phi_\alpha}(\rho^{-1/2}\Lambda_2) \leq E_{\phi_\alpha}(\Gamma), \quad \Gamma \in \Xi.$$

In other words, the hexagonal lattice Λ_2 is universally optimal in Ξ .

Remark. We also derive a minimality result for the hexagonal lattice in the set

$$\begin{aligned} \Gamma_{x,x^*} &= \{\Gamma \subset \mathbb{R}^2 \mid \Gamma = \Lambda \cup (\Lambda + x) \cup (\Lambda + x^*), \\ &\quad x \in [0, 1]^2, x^* = (1, 1) - x, \Lambda \in \{\mathbb{Z}^2, \Lambda_2\}\}. \end{aligned}$$

For Λ_2 the result is not too interesting as the optimal configuration is a scaled version of Λ_2 . For $\mathbb{Z}^2$, we prove that the pair $x_0 = (1/3, 1/3)$ and $x_0^* = (2/3, 2/3)$ as well as the pair $\tilde{x}_0 = (1/3, 2/3)$ and $\tilde{x}_0^* = (2/3, 1/3)$ are always critical points.

The optimality of Λ_2 among lattices already follows from the seminal work of Montgomery [17]. Corollary 1.3 provides more evidence for the widely conjectured universal optimality of the hexagonal lattice Λ_2 . Recently, it was noted by Coulangeon and Schürmann [12] that their proof of the hexagonal lattice being a local minimizer within the set of periodic configurations [11] unfortunately contains a non-repairable flaw. There is practically no doubt on the correctness of their statement, after all the hexagonal lattice is conjectured to be the unique *global* minimizer among periodic configurations. However, to the best of our knowledge, the status of universal optimality of the hexagonal lattice has been set back to the result of Montgomery [17] from 1988. In this note, we now extend the currently known result by enlarging the set of configurations within which the hexagonal lattice is optimal.

Universal optimality is a rare property and so far it was only proved to exist in dimensions 1, 8, and 24. Moreover, it provably fails for lattices in dimensions 3, 5, 6, and 7 (cf. [7, Sec. 9]). The reason is quite simple to state: to be universally optimal a lattice needs to solve the sphere

packing problem, which can be derived as a limiting case of energy minimization. Among lattices, the optimal sphere packing solutions in the named dimensions are known [10, Table 1.1]. By the Poisson summation formula, one can then prove that if a lattice is optimal for a parameter α (density fixed to 1), then its dual lattice is optimal for the parameter $1/\alpha$. As none of the optimal sphere packing lattices in the named dimensions are self-dual (up to scaling), they cannot be universally optimal. For a general potential, the circumstance that at different densities different configurations are energy-minimizing is called *phase transition*. Now, universal optimality includes the statement that phase transition cannot happen for completely monotone potentials.

The integer lattice Λ_1 , the hexagonal lattice Λ_2 (A_2 -root lattice), the D_4 root lattice Λ_4 , the E_8 root lattice Λ_8 and the Leech lattice Λ_{24} are self-dual (up to scaling). Here, we used the Λ_d notation of [10, Table 1.2]. It was proven by Cohn and Kumar that the (scaled) integer lattice is universally optimal in dimension 1 [7]. Recently, Cohn, Kumar, Miller, Radchenko, and Viazovska [8] proved universal optimality of Λ_8 and Λ_{24} . In dimension 4, numerical evidence was provided by Cohn, Kumar, and Schürmann [9] that Λ_4 may be universally optimal as well.

2. NOTATION

We are interested in the case when the potential is a Gaussian of width α , i.e.,

$$\phi_\alpha(r) = e^{-\alpha r^2}, \quad r \in \mathbb{R}_+, \alpha > 0.$$

The Fourier transform of a (suitable) function f is given by

$$\mathcal{F}f(\omega) = \int_{\mathbb{R}^d} f(t) e^{-2\pi i \omega \cdot t} dt.$$

The dot $\cdot$ denotes the Euclidean inner product of two vectors (always seen as column vectors). The Poisson summation formula for a lattice Λ with dual lattice $\Lambda^\perp$ reads

$$\sum_{\lambda \in \Lambda} f(\lambda + x) = \frac{1}{\text{vol}(\mathbb{R}^d/\Lambda)} \sum_{\lambda^\perp \in \Lambda^\perp} \mathcal{F}f(\lambda^\perp) e^{2\pi i \lambda^\perp \cdot x}.$$

Recall that the behavior of the Gaussian under the Fourier transform is

$$\mathcal{F}\phi_{\pi\alpha} = \alpha^{-1/2} \phi_{\pi/\alpha}.$$

Since the additional factor π is most often not relevant for our results we generally omit it. When it becomes important in a specific calculation we will explicitly write it. Moreover, fixing the density to 1 is not a restriction and we may choose any other fixed density. For brevity, we demonstrate this for a lattice $\Lambda = \rho^{-d/2} \Lambda_0 \subset \mathbb{R}^d$ of density ρ , where Λ_0 has density 1. For all $\alpha > 0$ we have

$$(2.1) \quad E_{\phi_\alpha}(\Lambda) = E_{\phi_\alpha}(\rho^{-d/2} \Lambda_0) = \sum_{\lambda \in \rho^{-d/2} \Lambda_0} e^{-\alpha |\lambda|^2} = \sum_{\lambda \in \Lambda_0} e^{-\frac{\alpha}{\rho^d} |\lambda|^2} = E_{\phi_{\alpha/\rho^d}}(\Lambda_0).$$

As the results shall hold for all $\alpha > 0$, we can choose any density we like and hide it in the exponent of the Gaussian. So, if a result holds for one fixed density and all α , then it already holds for all densities (and all α). In the sequel, we set $E(\Gamma) = E_{\phi_\alpha}(\Gamma)$ to simplify notation. Moreover, we will identify configurations which only differ by rotation or reflection (as they have the same energy). We write

$$\Gamma \cong \tilde{\Gamma} \quad \text{if there exists an orthogonal matrix } \mathcal{Q} \text{ such that } \Gamma = \mathcal{Q} \tilde{\Gamma}.$$

So, most often, explicit co-ordinates are not relevant but for some specific situations picking a certain lattice basis simplifies calculations. To this end, we note that we prefer the hexagonal lattice to be written as (recall that we use column vectors)

$$\Lambda_2 = \sqrt{\frac{2}{\sqrt{3}}} \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} \end{pmatrix} \mathbb{Z}^2.$$

3. AUXILIARY RESULTS

3.1. Jacobi theta function. We will need a result on Jacobi theta functions in the sequel. For argument $z \in \mathbb{C}$ and parameter $\tau \in \mathbb{H} = \{z \in \mathbb{C} \mid \Im(z) > 0\}$, denote the classical Jacobi theta-3 function by

$$\vartheta_3(z; \tau) = \sum_{k \in \mathbb{Z}} e^{\pi i \tau k^2} e^{2\pi i k z}.$$

It has a product representation known as the Jacobi triple product [18], [19];

$$\begin{aligned} \vartheta_3(z; \tau) &= \prod_{k \geq 1} (1 - e^{2k\pi i \tau}) \left(1 + e^{(2k-1)\pi i \tau} e^{2\pi i z}\right) \left(1 + e^{(2k-1)\pi i \tau} e^{-2\pi i z}\right) \\ &= \prod_{k \geq 1} (1 - e^{2k\pi i \tau}) \left(1 + 2e^{(2k-1)\pi i \tau} \cos(2\pi z) + e^{2(2k-1)\pi i \tau}\right). \end{aligned}$$

As we will only need the real-valued version of ϑ_3 , we set

$$\theta_3(x, \alpha) = \vartheta_3(x, i\alpha), \quad x \in \mathbb{R}, \alpha > 0.$$

Lemma 3.1. *Let $\alpha > 0$ be fixed. Then,*

$$\theta_3(0; \alpha) \geq \theta_3(x; \alpha) \geq \theta_3\left(\frac{1}{2}; \alpha\right),$$

with equality if and only if $x \in \mathbb{Z}$ for the first and $x \in \mathbb{Z} + \frac{1}{2}$ for the second inequality.

Proof. The proof follows easily from the product representation and the fact that $\cos(2\pi x)$ assumes its maximum for $x \in \mathbb{Z}$ and its minimum for $x \in \mathbb{Z} + \frac{1}{2}$. $\square$

This fact has already been noted and used by Janssen in his work on Gabor frame bounds [16]. A finer analysis of the behavior of $\theta_3(x; \alpha)$ has been carried out by Montgomery [17].

3.2. The result of Cohn and Kumar. In dimension 1, the considered energy minimization problem was completely settled by Cohn and Kumar [7]. Their result shows that at any density the scaled integers are universally optimal in dimension 1.

Theorem 3.2 (Cohn and Kumar, 2007). *Among all periodic configurations of the form $\Gamma = \bigcup_{k=1}^N \delta\mathbb{Z} + x_k$ of fixed density, with $x_k \in [0; \delta)$, $x_k \neq x_j$ for $k \neq j$*

$$E(\Gamma) = \frac{1}{N} \sum_{j,k=1}^N \sum_{\ell \in \mathbb{Z}} \phi_\alpha(|\ell + x_k - x_j|)$$

is minimal if and only if $x_k = \frac{k-1}{N} + x$ for a global shift $x \in \mathbb{R}$, i.e., Γ is equispaced.

The term *universally optimal* refers to the fact that the result for the Gaussian function ϕ_α implies that the same holds for any completely monotone function of squared distance. A function $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is completely monotone if and only if

$$(-1)^k f^{(k)}(r) \geq 0 \quad \text{for all } k \in \mathbb{N}.$$

By the Bernstein-Widder theorem ([2], [20]) any completely monotone function can also be written as the Laplace transform of a non-negative Borel measure μ .

Theorem 3.3 (Bernstein-Widder, 1928/1941). *Let f be completely monotone. Then there exists a non-negative Borel measure μ_f such that*

$$f(r) = \int_0^\infty e^{-\alpha r} d\mu_f(\alpha).$$

It follows that, for any dimension d , if a configuration is optimal for the Gaussian ϕ_α for all $\alpha > 0$, then it is also optimal for all completely monotone potential functions p of squared distance:

$$\frac{1}{N} \sum_{j,k=1}^N \sum_{\lambda \in \Lambda \setminus \{0\}} p(|\lambda + x_j - x_k|^2) = \frac{1}{N} \sum_{j,k=1}^N \sum_{\lambda \in \Lambda \setminus \{0\}} \int_0^\infty e^{-\alpha |\lambda + x_j - x_k|^2} d\mu_p(\alpha)$$

$$= \frac{1}{N} \int_0^\infty \left(\sum_{j,k=1}^N \sum_{\lambda \in \Lambda \setminus \{0\}} e^{-\alpha|\lambda + x_j - x_k|^2} \right) d\mu_p(\alpha).$$

Note that, by Bernstein [2, Thm. 9.16], the above integral is indeed convergent and the interchange of summation and integration is justified by monotone convergence: the potential is an increasing limit of weighted sums of Gaussians (cf. [8, Sec. 1]).

3.3. The result of Montgomery. In dimension 2, the energy minimization problem is solved among *lattices* by the result of Montgomery [17]. To the best of our knowledge, this is the best available result in dimension 2 after the appearance of [12]. Before stating the result, we introduce the following (theta-like) function:

$$(3.1) \quad f_\Gamma(x; \alpha) = \sum_{\gamma \in \Gamma} \phi_\alpha(|x - \gamma|) = \sum_{\gamma \in \Gamma} e^{-\alpha|x - \gamma|^2}, \quad \alpha > 0.$$

We will often write $f_\Gamma(x)$ instead of $f_\Gamma(x; \alpha)$ if we do not need a specific α .

Theorem 3.4 (Montgomery, 1988). *For all $\alpha > 0$ and among all lattices of any fixed density, the following inequality (for lattice theta functions) holds*

$$f_\Lambda(0) \geq f_{\Lambda_2}(0), \quad \forall \alpha > 0,$$

with equality if and only if Λ is the hexagonal lattice Λ_2 .

As the energy of a lattice Λ can be written as $E_{\phi_\alpha}(\Lambda) = f_\Lambda(0) - 1$, we see that Montgomery's result implies energy minimization for the Gaussian. By Theorem 3.3, the same holds for completely monotone potentials of squared distance.

4. PROOFS

4.1. Proof of Theorem 1.2.

Proof of Theorem 1.2. If $\Gamma_\times$ is as in (1.2), i.e., it is the Cartesian product of 1-dimensional periodic configurations Γ_l , $l = 1, \dots, d$, then it can be described as

$$\Gamma_\times = \bigcup_{l=1}^d \bigcup_{j_l=1}^{J_l} \delta \mathbb{Z}^d + (x_{1,j_1}, x_{2,j_2}, \dots, x_{d,j_d}),$$

where $\delta = \text{diag}(\delta_1, \dots, \delta_d)$ denotes the diagonal matrix with entries $\delta_1, \dots, \delta_d$. Using the factorization properties of the exponential function, the energy (1.1) for $\Gamma_\times$ can be written as follows:

$$\begin{aligned} E(\Gamma_\times) &= \frac{1}{\prod_{m=1}^d J_m} \sum_{l=1}^d \sum_{j_l, k_l=1}^{J_l} \sum_{\lambda \in \mathbb{Z}^d \setminus \{0\}} \phi_\alpha(|\delta \lambda + (x_{1,j_1} - x_{1,k_1}, \dots, x_{d,j_d} - x_{d,k_d})|) \\ &= \frac{1}{\prod_{m=1}^d J_m} \sum_{l=1}^d \sum_{j_l, k_l=1}^{J_l} \sum_{\lambda \in \mathbb{Z}^d \setminus \{0\}} \prod_{m=1}^d \phi_\alpha(|\delta_m \lambda_m + x_{m,j_m} - x_{m,k_m}|) \\ &= \prod_{m=1}^d \left(\frac{1}{J_m} \sum_{j_m, k_m=1}^{J_m} \sum_{\lambda \in \mathbb{Z} \setminus \{0\}} \phi_\alpha(|\delta_m \lambda + x_{m,j_m} - x_{m,k_m}|) \right). \end{aligned}$$

This shows that the problem reduces to finding the optimal point configuration in each coordinate, i.e., we can refer to the one-dimensional solution. By the result of Cohn and Kumar [7] (Theorem 3.2) it follows that the optimal solutions need to be equispaced in each of the Γ_l , i.e., it has to be a cuboid-shaped lattice of the form

$$\Lambda_\beta = \beta_1 \mathbb{Z} \times \dots \times \beta_d \mathbb{Z},$$

with $\beta_m = \delta_m/J_m$, $m = 1, \dots, d$. For the lattice Λ_β , the energy for the Gaussian potential can be written as a product of Jacobi theta functions:

$$E(\Lambda_\beta) = \prod_{j=1}^d \sum_{k \in \mathbb{Z}} e^{-\alpha \beta_j^2 k^2} - 1 = \prod_{j=1}^d \theta_3 \left(0; \frac{\alpha \beta_j^2}{\pi} \right) - 1, \quad \alpha > 0.$$

It was also proven by Montgomery [17] (see also [6], [15]) that the above product is minimal if and only if $\beta_1 = \dots = \beta_d$, when $\prod_{k=1}^d \beta_k$ is fixed. $\square$

4.2. The union of 2 lattices. Lattices are the simplest periodic structures. The next natural step is to consider the union of two lattices in $\mathbb{R}^d$;

$$\Gamma = \Lambda \cup (\Lambda + x), \quad x \notin \Lambda.$$

This already turns out to be tricky very quickly as, in general, it is awfully difficult to determine the shift (which depends on $(d^2 + d - 2)/2$ lattice parameters and α) of the second lattice which minimizes the energy of the configuration.

4.2.1. The union of two cuboid-shaped lattices. The simplest case for the union of 2 lattices is the following: let $\Lambda = \delta \mathbb{Z}^d$, where $\delta = \text{diag}(\delta_1, \dots, \delta_d)$. Consider the periodic configuration of the following form:

$$(4.1) \quad \Gamma_x = \Lambda \cup (\Lambda + \delta x), \quad x \notin \mathbb{Z}^d.$$

Then, as we will show, the energy of the system can be written as

$$(4.2) \quad E(\Gamma_x) = E(\Lambda) + f_\Lambda(x),$$

where $f_\Lambda(x)$ is defined as in (3.1). Equation (4.2) actually holds for any lattice, not only cuboid-shaped ones. For a lattice of type (4.1), it follows from the behavior of f_Λ and Lemma 3.1 (cf. [17, Lemma 1]) that

$$E(\Gamma_x) \geq E(\Lambda) + f_\Lambda(\delta \frac{1}{2}), \quad \frac{1}{2} = (\frac{1}{2}, \dots, \frac{1}{2})$$

with equality being attained if and only if $x \in \frac{1}{2} + \mathbb{Z}^d$, as we will now show.

Proposition 4.1. *Let Γ_x be defined as in (4.1). The energy of Γ_x is given by (4.2) and is minimal if and only if $x \in \frac{1}{2} + \mathbb{Z}^d$. Hence, the minimizer is a lattice.*

Proof. First, allow Λ to be any lattice. Set $\Gamma_x = \Lambda \cup (\Lambda + x)$, $x_1 = 0$ and $x_2 = x \notin \Lambda$. Note that $f_\Lambda(x) = f_\Lambda(-x)$ for any lattice Λ . Then, we can write the energy as

$$\begin{aligned} E(\Gamma_x) &= \frac{1}{2} \sum_{\gamma, \gamma' \in \Gamma_x} e^{-\alpha |\gamma - \gamma'|^2} - 1 = \frac{1}{2} \sum_{j,k=1}^2 \sum_{\lambda \in \Lambda} e^{-\alpha |\lambda + x_j - x_k|^2} - 1 \\ &= \sum_{\lambda \in \Lambda} e^{-\alpha |\lambda|^2} - 1 + \sum_{\lambda \in \Lambda} e^{-\alpha |\lambda + x|^2} = E(\Lambda) + \sum_{\lambda \in \Lambda} e^{-\alpha |\lambda + x|^2} \\ &= E(\Lambda) + f_\Lambda(x). \end{aligned}$$

Due to the special structure of $\Lambda = \delta \mathbb{Z}^d$, we can write the energy as

$$E(\Gamma_x) = E(\Lambda) + \prod_{k=1}^d \sum_{n \in \mathbb{Z}} e^{-\alpha \delta_k^2 (n + x_k)^2} = E(\Lambda) + \prod_{k=1}^d \sqrt{\frac{\pi}{\alpha \delta_k^2}} \theta_3 \left(x_k; \frac{\pi}{\alpha \delta_k^2} \right).$$

For the second equality, we used the Poisson summation formula. Now, we can use Lemma 3.1 to see that $E(\Gamma_x)$ is minimal if and only if $x \in \frac{1}{2} + \mathbb{Z}^d$. $\square$

4.2.2. *The union of two hexagonal lattices.* As the hexagonal lattice is outstanding in the set of 2-dimensional lattices, we consider periodic configurations of the form

$$(4.3) \quad \Gamma_2(x) = \Lambda_2 \cup (\Lambda_2 + x).$$

Shifting Λ_2 by a deep hole x_0 gives the so-called honeycomb structure (see Figure 1)

$$\Gamma_{\circ} = \Lambda_2 \cup (\Lambda_2 + x_0).$$

Coordinates of a *deep hole* (cf. [10, Chap. 1]) in the hexagonal lattice are given by

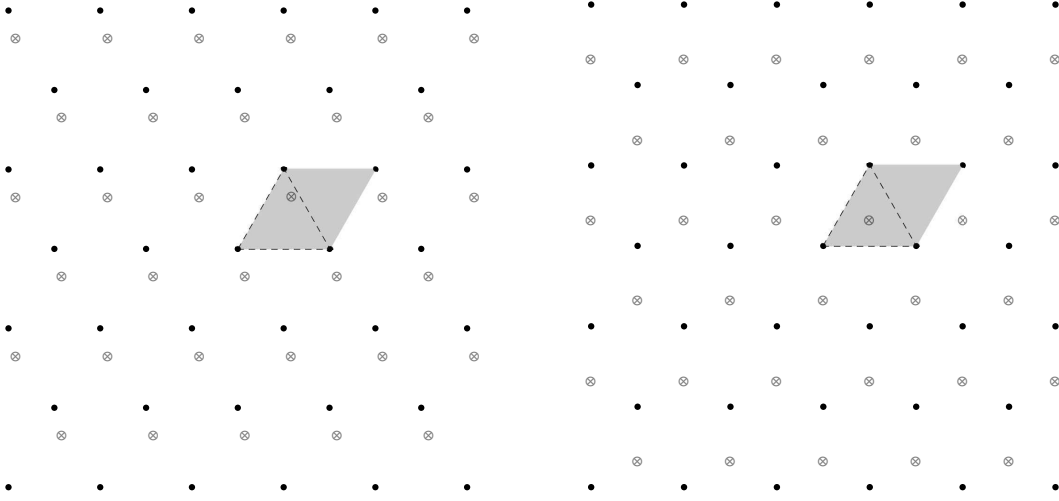


Figure 1. Union of two hexagonal lattices by an arbitrary shift (left) and a shift by a deep hole (right). A deep hole is the center of a fundamental triangle and we have two deep holes in a fundamental parallelogram. The resulting configuration is known as the honeycomb structure, which is not a lattice.

$$(4.4) \quad x_0 = \sqrt{\frac{2}{\sqrt{3}}} \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} k + \frac{m}{3} \\ l + \frac{m}{3} \end{pmatrix}, \quad k, l \in \mathbb{Z}, m \in \{1, 2\}.$$

Proposition 4.2. *Among all configurations $\Gamma_2(x)$ of the form (4.3), the honeycomb structure $\Gamma_{\circ}$ is the unique minimizer of $E(\Gamma_2(x))$.*

Proof. The proof requires a result of Baernstein [1] on the minimal temperature of the heat kernel on the hexagonal flat torus $\mathbb{R}^2/\Lambda_2$. It states that

$$f_{\Lambda_2}(x) = \sum_{\lambda \in \Lambda_2} e^{-\alpha|\lambda+x|^2} \geq \sum_{\lambda \in \Lambda_2} e^{-\alpha|\lambda+x_0|^2} = f_{\Lambda_2}(x_0),$$

with equality if and only if x is another deep hole of Λ_2 . The result now follows as

$$E(\Gamma_2(x)) = E(\Lambda_2) + f_{\Lambda_2}(x) \geq E(\Lambda_2) + f_{\Lambda_2}(x_0) = E(\Gamma_{\circ}).$$

□

We remark that the hexagonal lattice cannot be written as a union of two shifted hexagonal lattices. Regarding the conjectured universal optimality of the hexagonal lattice, the honeycomb $\Gamma_{\circ}$ is probably the most outstanding rival in the set of periodic configurations which is not a lattice. We will now prove that, scaled to unit density, the honeycomb structure cannot beat the hexagonal lattice Λ_2 .

4.3. Proof of Theorem 1.1. We start by proving properties of two auxiliary functions. These will be used for refined convexity estimates of the energy.

$$(4.5) \quad g(s) = e^{-s} - \frac{1}{2}e^{-2s} - \frac{1}{2}e^{-2/3s},$$

$$(4.6) \quad h(s) = e^{-s} - \frac{1}{4}e^{-s/2} - \frac{3}{4}e^{-3/2s}.$$

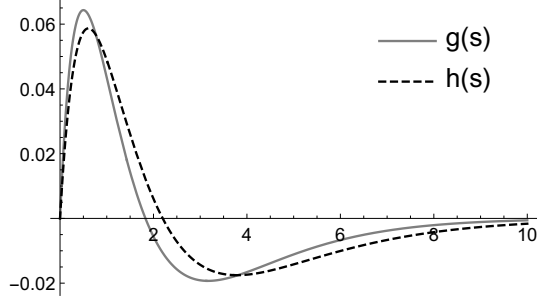


Figure 2. The auxiliary functions $g(s)$ and $h(s)$.

Lemma 4.3. *The following estimates hold:*

$$g(s) < 0 \quad \text{for } 2 < s < \infty \quad \text{and} \quad h(s) < 0 \quad \text{for } 6 < s < \infty.$$

Proof. We start with the inequality for g . Setting $t = e^{-s/3}$, it suffices to verify

$$-t^6 + 2t^3 - t^2 < 0 \quad \text{for } t \in (0, e^{-2/3}).$$

Note that for $0 < t \leq e^{-2/3} < 1$ we have

$$-1 < t^3 + t^2 + t - 1 < e^{-2} + e^{-4/3} + e^{-2/3} - 1 < -0.087.$$

Therefore, since

$$-t^6 + 2t^3 - t^2 = -t^2(t-1)(t^3 + t^2 + t - 1),$$

we get the required inequality for g . For the function h , we set $t = e^{-s/2}$ and verify

$$-3t^3 + 4t^2 - t < 0 \quad \text{for } t \in (0, e^{-3}).$$

Since $(0, e^{-3}) \subset (0, 1/3)$ we obtain

$$-3t^3 + 4t^2 - t = -t(1-t)(1-3t) < 0.$$

□

Proof of Theorem 1.1. This time we write the energies with π in the exponent;

$$\begin{aligned} E_{\phi_{\pi\alpha}}(\Lambda_2) &= \sum_{k,l \in \mathbb{Z}} e^{-\frac{2}{\sqrt{3}}\pi\alpha(k^2+kl+l^2)} - 1, \\ E_{\phi_{\pi\alpha}}(\Gamma_0) &= E_{\phi_{\pi\alpha}}(\sqrt{2}\Lambda_2) + \sum_{\lambda \in \Lambda_2} e^{-\pi(2\alpha)|\lambda+x_0|^2} \\ &= E_{\phi_{\pi\alpha}}(\sqrt{2}\Lambda_2) + \frac{1}{2\alpha} \sum_{\lambda \in \Lambda_2} e^{-\frac{\pi}{2\alpha}|\lambda|^2} e^{2\pi i\sigma(\lambda, x_0)}. \end{aligned}$$

For the last equality, we used the symplectic version of the Poisson summation formula. This is a nice bookkeeping trick in dimension 2 (see [4, Sec. 1.1], [5, App. B], [14, Sec. 4.4]) as the dual lattice is replaced by a 90 degrees rotated version called the adjoint lattice Λ° . In particular, in dimension 2 any lattice (of density 1) is its own adjoint (up to scaling for density other than 1). This also requires a change of variables in the complex exponential. The Euclidean inner product $\lambda \cdot x$ is replaced by the skew-symmetric form $\sigma(\lambda, x)$;

$$\sigma(\lambda, x) = \lambda \cdot Jx, \quad \text{where } \lambda, x \in \mathbb{R}^2, \quad J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \Lambda^\circ = J\Lambda^\perp.$$

Now we can use the cubic AGM of Borwein and Borwein [3] to write the (symplectic) Fourier series in terms of the energy on scaled hexagonal lattices. For the sake of self-containment, we present necessary parts from [3]. For $|q| < 1$, $\zeta_3^3 = 1$, $\zeta_3 \neq 1$ set

$$a(q) = \sum_{k,l \in \mathbb{Z}} q^{k^2+kl+l^2}, \quad b(q) = \sum_{k,l \in \mathbb{Z}} q^{k^2+kl+l^2} \zeta_3^{k-l}, \quad \text{then} \quad 3a(q^3) = a(q) + 2b(q).$$

At this point, we also refer to [14] for some geometric explanations. By using the above equation, which is the arithmetic part of the cubic AGM, we get

$$\begin{aligned} E_{\phi_{\pi\alpha}}(\Gamma_{\circ}) &= E_{\phi_{\pi\alpha}}(\sqrt{2}\Lambda_2) + \frac{1}{4\alpha} (3(E_{\phi_{3\pi/(2\alpha)}}(\Lambda_2) + 1) - (E_{\phi_{\pi/(2\alpha)}}(\Lambda_2) + 1)) \\ &= E_{\phi_{\pi\alpha}}(\sqrt{2}\Lambda_2) + \frac{1}{2} ((E_{\phi_{(2\alpha)\pi/3}}(\Lambda_2) + 1) - (E_{\phi_{(2\alpha)\pi}}(\Lambda_2) + 1)) \\ &= E_{\phi_{\pi\alpha}}(\sqrt{2}\Lambda_2) + \frac{1}{2} (E_{\phi_{\alpha\pi}}(\sqrt{2/3}\Lambda_2) - E_{\phi_{\alpha\pi}}(\sqrt{2}\Lambda_2)) \\ &= \frac{1}{2} (E_{\phi_{\pi\alpha}}(\sqrt{2}\Lambda_2) + E_{\phi_{\alpha\pi}}(\sqrt{2/3}\Lambda_2)), \end{aligned}$$

where we used the Poisson summation formula in the second step and the “density trick” from (2.1) in the third step. We need to show that, for all $\alpha > 0$, we have

$$(4.7) \quad E_{\phi_{\alpha\pi}}(\Lambda_2) < \frac{1}{2} (E_{\phi_{\pi\alpha}}(\sqrt{2}\Lambda_2) + E_{\phi_{\alpha\pi}}(\sqrt{2/3}\Lambda_2)),$$

which by the Poisson summation formula is equivalent to

$$(4.8) \quad E_{\phi_{\pi/\alpha}}(\Lambda_2) < \frac{1}{2} \left(\frac{1}{2} E_{\phi_{\pi/\alpha}}(\sqrt{1/2}\Lambda_2) + \frac{3}{2} E_{\phi_{\pi/\alpha}}(\sqrt{3/2}\Lambda_2) \right).$$

So, the hexagonal lattice Λ_2 of density 1 has smaller energy than a weighted average of its scaled versions of density $1/2$ and $3/2$. This is almost a consequence of the convexity of the energy in α , but not quite. Writing out the formulas, one sees that we need a kind of convexity result in $1/\alpha$. We proceed by comparing terms where the quadratic form $k^2 + kl + l^2$ is constant in inequalities (4.7) and (4.8). Observe that the inequality in (4.7) can be rewritten as

$$(4.9) \quad \sum_{(k,l) \in \mathbb{Z}^2 \setminus \{(0,0)\}} g\left(\frac{2\pi\alpha}{\sqrt{3}}(k^2 + kl + l^2)\right) < 0,$$

where f is defined as in (4.5) and the inequality in (4.8) is equivalent to

$$(4.10) \quad \sum_{(k,l) \in \mathbb{Z}^2 \setminus \{(0,0)\}} h\left(\frac{2\pi}{\sqrt{3}\alpha}(k^2 + kl + l^2)\right) < 0,$$

where g is defined as in (4.6). It clearly suffices to prove that each summand is negative to conclude the claim. We make two case distinctions.

Case 1: $\alpha \geq \frac{\sqrt{3}}{\pi}$. In this case, for all $(k, l) \in \mathbb{Z}^2 \setminus \{(0,0)\}$ we have

$$\frac{2\pi\alpha}{\sqrt{3}}(k^2 + kl + l^2) \geq \frac{2\pi}{\sqrt{3}} \cdot \frac{\sqrt{3}}{\pi} = 2,$$

so by Lemma 4.3, the inequality (4.9) holds for all $\alpha \geq \frac{\sqrt{3}}{\pi}$.

Case 2: $\alpha \leq \frac{\sqrt{3}}{\pi}$. In this case,

$$\frac{2\pi}{\sqrt{3}\alpha}(k^2 + kl + l^2) \geq \frac{2\pi}{\sqrt{3}} \cdot \frac{\pi}{\sqrt{3}} = \frac{2\pi^2}{3} > 6,$$

so by Lemma 4.3, the inequality (4.10) holds for all $\alpha \leq \frac{\sqrt{3}}{\pi}$. In combination, this means that (4.7) and (4.8) hold for all $\alpha > 0$. $\square$

Proof of Corollary 1.3. The proof is just accumulating existing and derived results.

- The set Λ : by Theorem 3.4, Λ_2 is optimal in Λ for all $\alpha > 0$.

- The set $\Gamma_{\times}$: by Theorem 1.2, which is a consequence of Theorem 3.2, we know that the optimal structure in the set $\Gamma_{\times}$ needs to be a rectangular lattice. Among these, the square lattice yields

the smallest energy ([17], see also [15]) and by Theorem 3.4 the hexagonal lattice has smaller energy.

- The set Γ_2 : let Γ_x be the union of two rectangular lattices as in (4.1). For $x_0 = \delta \frac{1}{2}$, it is not hard to see that Γ_{x_0} is a lattice

$$\Gamma_{x_0} = \delta \mathbb{Z}^2 \cup \delta \left(\mathbb{Z}^2 + \frac{1}{2} \right) = \begin{pmatrix} \delta_1 & \frac{\delta_1}{2} \\ 0 & \frac{\delta_2}{2} \end{pmatrix} \mathbb{Z}^2.$$

By Theorem 3.4 the hexagonal lattice minimizes energy among lattices and by Proposition 4.1, the minimizing configuration Γ_x needs to be the lattice Γ_{x_0} . The hexagonal lattice can be written as a union of two rectangular lattices with $\delta_2/\delta_1 = \sqrt{3}$. Now let $\Gamma_2(x)$ be the union of two hexagonal lattices. By Proposition 4.2, the optimal structure is the honeycomb $\Gamma_\circ$ and by Theorem 1.1, it has higher energy than the hexagonal lattice Λ_2 for all $\alpha > 0$. $\square$

4.4. Union of 3 lattices. Analyzing the union of 3 arbitrary lattices is a too challenging endeavor for this note. So we will focus on some particular cases. Note that the union of 3 hexagonal lattices, i.e., $\Lambda_2 \cup (\Lambda_2 + x) \cup (\Lambda_2 + y)$ does not produce too many interesting results: the optimal shifts are the deep holes $x = x_0$ and $y = -x_0$ defined by (4.4). This means that the optimal configuration of 3 shifted hexagonal lattices is again a hexagonal lattice (of thrice the density). Therefore, we will concentrate on the union of 3 square lattices. In this case, the problem becomes surprisingly challenging. The energy of the set

$$(4.11) \quad \Gamma_{x,y} = \mathbb{Z}^2 \cup (\mathbb{Z}^2 + x) \cup (\mathbb{Z}^2 + y)$$

is, for $x = (x_1, x_2)$, $y = (y_1, y_2)$ and $f(x) = f_{\mathbb{Z}}(x; \alpha)$ ($f_{\mathbb{Z}}$ defined by (3.1)), given by

$$E_{\phi_{\alpha\pi}}(\Gamma_{x,y}) = f(0)^2 + \frac{2}{3} (f(x_1)f(x_2) + f(y_1)f(y_2) + f(x_1 - y_1)f(x_2 - y_2)) - 1.$$

Thus, it suffices to determine the minimizers $(x, y) \in [0, 1]^2$ of the objective function

$$F(x, y) = f(x_1)f(x_2) + f(y_1)f(y_2) + f(x_1 - y_1)f(x_2 - y_2).$$

The critical points of F are the solutions to the system of equations

$$\begin{aligned} f'(x_1)f(x_2) &= -f'(x_1 - y_1)f(x_2 - y_2), & f(x_1)f'(x_2) &= -f(x_1 - y_1)f'(x_2 - y_2), \\ f'(y_1)f(y_2) &= f'(x_1 - y_1)f(x_2 - y_2), & f(y_1)f'(y_2) &= f(x_1 - y_1)f'(x_2 - y_2). \end{aligned}$$

Adding the two left as well as the two right equations, we see that

$$(4.12) \quad f'(y_1)f(y_2) = -f'(x_1)f(x_2) \quad \text{and} \quad f(y_1)f'(y_2) = -f(x_1)f'(x_2).$$

Since $f(x) = f(-x)$, we have $f'(x) = -f'(-x)$ and a trivial solution of the equations in (4.12) is clearly given by $x = -y$. This prompted us to restrict to configurations of type $\Gamma_{x,-x}$. Note that the solutions of (4.12) are periodic with respect to $\mathbb{Z}^2$. So we may assume $x \in [0, 1]^2$ and to stay in the unit square, instead of $-x$, we use the solution $x^* = (1, 1) - x$. The remaining restrictions are then given by

$$(4.13) \quad f'(x_1)f(x_2) + f'(2x_1)f(2x_2) = 0 \quad \text{and} \quad f(x_1)f'(x_2) + f(2x_1)f'(2x_2) = 0.$$

Due to symmetry and periodicity, solutions are given by the pairs

$$(4.14) \quad x_0 = \left(\frac{1}{3}, \frac{1}{3}\right) \text{ and } x_0^* = \left(\frac{2}{3}, \frac{2}{3}\right) \quad \text{or by} \quad \tilde{x}_0 = \left(\frac{1}{3}, \frac{2}{3}\right) \text{ and } \tilde{x}_0^* = \left(\frac{2}{3}, \frac{1}{3}\right).$$

Thus, it is worthwhile to examine the energy of

$$\Gamma_{x_0, x_0^*} = \mathbb{Z}^2 \cup (x_0 + \mathbb{Z}^2) \cup (x_0^* + \mathbb{Z}^2) \cong \Gamma_{\tilde{x}_0, \tilde{x}_0^*}.$$

4.4.1. Open problem. Is it true that (by periodicity $x, y \in [0, 1]^2$) the configuration minimizing $E(\Gamma_{x,y})$ in (4.11) shows the symmetry $y = x^* = (1, 1) - x$? Assume that it indeed does. Then the pairs x_0, x_0^* , and $\tilde{x}_0, \tilde{x}_0^*$ always yield critical configurations. What are other critical configurations? Take, e.g., $\alpha = 3.5$, $z = (1/4, 1/2)$ and $z^* = (1, 1) - z$ (note that Γ_{z,z^*} is not critical). Using *Mathematica* [21] and its built-in function `EllipticTheta` (which can be evaluated to any precision), we get

$$E_{\phi_{3.5\pi}}(\Gamma_{z,z^*}) \approx 0.17159 < 0.18279 \approx E_{\phi_{3.5\pi}}(\Gamma_{x_0, x_0^*}).$$

4.4.2. *Asymptotic solution for small parameters.* We are now assuming that $y = x^*$. If α tends to 0, then we can prove that the only solutions in $[0, 1]^2 \times [0, 1]^2$ are indeed the pairs (x_0, x_0^*) or $(\tilde{x}_0, \tilde{x}_0^*)$. We start with an auxiliary result for $f(x) = f_{\mathbb{Z}}(x; \alpha)$.

Lemma 4.4. *We define the following functions for $\alpha > 0$.*

$$g(\alpha) = \begin{cases} \alpha^{-\frac{1}{2}} \left(1 - \frac{22002}{11000} e^{-\frac{\pi}{\alpha}}\right) & \alpha < 1, \\ 2e^{-\frac{\pi\alpha}{4}} & 1 \leq \alpha. \end{cases} \quad h(\alpha) = \begin{cases} \alpha^{-\frac{1}{2}} \left(1 + \frac{22002}{11000} e^{-\frac{\pi}{\alpha}}\right) & \alpha < 1, \\ 1 + \frac{22002}{11000} e^{-\pi\alpha} & 1 \leq \alpha. \end{cases}$$

Then, for any $x \in \mathbb{R}$ the following inequality is true: $g(\alpha) \leq f_{\mathbb{Z}}(x; \alpha) \leq h(\alpha)$.

Proof. As we have $f_{\mathbb{Z}}(1/2; \alpha) \leq f_{\mathbb{Z}}(x; \alpha) \leq f_{\mathbb{Z}}(0; \alpha)$ (Lemma 3.1), it suffices to show the estimates for $x = 1/2$ and $x = 0$, respectively. We distinguish two cases:

($\alpha \geq 1$): For the upper estimate, we have

$$\begin{aligned} f_{\mathbb{Z}}(0; \alpha) &= 1 + 2 \sum_{k \geq 1} e^{-\pi\alpha k^2} \leq 1 + 2e^{-\pi\alpha} \sum_{k \geq 1} e^{-\pi(k^2-1)} \leq 1 + 2e^{-\pi\alpha} \left(1 + \sum_{k \geq 3} e^{-\pi k}\right) \\ &= 1 + 2e^{-\pi\alpha} \left(1 + \frac{e^{-3\pi}}{1-e^{-\pi}}\right) \leq 1 + 2e^{-\pi\alpha} \left(1 + \frac{1}{11000}\right). \end{aligned}$$

As a lower estimate we simply use the leading term(s);

$$f_{\mathbb{Z}}(\frac{1}{2}; \alpha) = 2 \sum_{k \geq 1} e^{-\pi\alpha(k-1/2)^2} \geq 2e^{-\frac{\pi\alpha}{4}}.$$

($\alpha < 1$): We use the Poisson summation formula and the fact from above that $\frac{e^{-3\pi}}{1-e^{-\pi}} \leq \frac{1}{11000}$. We start with the upper estimate:

$$\begin{aligned} \sqrt{\alpha} f_{\mathbb{Z}}(0; \alpha) &= 1 + 2 \sum_{k \geq 1} e^{-\frac{\pi}{\alpha} k^2} \leq 1 + 2e^{-\frac{\pi}{\alpha}} \sum_{k \geq 1} e^{-\frac{\pi}{\alpha}(k^2-1)} \\ &\leq 1 + 2e^{-\frac{\pi}{\alpha}} \left(1 + \sum_{k \geq 3} e^{-\pi k}\right) \leq 1 + \frac{22002}{11000} e^{-\frac{\pi}{\alpha}}. \end{aligned}$$

For the lower estimate we obtain

$$\sqrt{\alpha} f_{\mathbb{Z}}(\frac{1}{2}; \alpha) = 1 + 2 \sum_{k \geq 1} (-1)^k e^{-\frac{\pi}{\alpha} k^2} \geq 1 - 2e^{-\frac{\pi}{\alpha}} \left(1 + \sum_{k \geq 3} e^{-\pi k}\right) \geq 1 - \frac{22002}{11000} e^{-\frac{\pi}{\alpha}}.$$

□

Setting (again) $f(x) = f_{\mathbb{Z}}(x; \alpha)$, we analyze $f'(x)$ to find solutions for (4.13). Using $\sqrt{\alpha}f(x) = \theta_3(x; 1/\alpha)$, we get a triple product representation for f (cf. [17]);

$$\sqrt{\alpha} f(x) = \prod_{k \geq 1} \left(1 - e^{-2k\pi/\alpha}\right) \left(1 + 2e^{-(2k-1)\pi/\alpha} \cos(2\pi x) + e^{-2(2k-1)\pi/\alpha}\right).$$

For the derivative with respect to x we introduce the function $Q(x; \alpha)$ from [17]:

$$\begin{aligned} Q(x; \alpha) &= 4\pi \sum_{l \geq 1} (1 - e^{-2\pi l\alpha}) e^{-(2l-1)\pi\alpha} \\ &\quad \times \prod_{\substack{k \geq 1 \\ k \neq l}} (1 - e^{-2\pi k\alpha}) \left(1 + 2e^{-(2k-1)\pi\alpha} \cos(2\pi x) + e^{-2(2k-1)\pi\alpha}\right). \end{aligned}$$

The function $f'(x)$ then satisfies

$$\sqrt{\alpha} f'(x) = -\sin(2\pi x) Q(x; \frac{1}{\alpha}).$$

The function $Q(x; \alpha)$ was analyzed in detail by Montgomery [17, Lem. 1]: it is 1-periodic, even, positive, strictly decreasing on $(0, 1/2)$ and strictly increasing on $(1/2, 1)$. In particular, it assumes global maxima in $\mathbb{Z}$ and global minima in $\mathbb{Z} + 1/2$. Moreover, we have the following estimates obtained by Montgomery [17, Lem. 2].

Lemma 4.5. *We define the following functions for $\alpha > 0$.*

$$A(\alpha) = \begin{cases} \alpha^{-3/2} e^{-\pi/(4\alpha)}, & \alpha < 1, \\ (1 - \frac{1}{3000}) 4\pi e^{-\pi\alpha}, & 1 \leq \alpha. \end{cases} \quad B(\alpha) = \begin{cases} \alpha^{-3/2}, & \alpha < 1, \\ (1 + \frac{1}{3000}) 4\pi e^{-\pi\alpha}, & 1 \leq \alpha. \end{cases}$$

Then, for all x the following estimates hold true: $A(\alpha) \leq Q(x; \alpha) \leq B(\alpha)$.

We use the factorization $\sqrt{\alpha} f'(x) = -\sin(2\pi x) Q(x; 1/\alpha)$ to find common points of the curves c_1 and c_2 (see Figure 3) to have solutions to (4.13):

$$c_1 : \frac{-f(x_2)}{f(2x_2)} = 2 \cos(2\pi x_1) \frac{Q(2x_1, \frac{1}{\alpha})}{Q(x_1, \frac{1}{\alpha})} \quad \text{and} \quad c_2 : \frac{-f(x_1)}{f(2x_1)} = 2 \cos(2\pi x_2) \frac{Q(2x_2, \frac{1}{\alpha})}{Q(x_2, \frac{1}{\alpha})}.$$

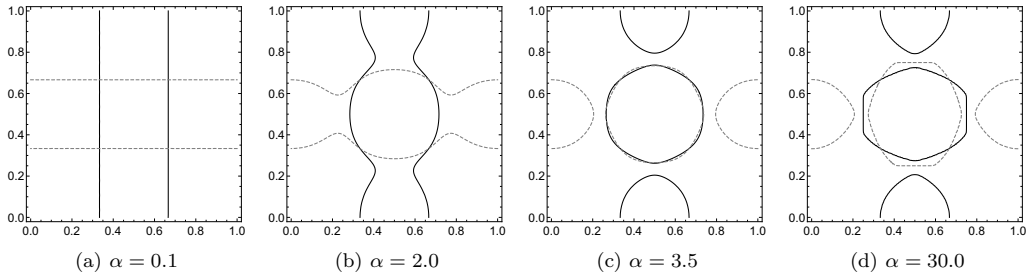


Figure 3. c_1 : —, and c_2 : ---, in $[0, 1]^2$. For most α we find four points of intersection, except for some α around 3.5. Plots created with *Mathematica* [21].

($\alpha \rightarrow 0$): Let us analyze the case when α is asymptotically 0. By Lemma 4.4 and Lemma 4.5, we conclude that for all $x \in \mathbb{R}$ we have

$$-\frac{f(x)}{f(2x)} \asymp -1 \quad \text{and} \quad 2 \cos(2\pi x) \frac{Q(2x, \frac{1}{\alpha})}{Q(x, \frac{1}{\alpha})} \asymp 2 \cos(2\pi x).$$

as α tends to 0. Hence, asymptotically we get the following curves:

$$c_1 : -1 \asymp 2 \cos(2\pi x_1), \forall x_2 \quad \text{and} \quad c_2 : -1 \asymp 2 \cos(2\pi x_2), \forall x_1.$$

This proves that indeed, as α tends to 0, we only have the solutions x_0 and x_0^* or the solutions $\tilde{x}_0$ and $\tilde{x}_0^*$ given by (4.14).

($\alpha \rightarrow \infty$): For α going to infinity we are not able to describe the exact shape of the limiting curves c_1 and c_2 (which are rotations of one another) in the unit square. The numerics tell us that c_1 and c_2 consist of two closed curves (seen periodically). Also, for any α the points $(\frac{1}{3}, 0)$, $(\frac{2}{3}, 0)$ are on c_1 and $(0, \frac{1}{3})$, $(0, \frac{2}{3})$ are part of c_2 .

4.4.3. Comparison. Note that Γ_{x_0, x_0^*} is a lattice and so by Theorem 3.4 it has lower energy than the hexagonal lattice Λ_2 . Scaling Γ_{x_0, x_0^*} to unit density, we may write

$$\Gamma_{x_0, x_0^*} = \sqrt{3} \begin{pmatrix} \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \mathbb{Z}^2 \cong \sqrt{\frac{2}{3}} \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{3}{2} \end{pmatrix} \mathbb{Z}^2,$$

where the second version is a clockwise rotation of Γ_{x_0, x_0^*} by 45 degrees. A question which arises is whether the square lattice $\mathbb{Z}^2$ may also have lower energy than Γ_{x_0, x_0^*} .

Theorem 4.6. *For all $\alpha > 0$, $\mathbb{Z}^2$ has lower energy than the configuration Γ_{x_0, x_0^*} .*

Proof. The energy of Γ_{x_0, x_0^*} can be estimated from below by

$$E_{\phi_{\pi\alpha}}(\Gamma_{x_0, x_0^*}) = \sum_{k, l \in \mathbb{Z}} e^{-\frac{2\pi\alpha}{3}(k^2 + kl + \frac{5}{2}l^2)} - 1 \geq 2e^{-\frac{2\pi\alpha}{3}}.$$

By Lemma 4.4, for $\alpha \geq 1$, it suffices to show

$$E_{\phi_{\pi\alpha}}(\mathbb{Z}^2) = f_{\mathbb{Z}}(0, \alpha)^2 - 1 \leq \left(1 + \frac{22002}{11000} e^{-\pi\alpha}\right)^2 - 1 < 2e^{-\frac{2\pi\alpha}{3}}.$$

By the change of variable $x = e^{-\frac{\pi\alpha}{3}}$, this reduces to showing for all $x \in (0, e^{-\frac{\pi}{3}}]$

$$0 > \left(1 + \frac{22002}{11000}x^3\right)^2 - 1 - 2x^2 = \frac{22002^2}{11000^2}x^2 \left(x^4 + 2 \cdot \frac{11000}{22002}x - 2 \cdot \frac{11000^2}{22002^2}\right).$$

Note that the quartic function in brackets is strictly monotonically increasing on the interval $(0, e^{-\frac{\pi}{3}}] \subseteq (0, 0.4)$, so that we can estimate as follows:

$$x^4 + 2 \cdot \frac{11000}{22002}x - 2 \cdot \frac{11000^2}{22002^2} \leq 0.4^4 + 2 \cdot \frac{11000}{22002} \cdot 0.4 - 2 \cdot \frac{11000^2}{22002^2} < 0.4^4 + 0.4 - 0.5 < 0.$$

This proves the claim for $\alpha \geq 1$. Since all lattices of unit density in $\mathbb{R}^2$ are symplectic and as 2-dimensional Gaussians are eigenfunctions of the symplectic Fourier transform (cf. [13, Sec. 2.2]), the case $\alpha < 1$ follows by the symplectic Poisson summation formula. Combining the cases then gives the full result for all $\alpha > 0$. $\square$

REFERENCES

- [1] A. Baernstein II. A minimum problem for heat kernels of flat tori. In *Extremal Riemann surfaces (San Francisco, CA, 1995)*, volume 201 of *Contemporary Mathematics*, page 227–243. American Mathematical Society, Providence, RI, 1997.
- [2] S. N. Bernstein. Sur les fonctions absolument monotones. *Acta Mathematica*, 52:1–66, 1928.
- [3] J. M. Borwein and P. B. Borwein. A Cubic Counterpart of Jacobi’s Identity and the AGM. *Transactions of the American Mathematical Society*, 332(2):691–701, 1991.
- [4] L. Bétermin and M. Faulhuber. Maximal theta functions – Universal optimality of the hexagonal lattice for Madelung-like lattice energies. *Journal d’Analyse Mathématique*, 149:307–341, 2023.
- [5] L. Bétermin, M. Faulhuber, and S. Steinerberger. A variational principle for Gaussian lattice sums. *arXiv preprint*, 2110.06008, 2021.
- [6] L. Bétermin and M. Petrache. Dimension reduction techniques for the minimization of theta functions on lattices. *Journal of Mathematical Physics*, 58:071902, 2017.
- [7] H. Cohn and A. Kumar. Universally optimal distribution of points on spheres. *Journal of the American Mathematical Society*, 20(1):99–148, 2007.
- [8] H. Cohn, A. Kumar, S. D. Miller, D. Radchenko, and M. S. Viazovska. Universal optimality of E_8 and Leech lattices and interpolation formulas. *Annals of Mathematics*, 196(3):983–1082, 2022.
- [9] H. Cohn, A. Kumar, and A. Schürmann. Ground states and formal duality relations in the Gaussian core model. *Physical Review E*, 80:061116, 2009.
- [10] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups*, volume 290 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 3. edition, 1998.
- [11] R. Coulangéon and A. Schürmann. Energy minimization, periodic sets and spherical designs. *International Mathematics Research Notices*, 4:829–848, 2012.
- [12] R. Coulangéon and A. Schürmann. Erratum: Energy Minimization, Periodic Sets, and Spherical Designs. *International Mathematics Research Notices*, 4:3198–3200, 2022.
- [13] M. Faulhuber. A short note on the frame set of odd functions. *Bulletin of the Australian Mathematical Society*, 98:481–493, 2018.
- [14] M. Faulhuber, A. Gumber, and I. Shafkulovska. The AGM of Gauss, Ramanujan’s corresponding theory, and spectral bounds of self-adjoint operators. *arXiv preprint*, 2022.
- [15] M. Faulhuber and S. Steinerberger. Optimal Gabor frame bounds for separable lattices and estimates for Jacobi theta functions. *Journal of Mathematical Analysis and Applications*, 445(1):407–422, 2017.
- [16] A. J. E. M. Janssen. Some Weyl-Heisenberg frame bound calculations. *Indagationes Mathematicae*, 7(2):165–183, 1996.
- [17] H. L. Montgomery. Minimal theta functions. *Glasgow Mathematical Journal*, 30:75–85, 1988.
- [18] E. Stein and R. Shakarchi. *Complex Analysis*. Princeton University Press, 2003.
- [19] E. T. Whittaker and G. N. Watson. *A Course of Modern Analysis*. Cambridge University Press, reprinted edition, 1969.
- [20] D. V. Widder. *The Laplace Transform*. Princeton University Press, 1941.
- [21] Wolfram Research, Inc., Champaign, IL. *Mathematica*, Version 12–13, 2019–2021.

3.5 From completeness of discrete translates to phaseless sampling of the short-time Fourier transform

Bibliographic Information

Grohs, P., Liehr, L., and Shafkulovska, I., From completeness of discrete translates to phaseless sampling of the short-time Fourier transform, *submitted* (2022)

Submission date

24 November 2022

Contribution

Lukas Liehr and Irina Shafkulovska contributed equally to the idea, concept, development, and writing of the paper. Philipp Grohs contributed scientific advice and proofreading.

E-Print

arxiv:2211.05687

FROM COMPLETENESS OF DISCRETE TRANSLATES TO PHASELESS SAMPLING OF THE SHORT-TIME FOURIER TRANSFORM

PHILIPP GROHS^{*,†,‡}, LUKAS LIEHR^{*}, AND IRINA SHAFKULOVSKA^{*,§}

ABSTRACT. This article bridges and contributes to two important research areas, namely the completeness problem for systems of translates in function spaces and the short-time Fourier transform (STFT) phase retrieval problem. As a first main contribution, we show that a complex-valued, compactly supported function can be uniquely recovered from samples of its spectrogram if certain density properties of an associated system of translates hold true. Secondly, we derive new completeness results for systems of discrete translates in spaces of continuous functions on compact sets. We finally combine these findings to deduce several novel recovery results from spectrogram samples. Our results hold for a large class of window functions, including Gaussians, all Hermite functions, as well as the practically highly relevant Airy disk function. To the best of our knowledge, our results constitute the first recovery guarantees for the sampled STFT phase retrieval problem with a non-Gaussian window.

Keywords. phase retrieval, discrete translates, completeness, sampling, short-time Fourier transform, lattice

AMS subject classifications. 41A30, 42C30, 94A12, 94A20

1. INTRODUCTION

The main aim of the present paper is to bridge and contribute to two active research areas in pure and applied mathematics: the problem of completeness of discrete translates and the discretization of the short-time Fourier transform (STFT) phase retrieval problem. Both problems have received significant attention in recent years:

- The completeness of translates problem constitutes a classical problem at the intersection of approximation theory, functional analysis,

^{*}FACULTY OF MATHEMATICS, UNIVERSITY OF VIENNA, OSKAR-MORGENSTERN-PLATZ 1, 1090 VIENNA, AUSTRIA

[†]RESEARCH NETWORK DATASCIENCE@UNIVIE, UNIVERSITY OF VIENNA, KOLINGASSE 14-16, 1090 VIENNA, AUSTRIA

[‡]JOHANN RADON INSTITUTE OF APPLIED AND COMPUTATIONAL MATHEMATICS, AUSTRIAN ACADEMY OF SCIENCES, ALTENBERGSTRASSE 69, 4040 LINZ, AUSTRIA

[§]ACOUSTICS RESEARCH INSTITUTE, AUSTRIAN ACADEMY OF SCIENCES, WOHLLEBENGASSE 12-14, 1040 VIENNA, AUSTRIA

E-mail addresses: philipp.grohs@univie.ac.at, lukas.liehr@univie.ac.at, irina.shafkulovska@univie.ac.at.

Date: Friday 11th November, 2022.

and harmonic analysis with a series of recent contributions, see for instance [53, 54].

- The study of the STFT phase retrieval problem has seen a surge in the mathematical community due to its relevance in a wide-ranging number of applications [1, 19, 42, 73]. The two recent survey papers [22, 28] contain an extensive list of results and references on this problem. More recent contributions on both the theoretical and numerical aspects can be found in [4, 6, 24, 33, 51].

We start by introducing both the completeness of translates problem as well as the STFT phase retrieval problem, followed by a detailed derivation of their interrelations.

1.1. Completeness of discrete translates. Let $(X, \|\cdot\|)$ be a Banach space over $\mathbb{C}$, let Λ be an index set and consider a subset $\{f_\lambda : \lambda \in \Lambda\} \subseteq X$. Then $\{f_\lambda : \lambda \in \Lambda\}$ is said to be complete in X if its linear span is dense in X . Determining whether a given system is complete lies at the very heart of approximation theory and numerical analysis alike. The traditionally most notable result is due to Weierstrass [68], later generalized by Stone [64, 65]. For a more modern exposition we refer to Pinkus' overview article on the completeness problem in approximation theory [55]. The practicality aspect further prompts one to look for structured complete sets. Indeed, suppose from now on that X is a function space on $\mathbb{R}^d$ and let $f_\lambda = f(\cdot - \lambda)$ be the λ -shift of a function $f \in X$. We define

$$(1) \quad \mathfrak{T}(f, \Lambda) := \{f(\cdot - \lambda) : \lambda \in \Lambda\}$$

and call $\mathfrak{T}(f, \Lambda)$ the *system of translates* of f by Λ . The *completeness of translates problem* asks for assumptions on X, f and Λ so that $\mathfrak{T}(f, \Lambda)$ is complete in X . The investigation of this problem produced a vast literature, a possible starting point being the classical results of Wiener [71]. In recent years, close attention was drawn to the situation where X is a space of functions defined on the entire real line and Λ is a *discrete* set, see for instance [11, 17, 48, 52, 53, 54]. However, the case where $X = C(K)$, $K \subseteq \mathbb{R}^d$ compact, is the space of continuous functions on K seems much less studied, notable exceptions include results of Landau [47], Zalik [72] and Akhiezer [55, Theorem 5.4.]. In this situation, the choice $\Lambda = \mathbb{R}^d$ leads to the theory of mean-periodic functions, a concept which will play a rather subtle but indispensable role in this paper.

Definition 1.1. A function $f \in C(\mathbb{R}^d)$ is said to be *mean-periodic* if there exists a compactly supported complex Radon measure μ , $\mu \neq 0$, such that

$$f * \mu = 0.$$

An application of the Hahn-Banach theorem implies that if f is not mean-periodic, then $\mathfrak{T}(f, \mathbb{R}^d)$ is complete in $C(K)$ for every compact set $K \subseteq \mathbb{R}^d$. It is precisely this compactness assumption that will be of central importance in the remainder of the present article.

1.2. Phaseless sampling of the short-time Fourier transform. Moving on to the second half of the title, we again consider a Banach space

$(X, \|\cdot\|)$, typically a Hilbert space $(X, \langle \cdot, \cdot \rangle)$. In this setting, we are given a linear operator acting as

$$\mathcal{A} : X \rightarrow \mathbb{C}^\Gamma, \quad f \mapsto (\mathcal{A}f(\gamma))_{\gamma \in \Gamma}.$$

Assuming $\mathcal{A}$ is injective, the phase retrieval problem asks whether the phaseless measurements

$$|\mathcal{A}| : X \rightarrow \mathbb{C}^\Gamma, \quad f \mapsto (|\mathcal{A}f(\gamma)|)_{\gamma \in \Gamma}$$

still define an injective operator. Since $\mathcal{A}$ is linear, for all complex numbers $\tau \in \mathbb{T}$ of modulus one and all $f \in X$ it holds that $|\mathcal{A}(\tau f)| = |\mathcal{A}f|$. Hence, we introduce the equivalence relation for $f, h \in X$

$$f \sim h \iff \exists \tau \in \mathbb{T} : f = \tau h.$$

Then the injectivity of $|\mathcal{A}|$ is to be understood on the quotient space $X/\sim$. The fundamental operator we are investigating here is the *short-time Fourier transform* (STFT) with respect to a *window function* $g \in L^2(\mathbb{R}^d)$, $g \neq 0$, given by

$$(2) \quad V_g f(x, \omega) := \int_{\mathbb{R}^d} f(t) \overline{g(t-x)} e^{-2\pi i \omega \cdot t} dt = \langle f, M_\omega T_x g \rangle, \quad (x, \omega) \in \mathbb{R}^{2d},$$

where $T_x g(t) = g(t-x)$ and $M_\omega g(t) = e^{2\pi i \omega \cdot t} g(t)$ denote the time resp. frequency shift, and $\langle v, w \rangle := \int_{\mathbb{R}^d} v(t) \overline{w(t)} dt$ denotes the L^2 -inner product of two functions $v, w \in L^2(\mathbb{R}^d)$, while $\omega \cdot t = \sum_{j=1}^d \omega_j t_j$ denotes the standard scalar product on $\mathbb{R}^d$. The square of the absolute value of the STFT is called the spectrogram of f ,

$$\text{SPEC}_g f(x, \omega) = |V_g f(x, \omega)|^2.$$

With respect to (STFT) phase retrieval, for a subspace $X \subseteq L^2(\mathbb{R}^d)$, subset $\Gamma \subseteq \mathbb{R}^{2d}$ and window function $g \in L^2(\mathbb{R}^d)$, we ask when

$$\text{SPEC}_g : X \rightarrow \mathbb{C}^\Gamma, \quad f \mapsto (|V_g f(\gamma)|^2)_{\gamma \in \Gamma}$$

is injective on $X/\sim$. Note that $|V_g f| \geq 0$, so having the samples of the spectrogram is equivalent to having the phaseless STFT samples. This setup fits within the abstract framework since the STFT is injective on $L^2(\mathbb{R}^d)$. In fact, there exists an associated Calderón-type reproducing formula. While this is a good starting point, the separability of $L^2(\mathbb{R}^d)$ motivates the search for discretizations of this representation. In particular, it would be advantageous to be able to uniquely determine f from the evaluation of its STFT (and its spectrogram, respectively) on a separated set Γ , i.e., a set with

$$\inf_{\substack{\gamma, \gamma' \in \Gamma \\ \gamma \neq \gamma'}} \|\gamma - \gamma'\|_2 > 0.$$

A particular type of separated sets are lattices, that is, sets of type $\Lambda = A\mathbb{Z}^{2d}$ for some invertible generating matrix $A \in \text{GL}(2d, \mathbb{R})$. The special structure comes with additional nice properties that allow us to understand the STFT sampling problem better through a range of duality theorems [26, 27]. Unavoidably, one encounters the reciprocal lattice

$$\Gamma^* := \left\{ \lambda \in \mathbb{R}^{2d} : \gamma \cdot \lambda \in \mathbb{Z} \text{ for all } \gamma \in \Gamma \right\}$$

and the fundamental domain of Γ , which is a measurable cross-section of $\mathbb{R}^{2d}/\Gamma$. The *volume* of the lattice $\text{vol}(\Gamma)$ is the Lebesgue measure of a fundamental domain of Γ and its *density* is its reciprocal $D(\Lambda) = \text{vol}(\Gamma)^{-1}$. With regards to the discretization of the STFT, the following meta-theorem holds [36].

Theorem 1.2. *Let g be a window function that is sufficiently well-localized in the time-frequency domain. Then for all lattices Γ of sufficient density, the restriction*

$$L^2(\mathbb{R}^d) \rightarrow \ell^2(\Gamma), \quad f \mapsto (V_g f(\gamma))_{\gamma \in \Gamma}$$

is a well-defined, bounded injective operator with a bounded inverse from its range to $L^2(\mathbb{R}^d)$.

The theorem not only ensures the injectivity of the restriction operator but also guarantees a stable reconstruction of the sampled STFT. In stark contrast to this, phase retrieval in infinite-dimensional Hilbert or Banach spaces is never uniformly stable [7, 18]. Moreover, the uniqueness properties of the operator SPEC when sampled on a lattice differs fundamentally from the setting above: if $X = L^2(\mathbb{R}^d)$, then for every $g \in L^2(\mathbb{R}^d)$ and all lattices $\Gamma \subseteq \mathbb{R}^{2d}$ which frequently appear in time-frequency analysis the operator $\text{SPEC}_g : X \rightarrow \mathbb{C}^\Gamma$ is *not* injective on $X/\sim$ [8, 30, 31].

1.3. Objectives of this paper. In an attempt to circumvent the discretization barrier, in the present paper we restrict ourselves to proper subspaces of $L^2(\mathbb{R}^d)$, namely various function spaces consisting of compactly supported functions. The goal of the present exposition is threefold:

- We derive new completeness results of systems of discrete translates for spaces of continuous functions on compact sets.
- We develop a systematic link between the problem of completeness of discrete translates and the phaseless sampling problem of the short-time Fourier transform. More precisely, given a certain function space $\mathcal{C} \subseteq L^2(\mathbb{R}^d)$ of compactly supported functions and a window function $g \in L^2(\mathbb{R}^d)$, we show how completeness of translates of a (suitably chosen) family of related functions $\{g_\omega : \omega \in \Omega\}$, Ω an index set, implies that a function $f \in \mathcal{C}$ is uniquely determined by phaseless the samples $|V_g f(\Gamma)|$, provided that Γ is a sufficiently dense lattice. Generalizations to non-uniform sampling sets Γ are also presented.
- Leveraging these results allows us to deduce several new sampling theorems for the STFT phase retrieval problem, among them the first ever sampling result for the Airy disk window function, which is of particular practical importance [16, Chapter 8.5.2].

The proofs are based on a combination of a duality principle with the theory of mean-periodic functions and uniqueness sets of multivariate entire functions of exponential type.

1.4. Main results. In this section we shall state the main results of the present article.

1.4.1. *From completeness to phaseless sampling.* We start with an abstract result that highlights an intimate relation between completeness properties of discrete translates and the property of every function $f \in L^2(K)$, $K \subseteq \mathbb{R}^d$ compact, being uniquely determined up to a global phase by phaseless STFT samples. As in previous publications, we use the tensor-product notation $g_\omega = (T_\omega g)\overline{g}$, $\omega \in \mathbb{R}^d$. Additionally, $\mathfrak{T}(f, \Lambda)$ denotes the system of Λ -translates of a function f as defined in Equation (1). Finally, $A \pm B$ denotes the standard set sum resp. set difference of $A, B \subseteq \mathbb{R}^d$,

$$A + B = \{a + b : a \in A, b \in B\}, \quad A - B = \{a - b : a \in A, b \in B\},$$

and $PW_S^p(\mathbb{R}^d)$ denotes the Paley-Wiener spaces subjected to a compact set $S \subseteq \mathbb{R}^d$,

$$PW_S^p(\mathbb{R}^d) = \left\{ f \in L^p(\mathbb{R}^d) : \text{supp } \hat{f} \subseteq S \right\}$$

Finally, $V_g f$ denotes the short-time Fourier transform as defined in Equation (2). To state the first main result, we recall the notion of *uniqueness sets*. Throughout this paper, by a uniqueness set for a function space X of continuous functions on $\mathbb{R}^d$ (or $\mathbb{C}^d$) we understand a subset $\Lambda \subseteq \mathbb{R}^d$ such that the samples of a function $f \in X$ at Λ uniquely determine f , that is, the following implication holds:

$$f(\lambda) = 0 \quad \forall \lambda \in \Lambda \quad \implies \quad f \equiv 0.$$

Theorem 1.3. *Let $\Lambda, \Gamma \subseteq \mathbb{R}^d$ and let $K \subseteq \mathbb{R}^d$ be a compact set. Let $g \in C(\mathbb{R}^d)$ be a window function such that for every $\omega \in K - K$ the system of translates $\mathfrak{T}(g_\omega, \Lambda)$ is complete in $(C(K), \|\cdot\|_\infty)$. If $\Gamma \subseteq \mathbb{R}^d$ is a uniqueness set for $PW_{K-K}^1(\mathbb{R}^d)$, then every $f \in L^2(K)$ is determined up to a global phase by $|V_g f(\Lambda \times \Gamma)|$. In particular, this holds whenever Γ is a lattice such that the difference set $K - K$ is contained in a fundamental domain of the reciprocal lattice Γ^* .*

The previous theorem is restated and proven in Section 3 as Theorem 3.2. In terms of discretization of STFT phase retrieval, it states the following: if one can pick a discrete set Λ so that the window function g has the property that $\mathfrak{T}(g_\omega, \Lambda)$ is complete in $(C(K), \|\cdot\|_\infty)$ for every $\omega \in K - K$, then every $f \in L^2(K)$ is automatically determined by discrete samples of its spectrogram. If Λ can be chosen as a lattice, so can the discrete set that determines f by spectrogram samples as well. It is worth mentioning that STFT phase retrieval on $L^2(K)$ is equivalent to STFT phase retrieval on $PW_K^2(\mathbb{R}^d)$: by the fundamental identity of time-frequency analysis, for all $x, \omega \in \mathbb{R}^d$ holds

$$V_g f(x, \omega) = e^{-2\pi i x \cdot \omega} V_{\mathcal{F}^{-1}g} \mathcal{F}^{-1} f(\omega, -x).$$

Therefore, any compactly supported function $f \in L^2(K)$ can be uniquely determined by $|V_g f(\Lambda \times \Gamma)|$ if and only if any band-limited function $h \in PW_K^2(\mathbb{R}^d)$ can be uniquely determined by $|V_{\mathcal{F}^{-1}g} h(\Gamma \times (-\Lambda))|$. Notice that while $\mathfrak{T}(g, \Lambda)$ may be complete, it is, in general, unclear how $\mathfrak{T}(g_\omega, \Lambda)$ behaves. Therefore, as windows, we consider a class of functions that satisfies completeness properties and is in addition closed under tensor products.

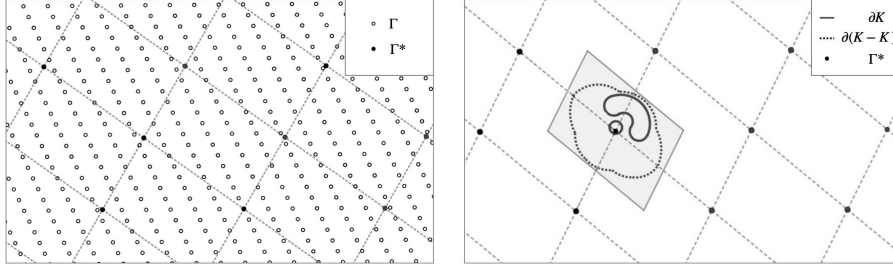


FIGURE 1. Left plot: A lattice Γ (empty circles) and its dual lattice Γ^* (full circles). Right plot: The boundary of a compact set K (full), the boundary of the set difference $K - K$ (dashed), the dual lattice Γ^* as in the left plot, and a fundamental domain of Γ^* containing $K - K$ (gray).

1.4.2. *The function class $\mathcal{P}_{\alpha,\beta}(\mathbb{C}^d)$.* The prime objective of the present section is to propose a unifying approach to known completeness results of discrete translates with Gaussian generators [72] and band-limited generators [55]. We provide both an extension and a synthesis of previous completeness results. This will lead to new uniqueness statements for the STFT phase retrieval problem.

Recall that an entire function $F : \mathbb{C}^d \rightarrow \mathbb{C}$ is said to be of exponential type no greater than $\sigma < \infty$ if for every $\varepsilon > 0$ it holds that

$$|F(z)| \lesssim e^{(\sigma+\varepsilon)\|z\|_1}, \quad z \in \mathbb{C}^d, \quad \|z\|_1 = \sum_{j=1}^d |z_j|.$$

The collection of all entire functions of exponential type no greater than σ is denoted by $\mathcal{E}_\sigma(\mathbb{C}^d)$. Among various properties, uniqueness sets have been explored in various setups [37, 49, 56, 58]. Observe that uniqueness sets for functions of exponential type can be chosen to be discrete, in particular, lattices. We now define a related function space, denoted with $\mathcal{P}_{\alpha,\beta}(\mathbb{C}^d)$, where we will draw window functions from.

Definition 1.4. *Let $0 \leq \alpha, \beta < \infty$. We define the class of functions $\mathcal{P}_{\alpha,\beta}(\mathbb{C}^d)$ via*

$$\mathcal{P}_{\alpha,\beta}(\mathbb{C}^d) := \left\{ z \mapsto p(z)e^{-z^T A z - b^T z} : \right. \\ \left. p \in \mathcal{E}_\alpha(\mathbb{C}^d), A \in \mathbb{C}^{d \times d}, b \in \mathbb{C}^d, \|A + A^T\|_1 \leq \beta \right\}.$$

We refer to Section 4 for the proofs of all forthcoming results related to $\mathcal{P}_{\alpha,\beta}(\mathbb{C}^d)$. Note that this class contains, for instance, (generalized) Gaussians, Hermite functions, band-limited functions and products of these. We will often use the same symbol for a function on $\mathbb{R}^d$ and its holomorphic extension (provided it exists) on $\mathbb{C}^d$. Conversely, we do not make a notational difference between a function on $\mathbb{C}^d$ and its restriction on $\mathbb{R}^d$. In addition to the previous concepts, we define the ℓ^∞ -diameter of a compact set K by

$$\Delta(K) := \max_{t, t' \in K} \|t - t'\|_\infty,$$

and arrive at the following completeness statement.

Theorem 1.5. *Let $K \subseteq \mathbb{R}^d$ be a compact set, $\varphi \in \mathcal{P}_{\alpha,\beta}(\mathbb{C}^d)$ not mean-periodic, and*

$$\sigma \geq \alpha + \frac{\beta}{2}\Delta(K).$$

If $\Lambda \subseteq \mathbb{R}^d$ is a uniqueness set for $\mathcal{E}_\sigma(\mathbb{C}^d)$, then $\mathfrak{T}(\varphi, \Lambda)$ is complete in $C(K)$.

There exists a vast literature on uniqueness sets for functions of exponential type. Consider for example the following classical result of Carlson [37] which easily transfers to the multivariate setting: $\mathbb{N}$ is a uniqueness set for the class of entire functions $F : \mathbb{C} \rightarrow \mathbb{C}$ which satisfy the growth condition $|F(z)| \lesssim e^{\beta|z|}$ for some $\beta < \pi$. With a standard projection to the univariate case, one can show that $\alpha\mathbb{N}^d$ is a uniqueness set for $\mathcal{E}_\sigma(\mathbb{C}^d)$ whenever $\alpha\sigma < \pi$. We refer to Section 4.2 for a discussion on uniqueness sets. We now arrive at the application to STFT phase retrieval in which we assume that the window function φ is square-integrable. Note that this is the standard assumption for the definition of the STFT.

Theorem 1.6. *Let $K \subseteq \mathbb{R}^d$ be a compact set, let $\varphi \in \mathcal{P}_{\alpha,\beta}(\mathbb{C}^d) \cap L^2(\mathbb{R}^d)$, $\varphi \neq 0$, and let*

$$\sigma \geq 2\alpha + \beta\Delta(K).$$

If $\Lambda \subseteq \mathbb{R}^d$ is a uniqueness set for $\mathcal{E}_\sigma(\mathbb{C}^d)$ and if Γ is a uniqueness set for $PW_{K-K}^1(\mathbb{R}^d)$, then every $f \in L^2(K)$ is determined up to a global phase by $|V_\varphi f(\Lambda \times \Gamma)|$. In particular, this holds if $K \subseteq [-\kappa, \kappa]^d$ for some $\kappa > 0$ and if $\sigma \geq 2\alpha + 2\beta\kappa$.

The sets Λ and Γ in the previous theorem can be chosen to be lattices where Λ depends on the growth variables α, β as well as the size of K which is governed by its diameter. For Hermite functions, we derive the following consequence.

Corollary 1.7. *Let $K \subseteq \mathbb{R}^d$ be a compact set, let $\mathfrak{h}$ be a Hermite function and let*

$$\sigma \geq 2\pi\Delta(K).$$

If $\Lambda \subseteq \mathbb{R}^d$ is a uniqueness set for $\mathcal{E}_\sigma(\mathbb{C}^d)$ and if $\Gamma \subseteq \mathbb{R}^d$ is a uniqueness set for $PW_{K-K}^1(\mathbb{R}^d)$, then every $f \in L^2(K)$ is determined up to a global phase by $|V_{\mathfrak{h}} f(\Lambda \times \Gamma)|$.

1.4.3. Band-limited windows and STFT-induced Paley-Wiener subspaces. Let $K' \subseteq \mathbb{R}^d$ be a compact set and let $p \in PW_{K'}^2(\mathbb{R}^d)$. It is evident that for $\alpha > 0$ sufficiently large, we have that $p \in \mathcal{E}_\alpha(\mathbb{C}^d)$. Hence, for such α it holds that $PW_{K'}^2(\mathbb{R}^d) \subseteq \mathcal{P}_{\alpha,0}(\mathbb{C}^d)$. Since band-limited functions are not mean-periodic, Theorem 1.5 applies and with a bit more effort we obtain the following statement (see Theorem 5.4).

Theorem 1.8. *Let $K, K' \subseteq \mathbb{R}^d$ be compact subsets and $g \in PW_{K'}^2(\mathbb{R}^d)$, $g \neq 0$. Then every $f \in L^2(K)$ is determined up to a global phase by $|V_g f(\mathcal{L})|$ provided that $\mathcal{L} \subseteq \mathbb{R}^{2d}$ is a lattice such that $(K' - K') \times (K - K)$ is contained in some fundamental domain of $\mathcal{L}^*$.*

Notice that a band-limited window function that is omnipresent in the application side of the STFT phase retrieval problem is the so-called Airy disk function (see, for instance, [16, Chapter 8.5.2]), which is given by

$$\mathbf{A}_a : \mathbb{R}^2 \rightarrow \mathbb{R}_{\geq 0}, \quad a > 0, \quad \mathbf{A}_a(t) = \left(\frac{a J_1(2\pi \|t\|_2 a)}{\|t\|_2} \right)^2,$$

where J_1 denotes the Bessel function of the first kind. For every $a > 0$, the map $\mathbf{A}_a$ is an element of $PW_K^2(\mathbb{R}^2)$ with $K = \{x \in \mathbb{R}^2 : \|x\|_2 \leq 2a\}$. We discuss this function in more detail in Section 5.

Furthermore, Theorem 1.8 highlights an intimate relation between STFT phase retrieval with band-limited window functions and phase retrieval theory in Paley-Wiener spaces: if $f \in L^2(K)$ for some compact set $K \subseteq \mathbb{R}^d$ and $g \in PW_{K'}^2(\mathbb{R}^d)$ for some compact set $K' \subseteq \mathbb{R}^d$, then $V_g f$ defines a band-limited function of $2d$ variables, namely $V_g f \in PW_{K' \times (-K)}^2(\mathbb{R}^{2d})$ (see Lemma 5.3). Before we progress with the discussion, let us make the following definition.

Definition 1.9 (STFT-induced Paley-Wiener subspace). *Let $K, K' \subseteq \mathbb{R}^{2d}$ be compact sets and let*

$$J := K' \times (-K).$$

We say that a subspace $\mathcal{S} \subseteq PW_J^2(\mathbb{R}^{2d})$ is STFT-induced by a band-limited window if there exists a band-limited function $g \in PW_{K'}^2(\mathbb{R}^d)$ such that

$$\mathcal{S} = V_g L^2(K) = \{V_g f : f \in L^2(K)\}.$$

We can now interpret the assertion of Theorem 1.8 as a result on phase retrieval in Paley-Wiener spaces: if $\mathcal{S} = V_g L^2(K)$ is an STFT-induced Paley-Wiener subspace, then every $F \in \mathcal{S}$ is determined up to a global phase by $|F(\mathcal{L})|$, provided that $\mathcal{L} \subseteq \mathbb{R}^{2d}$ is a lattice such that the Cartesian product $(K' - K') \times (K - K)$ is contained in a fundamental domain of $\mathcal{L}^*$. The question of phase retrieval in Paley-Wiener spaces was considered by several authors during the last years [6, 9, 15, 41, 46, 66]. Notice that usual phase retrieval in Paley-Wiener spaces is impossible in the following sense: for every $d \in \mathbb{N}$ and every compact $C \subseteq \mathbb{R}^d$ there exists $f, h \in PW_C^2(\mathbb{R}^d)$ such that $|f(t)| = |h(t)|$ for every $t \in \mathbb{R}^d$ but $f \not\sim h$. The entirety of functions f, h with the previous property can be constructed via the classical zero-flipping approach, as the fundamental papers of Akutowicz [2, 3], Hofstetter [39], and Walther [67] reveal. On the other hand, it was shown in [6, 66] that a restriction to real-valued functions in $PW_C^2(\mathbb{R}^d)$ yields uniqueness. However, it is unclear which infinite-dimensional subspaces consisting of *complex*-valued function obey the uniqueness property. To the best of our knowledge, STFT-induced Paley-Wiener subspaces constitute the first example of such subspaces.

1.4.4. Gaussian windows. A classical choice of a window function are Gaussians. Rather remarkably, no density assumption is required for Gaussian windows, but only a semigroup structure. This is hard to be expected from other windows as Proposition 4.2 suggests, adding to the Gaussian's special status within time-frequency analysis. We recall that the STFT with respect

to the Gaussian window is (up to a weight factor) the Bargmann transform, providing a unitary transformation from $L^2(\mathbb{R}^d)$ to the Fock space [36]. Moreover, the Gaussian is the only window with an analytic model space [10]. This connection to analytic functions leads to the most extensive results in STFT phase retrieval with Gaussian windows [4, 5, 8, 29, 32, 34, 35] and to the best of our knowledge, there are no results available for window functions varying from a Gaussian.

Theorem 1.10. *Let $g \in C(\mathbb{R}^d)$ be a multivariate complex-valued Gaussian*

$$g(x) = \exp\left(-(x - \nu)^T A(x - \nu)\right),$$

where $\nu \in \mathbb{C}^d$ and $A \in \mathbb{C}^{d \times d}$ is a Hermitian matrix mit invertible real part $\operatorname{Re} A$. If $K \subseteq \mathbb{R}^d$ is a compact set and Λ is a semigroup in $\mathbb{R}^d$ that contains a spanning set, then the system $\mathfrak{T}(g, \Lambda)$ is complete in $(C(K), \|\cdot\|_\infty)$. In particular, this holds when Λ is a lattice.

The proof can be found in Section 6. An immediate consequence is the following uniqueness result for Gabor phase retrieval.

Corollary 1.11. *Let $g \in C(\mathbb{R}^d)$ be a multivariate complex-valued Gaussian*

$$g(x) = \exp\left(-(x - \nu)^T A(x - \nu)\right),$$

where $\nu \in \mathbb{C}^d$ and where $A \in \mathbb{C}^{d \times d}$ is a positive definite Hermitian matrix. Suppose that $K \subseteq \mathbb{R}^d$ is a compact set. If $\Lambda, \Gamma \subseteq \mathbb{R}^d$ are lattices such that $K - K$ is contained in a fundamental domain of the reciprocal lattice Γ^* , then every $f \in L^2(K)$ is determined up to a global phase by $|V_g f(\Lambda \times \Gamma)|$.

We now combine the completeness properties of discrete Gaussian translates with the completeness theory of complex exponentials. To that end, we recall a result of Kahane which states the existence of a sequence $\mathcal{K} = \{\kappa_n : n \in \mathbb{Z}\} \subseteq \mathbb{R}$ with the property that the system of complex exponentials $\{e^{2\pi i \kappa_n \cdot} : n \in \mathbb{Z}\}$ is complete in $C(I)$ for every compact interval $I \subseteq \mathbb{R}$ while $\mathcal{K}$ has density zero. We refer to $\mathcal{K}$ as *Kahane's sequence*. In the context of STFT phase retrieval, the following can be said.

Theorem 1.12. *Let $g \in C(\mathbb{R}^d)$ be a real, multivariate Gaussian window function. Let $\mathcal{K} \subseteq \mathbb{R}$ be Kahane's sequence and let $\Lambda \subseteq \mathbb{R} \setminus \{0\}$ be a sequence of distinct real numbers of density zero such that $\sum_{\lambda \in \Lambda} \frac{1}{|\lambda|} = \infty$. Further, let*

$$\mathcal{V} := \bigcup_{\kappa > 0} L^4[-\kappa, \kappa]^d.$$

Then the following holds:

- i) $\mathcal{V}$ is dense in $L^2(\mathbb{R}^d)$,*
- ii) every $f \in \mathcal{V}$ is determined up to a global phase by $|V_g f(\Lambda^d \times \mathcal{K}^d)|$,*
- iii) $\Lambda^d \times \mathcal{K}^d$ has density zero.*

In particular, the theorem's conclusion holds true if Λ is the set of prime numbers.

Note that there exists no window function such that every $f \in L^2(\mathbb{R}^d)$ is determined up to a global phase by spectrogram samples on equidistant parallel hyperplanes of the form $\Lambda \times \mathbb{R}^d$, $\Lambda \subseteq \mathbb{R}^d$ lattice [31]. In particular,

there exists no window function and no separable lattice such that every $f \in L^2(\mathbb{R}^d)$ is determined up to a global phase by spectrogram samples on the lattice, irrespective of the density of the lattice. By contrast, Theorem 1.12 states that uniqueness can be achieved on a dense subset $\mathcal{V} \subseteq L^2(\mathbb{R}^d)$ from sampling sets which have density zero. This riveting conundrum once more underlines the sensitivity of the phase retrieval problem and the high informational value of the phase factor in time-frequency analysis.

1.4.5. Semi-discrete sampling sets. The results presented so far deal with discrete sampling sets and require some form of growth control on the window function. If one transfers to semi-discrete sampling sets as addressed by the discretization barriers mentioned above, then very weak assumptions on a function yield uniqueness. This is due to the fact that the problem of $\mathfrak{T}(g_\omega, \Lambda)$ being complete in $(C(K), \|\cdot\|_\infty)$ significantly simplifies if $\Lambda = \mathbb{R}^d$ (in this case Theorem 1.3 yields a uniqueness result for semi-discrete sampling sets which are for $d = 1$ parallel lines in the time-frequency plane). For semi-discrete sampling sets, we have the following statement.

Theorem 1.13. *Let $K \subseteq \mathbb{R}^d$ be a compact set. Suppose that $g : \mathbb{R}^d \rightarrow \mathbb{C}$ is a window function which satisfies the following two properties:*

- i) $g \in C(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$,*
- ii) for every $\omega \in K - K$ it holds that g_ω does not vanish identically.*

If Γ is a lattice such that $K - K$ is contained in a fundamental domain of the reciprocal lattice Γ^* , then every $f \in L^2(K)$ is determined up to a global phase by $|V_g f(\mathbb{R}^d \times \Gamma)|$.

1.5. Outline. The paper is divided as follows. In Section 2, we go over several elementary notions. In Section 3 we derive the connection between completeness of translates and STFT phase retrieval of compactly supported, square-integrable functions. Section 4 is dedicated to a detailed study of the function class $\mathcal{P}_{\alpha,\beta}(\mathbb{C}^d)$, its completeness properties, and consequentially, STFT phase retrieval results tied to $\mathcal{P}_{\alpha,\beta}(\mathbb{C}^d)$. We then consider STFT-induced Paley Wiener spaces and band-limited window functions in Section 1.4.3. We conclude with Section 6, which deals with the Gaussian window function and sampled STFT phase retrieval on dense subspaces of $L^2(\mathbb{R}^d)$.

2. PRELIMINARIES AND BASIC DEFINITIONS

In this section, we shall provide some basic definitions and settle the notation which is used throughout the remainder of the article.

2.1. (Short-time) Fourier transform. The Fourier transform of an integrable function $f \in L^1(\mathbb{R}^d)$ is given by

$$\mathcal{F}f(\omega) = \hat{f}(\omega) = \int_{\mathbb{R}^d} f(t) e^{-2\pi i \omega \cdot t} dt.$$

This defines a bounded operator from $L^1(\mathbb{R}^d)$ to $L^\infty(\mathbb{R}^d)$ and extends from $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ to a unitary operator on $L^2(\mathbb{R}^d)$ by standard density arguments. The short-time Fourier transform (STFT) $V_g f : \mathbb{R}^{2d} \rightarrow \mathbb{C}$ of a

function $f \in L^2(\mathbb{R}^d)$ with respect to a window function $g \in L^2(\mathbb{R}^d)$ is defined as in Equation (2). The STFT defines a uniformly continuous function on $\mathbb{R}^{2d}$ and satisfies

$$V_g f(x, \omega) = \mathcal{F}(f \overline{T_x g})(\omega) = (\mathcal{F} f * (M_{-x} \overline{\mathcal{R} \mathcal{F} g}))(\omega),$$

where $\mathcal{R}u(t) = u(-t)$ denotes the reflection operator. Further, we denote by $\langle v, w \rangle$ the integral

$$\langle v, w \rangle = \int_{\mathbb{R}^d} v(t) \overline{w(t)} dt, \quad v \in L^p(\mathbb{R}^d), \quad w \in L^q(\mathbb{R}^d),$$

where $p, q \in [1, \infty]$ are Hölder conjugate exponents.

2.2. Paley-Wiener spaces. A central role in this paper is played by Paley-Wiener spaces and uniqueness sets for these spaces. We primarily consider the Paley-Wiener spaces of integrable and square-integrable functions. Therefore, we define for $p \in \{1, 2\}$ and for a compact set $K \subseteq \mathbb{R}^d$ the Paley-Wiener space $PW_K^p(\mathbb{R}^d)$ via

$$PW_K^p(\mathbb{R}^d) = \left\{ f \in L^p(\mathbb{R}^d) : \text{supp}(\mathcal{F} f) \subseteq K \right\}.$$

The inclusion $PW_K^1(\mathbb{R}^d) \subseteq PW_K^2(\mathbb{R}^d)$ is well-known [38]. Recall that the classical theorem of Shannon, Whittaker and Kotelnikov asserts that if the subset $\Gamma \subseteq \mathbb{R}^d$ is a lattice, $K \subseteq \mathbb{R}^d$ a compact set, and $f \in PW_K^2(\mathbb{R}^d)$, then f can be uniquely retrieved from its samples at Γ if and only if K is contained in a fundamental domain D of the reciprocal lattice Γ^* (see, for instance, [38, Theorem 14.4] or [45, Theorem 4.4]). In particular, lattices of the previous type are examples of uniqueness sets for $PW_K^2(\mathbb{R}^d)$. As a matter of fact, f can be reconstructed via

$$f(t) = \sum_{\gamma \in \Gamma} f(\gamma) \mathfrak{r}(t - \gamma),$$

where the reconstruction function $\mathfrak{r}$ is given by

$$\mathfrak{r}(t) := \text{vol}(\Gamma^*) \int_D e^{2\pi i x \cdot t} dx.$$

The inclusion $PW_K^2(\mathbb{R}^d) \subseteq \mathcal{E}_{2\pi C}(\mathbb{C}^d)$, where $C = \max_{x \in K} \|x\|_\infty$, relates Paley-Wiener spaces to the spaces of entire functions of exponential type and uniqueness sets for $\mathcal{E}_{2\pi C}(\mathbb{C}^d)$ to uniqueness sets for $PW_K^2(\mathbb{R}^d)$.

2.3. Complex measures. For a locally compact Hausdorff space X we denote by $M(X)$ the set of all Radon measures on X . With the usual definition of total variation $|\mu|$ of $\mu \in M(X)$ [59, Chapter 6], the space $M(X)$ is a Banach space when endowed with the total variation norm given by $\|\mu\|_{TV} := |\mu|(X)$. Let $C_0(X) = (C_0(X), \|\cdot\|_\infty)$ be the Banach space of all complex-valued, continuous functions on X which vanish at infinity, that is, for every $\varepsilon > 0$ there exists a compact set $K \subseteq X$ such that $|f(x)| \leq \varepsilon$ for every $x \in K^c$. The classical Riesz representation theorem states that $C_0(X)^* \simeq M(X)$ in the following sense: $\Psi \in C_0(X)^*$ if and only if it can be represented in the form

$$\Psi(f) = \int_X f d\mu,$$

where $\mu \in M(X)$ is the unique Radon measure corresponding to Ψ . Moreover, the operator norm of Ψ is the total variation of μ , i.e., $\|\Psi\| = \|\mu\|_{TV}$. If $K \subseteq \mathbb{R}^d$ is a compact set, then $C_0(K) = C(K)$ and therefore its dual is $C(K)^* \simeq M(K)$.

3. ABSTRACT THEORY

3.1. Convolution operators and systems of translates. We dedicate this section to a conceptual take on STFT phase retrieval. Instead of showing that explicit functions together with (separated) sets can retrieve compactly supported functions, we provide general sufficient conditions in terms of completeness properties of discrete translates that can be tested on diverse families of window functions.

Theorem 3.1. *Let $p \in [1, \infty]$, $\Lambda, \Gamma \subseteq \mathbb{R}^d$ and $K \subseteq \mathbb{R}^d$ be a compact set. Let $g \in L_{\text{loc}}^\infty(\mathbb{R}^d)$ be a window function which has the property that for every $\omega \in K - K$ the linear operator*

$$\mathcal{C}_\omega : L^p(K) \rightarrow \mathbb{C}^\Lambda, \quad f \mapsto \left(\int_K f(t) T_\lambda g_\omega(t) dt \right)_{\lambda \in \Lambda}$$

is injective. Further, let Γ be a uniqueness set for $PW_{K-K}^1(\mathbb{R}^d)$. Then every $f \in L^{2p}(K)$ is determined up to a global phase by $|V_g f(\Lambda \times \Gamma)|$. In particular, this holds whenever $\Gamma \subseteq \mathbb{R}^d$ is a lattice such that $K - K$ is contained in a fundamental domain of the reciprocal lattice Γ^* .

Proof. Let $f \in L^{2p}(K)$ be a $2p$ -integrable function on K . Since $p \geq 1$, and K has bounded measure, we have $f \in L^2(K)$. Additionally, $g \in L_{\text{loc}}^\infty(\mathbb{R}^d)$, so it follows that $f \overline{T_x g} \in L^2(K)$ for every $x \in \mathbb{R}^d$. Using the convolution identity of the Fourier transform and the fact that $\mathcal{F}h = \mathcal{F}^{-1}\mathcal{R}h$, $\overline{\mathcal{F}h} = \mathcal{F}^{-1}\overline{h}$ for every $h \in L^2(\mathbb{R}^d)$, we can represent $|V_g f(x, \cdot)|^2$ as

$$\begin{aligned} |V_g f(x, \cdot)|^2 &= \mathcal{F}(f \overline{T_x g}) \overline{\mathcal{F}(f \overline{T_x g})} = \mathcal{F}^{-1}\mathcal{R}(f \overline{T_x g}) \mathcal{F}^{-1}(\overline{f \overline{T_x g}}) \\ &= \mathcal{F}^{-1} \left(\mathcal{R}(f \overline{T_x g}) * (\overline{f \overline{T_x g}}) \right). \end{aligned}$$

Since f has support in K , the latter convolution can be written as

$$\begin{aligned} \mathcal{R}(f \overline{T_x g}) * (\overline{f \overline{T_x g}})(s) &= \int_K f(-(s-t)) \overline{g(-(s-t)-x)} \overline{f(t)} g(t-x) dt \\ &= \int_K f_s(t) \overline{T_x g_s(t)} dt. \end{aligned}$$

Therefore, we obtain the identity

$$(3) \quad \mathcal{F} \left(|V_g f(x, \cdot)|^2 \right) (\omega) = \int_K f_\omega(t) \overline{T_x g_\omega(t)} dt.$$

Suppose that $\omega \in \mathbb{R}^d$ is a vector such that $(K + \omega) \cap K \neq \emptyset$. Then there exist $k, k' \in K$ such that $k + \omega = k'$, implying that $\omega \in K - K$. By contraposition, we obtain the implication

$$\omega \notin K - K \implies (K + \omega) \cap K = \emptyset.$$

Since $\text{supp}(T_\omega f) \subseteq K + \omega$ and $\text{supp}(f) \subseteq K$, the previous implication shows that $f_\omega = (T_\omega f) \overline{f} = 0$ almost everywhere provided $\omega \notin K - K$. In view of

Equation (3), the map

$$\omega \mapsto \mathcal{F} \left(|V_g f(x, \cdot)|^2 \right) (\omega)$$

is supported on $K - K$. Further, since $f \overline{T_x g} \in L^2(K)$ and the fact that the Fourier transform is an L^2 -isomorphism, we have

$$\left| \mathcal{F}(f \overline{T_x g}) \right|^2 = |V_g f(x, \cdot)|^2 \in L^1(\mathbb{R}^d),$$

and therefore, $|V_g f(x, \cdot)|^2 \in PW_{K-K}^1(\mathbb{R}^d)$. Now let $h \in L^2(K)$ be a second function such that the spectrogram samples of f and h agree on $\Lambda \times \Gamma$, i.e.,

$$|V_g f(\Lambda \times \Gamma)| = |V_g h(\Lambda \times \Gamma)|.$$

Since $|V_g f(x, \cdot)|^2, |V_g h(x, \cdot)|^2 \in PW_{K-K}^1(\mathbb{R}^d)$ and since Γ is a uniqueness set for $PW_{K-K}^1(\mathbb{R}^d)$, we conclude

$$\left| V_g f(\Lambda \times \mathbb{R}^d) \right| = \left| V_g h(\Lambda \times \mathbb{R}^d) \right|.$$

In particular, Identity (3) yields

$$\int_K f_\omega(t) \overline{T_\lambda g_\omega(t)} dt = \int_K h_\omega(t) \overline{T_\lambda g_\omega(t)} dt \quad \forall \lambda \in \Lambda, \forall \omega \in \mathbb{R}^d.$$

Given that $f, h \in L^2(K)$, we have $f_\omega - h_\omega \in L^p(K)$ by Hölder's inequality. The injectivity assumption on the operator $\mathcal{C}_\omega$ gives $f_\omega = h_\omega$ a.e. for all $\omega \in K - K$. Since $f_\omega = h_\omega = 0$ a.e. for all $\omega \notin K - K$, it follows that $f_\omega = h_\omega$ a.e. for every $\omega \in \mathbb{R}^d$. Further we have $f_\omega, h_\omega \in L^1(\mathbb{R}^d)$ and obtain $\mathcal{F}(f_\omega)(x) = \mathcal{F}(h_\omega)(x)$ for every $x \in \mathbb{R}^d$ and every $\omega \in \mathbb{R}^d$. Equivalently, it holds that $V_f f$ agrees with $V_h h$ everywhere in $\mathbb{R}^{2d}$, whence $f \sim h$ by standard results on time-frequency analysis (see for instance [12, 36]).

If we assume that Γ is a lattice and $K - K$ is contained in a fundamental domain of the reciprocal lattice Γ^* , then Shannon's sampling theorem implies that Γ is a sampling set for $PW_{K-K}^2(\mathbb{R}^d) \supseteq PW_{K-K}^1(\mathbb{R}^d)$, in particular, a uniqueness set for $PW_{K-K}^1(\mathbb{R}^d)$. $\square$

Notice that confirming the injectivity of the *family* of convolution operators $(\mathcal{C}_\omega)_{\omega \in K-K}$ is not a trivial matter. By Steinhaus' Theorem [63, 69], the set $K - K$ contains an open neighborhood of zero whenever the compactum K has positive Lebesgue measure, implying that one should guarantee the injectivity of uncountably many operators. If one aims at retrieving square-integrable functions, then the injectivity of $\mathcal{C}_\omega$ can be achieved via a completeness property of translates of $(g_\omega)_{\omega \in K-K}$. Recall that for a continuous function $g \in C(\mathbb{R}^d)$ and a set $\Lambda \subseteq \mathbb{R}^d$ we denote by $\mathfrak{T}(g, \Lambda) := \{T_\lambda g : \lambda \in \Lambda\}$ the system of Λ -translates of g . The relation between the injectivity conditions of Theorem 3.1 to completeness properties of translates reads as follows.

Theorem 3.2. *Let $p \in [1, \infty]$, let $\Lambda, \Gamma \subseteq \mathbb{R}^d$ and let $K \subseteq \mathbb{R}^d$ be a compact set. Let $g \in C(\mathbb{R}^d)$ be a window function such that for every $\omega \in K - K$ the system of translates $\mathfrak{T}(g_\omega, \Lambda)$ is complete in $(C(K), \|\cdot\|_\infty)$. If Γ is a uniqueness set for $PW_{K-K}^1(\mathbb{R}^d)$, then every $f \in L^2(K)$ is determined up to a global phase by $|V_g f(\Lambda \times \Gamma)|$. In particular, this holds whenever Γ is*

a lattice such that $K - K$ is contained in a fundamental domain of the reciprocal lattice Γ^* .

Proof. According to Theorem 3.1, it suffices to prove that the convolution operator

$$\mathcal{C}_\omega : L^1(K) \rightarrow \mathbb{C}^\Lambda, \quad f \mapsto \left(\int_K f(t) T_\lambda g_\omega(t) dt \right)_{\lambda \in \Lambda}$$

is injective for all $\omega \in K - K$. To that end, we define for $f \in L^1(K)$ the operator

$$\Phi_f : C(K) \rightarrow \mathbb{C}, \quad \Phi_f(q) = \int_K f(t) q(t) dt.$$

Then Φ_f is a continuous linear functional on $C(K)$ and by the Riesz representation theorem, there exists a unique $\mu \in M(K)$ such that the functional is given by $\Phi_f(q) = \int_K q d\mu$ and $\|\Phi_f\| = \|\mu\|_{TV}$. By uniqueness, we have $d\mu = f dm$, where m denotes the Lebesgue measure on $\mathbb{R}^d$, and therefore $\|\mu\|_{TV} = \|f\|_{L^1(K)}$. Thus, the completeness of $\mathfrak{T}(g_\omega, \Lambda)$ in $(C(K), \|\cdot\|_\infty)$ gives rise to the implication

$$\Phi_f(T_\lambda g_\omega) = 0 \quad \forall \lambda \in \Lambda \implies \Phi_f = 0 \implies f = 0 \text{ a.e.}$$

This shows that $\mathcal{C}_\omega$ is injective on $L^1(K)$. The statement follows from Theorem 3.2. $\square$

Remark 3.3 (Multiplication by periodic functions). Consider the operator $\mathcal{C}_\omega : L^p(K) \rightarrow \mathbb{C}^\Lambda$ as defined in Theorem 3.1,

$$\mathcal{C}_\omega(f) = \left(\int_K f(t) T_\lambda g_\omega(t) dt \right)_{\lambda \in \Lambda},$$

with $g \in L^\infty_{\text{loc}}(\mathbb{R}^d)$ and $\Lambda \subseteq \mathbb{R}^d$. Suppose that $\Lambda \subseteq \mathbb{R}^d$ is a lattice, i.e., $\Lambda = A\mathbb{Z}^d$ for some $A \in \text{GL}(d, \mathbb{R})$, and let $\chi \in L^\infty(\mathbb{R}^d)$ be an Λ -periodic function (i.e. $T_\lambda \chi = \chi$ for all $\lambda \in \Lambda$) such that $\chi(t) \neq 0$ for almost every $t \in \mathbb{R}^d$. Then the injectivity of $\mathcal{C}_\omega$ is unaffected if g is multiplied by χ . For, if $\mathcal{C}_\omega$ is injective, then

$$0 = \int_K f(t) T_\lambda (g\chi)_\omega(t) dt = \int_K f(t) \chi_\omega(t) T_\lambda g_\omega(t) dt \quad \forall \lambda \in \Lambda$$

and therefore, $f\chi_\omega = 0$, where we used that χ_ω is Λ -periodic and $f\chi_\omega \in L^p(K)$ provided $f \in L^p(K)$ and $\chi \in L^\infty(\mathbb{R}^d)$. Since for every $\omega \in \mathbb{R}^d$ it holds that $\chi_\omega(t) \neq 0$ for almost every $t \in \mathbb{R}^d$, we conclude that $f = 0$.

Remark 3.4 (Hermitian functions). Let $K \subseteq \mathbb{R}^d$ be a measurable set. A function $f \in L^2(K)$ is called Hermitian if $\mathcal{R}f = \overline{f}$. In the following, we shall denote the space of Hermitian functions by $L^2_{\mathcal{H}}(K)$. It is evident that a function is Hermitian if and only if its Fourier transform is real-valued. Now suppose that $K \subseteq \mathbb{R}^d$ is a compact set and let $g \in L^2_{\mathcal{H}}(\mathbb{R}^d)$ be a Hermitian window function such that $\hat{g} \in L^1(\mathbb{R}^d)$. If $f \in L^2_{\mathcal{H}}(K)$, then the fundamental identity of time-frequency analysis (see for instance, [36, Equation 3.10]) shows that for every $\omega \in \mathbb{R}^d$ one has

$$|V_g f(0, \omega)| = |V_{\hat{g}} \hat{f}(\omega, 0)| = |(\mathcal{R}\hat{g} * \hat{f})(\omega)|,$$

where we used that $\hat{g}$ is real-valued. The function $\mathcal{R}\hat{g} * \hat{f}$ is real-valued, square-integrable (Young's inequality), and its Fourier transform has support in K . In other words, $\mathcal{R}\hat{g} * \hat{f}$ is a real-valued function belonging to $PW_K^2(\mathbb{R}^d)$. It is well-known that a function of this type is determined up to a global sign-factor by its modulus sampled on a sufficiently dense lattice $\Gamma \subseteq \mathbb{R}^d$, see Thakur's theorem for the univariate setting and corresponding multivariate extensions by Daubechies et al. [5, 66]. Consequently, if $h \in L_{\mathcal{H}}^2(\mathbb{R}^d)$ is a second function such that

$$|V_g f(\{0\} \times \Gamma)| = |V_g h(\{0\} \times \Gamma)|,$$

then $\mathcal{R}\hat{g} * \hat{f} \sim \mathcal{R}\hat{g} * \hat{h}$. The assumption $\mathcal{R}g(t) \neq 0$ for almost every $t \in K$ implies that $f \sim h$. The uniqueness of STFT phase retrieval problem for the space $L_{\mathcal{H}}^2(K)$ follows therefore from the uniqueness of a phaseless deconvolution problem. A similar observation was done in [9] for the univariate setting. The uniqueness of this problem is a simple application of uniqueness results in real Paley-Wiener spaces.

3.2. Mean-periodicity. The notion of mean-periodicity goes back to Del-sarte [20, 21] and was later further developed by Schwartz [60]. For an extensive overview of mean-periodic functions, we refer to Kahane's exposition [43]. In order to motivate the notation of mean-periodicity, assume that $f \in C(\mathbb{R})$ and let $a > 0$. Suppose that f is an a -periodic function, that is, $f(x) = f(x - a)$ for all $x \in \mathbb{R}$. Clearly, an a -periodic function cannot be an $a\mathbb{Z}$ -generator for $C(K)$, $K \subseteq \mathbb{R}$ compact. The a -periodicity of f can be equivalently written as the requirement that

$$f * \mu(x) := \int_{\mathbb{R}} f(x - y) d\mu(y) = 0$$

for all $x \in \mathbb{R}$, where $\mu = \delta_0 - \delta_a$ is a sum of two Dirac measures. This motivates the following definition.

Definition 3.5. *A function $f \in C(\mathbb{R}^d)$ is said to be mean-periodic if there exists a compactly supported complex Radon measure μ , $\mu \neq 0$, such that $f * \mu = 0$.*

Besides the motivating example, the set of mean-periodic functions is rather sparse. One can easily verify that exponential polynomials, i.e., products of type $z \mapsto p(z)e^{\omega^T z}$, where p is a multivariate polynomial and $\omega \in \mathbb{C}^d$, are also mean-periodic. In fact, any mean periodic function can be given in terms of related exponential polynomials [13]. Several equivalent definitions of mean-periodicity are known. For our purposes, the most striking one asserts that f is mean-periodic if and only if $\mathfrak{T}(f, \mathbb{R}^d)$ is *not* dense in $C(\mathbb{R}^d)$ with respect to the topology of uniform convergence on compact sets [43]. The perhaps most direct way to eliminate the property of a function being mean-periodic is to require an integrability condition.

Proposition 3.6. *Let $f \in C(\mathbb{R}^d)$ such that $f \neq 0$. If $f \in \mathcal{S}'(\mathbb{R}^d)$ and its Fourier transform $\mathcal{F}f$ is induced by a measurable function, then f is not mean-periodic. In particular, if $f \in C(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$ for some $p \in [1, 2]$, then f is not mean-periodic.*

Proof. A function is understood as a tempered distribution via the induced functional

$$L_f : \mathcal{S}(\mathbb{R}^d) \rightarrow \mathbb{C}, \quad h \mapsto \int_{\mathbb{R}^d} h(t) f(t) dt.$$

The definition is compatible with the Fourier transform in the sense that, whenever the Fourier transform of f exists, e.g., when $f \in L^1(\mathbb{R}^d) \cup L^2(\mathbb{R}^d)$, then the Fourier transform of the distribution is given by $\mathcal{F}L_f = L_{\mathcal{F}f}$. The analogous (in restricted form) holds for products and convolutions and can be found in the textbooks [14, 25, 40].

To prove the above Proposition, we assume for the sake of contradiction that there exists a non-trivial compactly supported measure $\mu \in M(\mathbb{R}^d)$ such that $f * \mu = 0$. Due to the compact support and the bounded variation of μ , its Fourier transform $\mathcal{F}\mu$ is a smooth function, has a holomorphic extension on $\mathbb{C}^d$ and all its derivatives on $\mathbb{R}^d$ have at most polynomial growth at infinity (see [40, Theorem 7.3.1]). Since f induces the tempered distribution L_f , it also holds $L_f * \mu = 0$ as a distribution. The Convolution Theorem for distributions [14, Theorem 8.5.1] implies

$$0 = \mathcal{F}(L_f * \mu) = \mathcal{F}L_f \cdot \mathcal{F}\mu = L_{\mathcal{F}f} \cdot \mathcal{F}\mu.$$

While this is initially to be understood in the distributional sense, since we already know that $\mathcal{F}L_f = L_{\mathcal{F}f}$, we can also interpret this in the pointwise a.e. sense, leading us to conclude

$$0 = L_{\mathcal{F}f} \cdot \mathcal{F}\mu = L_{\mathcal{F}f \cdot \mathcal{F}\mu}.$$

This holds if and only if

$$\mathcal{F}f \cdot \mathcal{F}\mu = 0 \text{ a.e.}$$

Since f does not vanish everywhere and the Fourier transform is an automorphism on $\mathcal{S}'(\mathbb{R}^d)$, $\mathcal{F}f$ also does not vanish identically, implying there exists a set $\Omega \subseteq \mathbb{R}^d$ of positive Lebesgue measure such that $\mathcal{F}\mu|_{\Omega} = 0$. Considering that $\mathcal{F}\mu$ has a holomorphic extension, $\mathcal{F}\mu$ must vanish identically [30, Lemma 3.4], thereby proving that also $L_\mu = 0$. This is the desired contradiction.

For the second part of the statement, suppose that $q \in [2, \infty]$ is the Hölder conjugate of $p \in [1, 2]$. Hausdorff-Young's inequality states that the Fourier transform has a continuous extension $L^p(\mathbb{R}^d) \rightarrow L^q(\mathbb{R}^d)$. In particular, the tempered distribution induced by f has a Fourier transform induced by a measurable function. $\square$

We note that the special case when $f \in C(\mathbb{R}^d)$, $d = 1$, is an integrable function that does not vanish identically was already proven in [55, Proposition 6.3]. Furthermore, this cannot be trivially extended for $L^p(\mathbb{R}^d)$, $p > 2$. The key role in the proof of Proposition 3.6 is the fact that the support of $\mathcal{F}f$ has positive Lebesgue measure. If $p > 2$, then there are $f \in L^p(\mathbb{R}^d)$ whose (distributional) Fourier transform is not induced by a measurable function. As a matter of fact, there are functions whose Fourier transform is supported on a set of measure zero. A nice example is the Fourier transform of the surface measure σ_d on the sphere S^{d-1} . In that case holds (cf. [25, Appendix

B.4])

$$f(\omega) = \mathcal{F}^{-1}\sigma_d(\omega) = \mathcal{F}\sigma_d(\omega) = 2\pi \|\omega\|_2^{-\frac{d-2}{2}} J_{\frac{d-2}{2}}(2\pi \|\omega\|_2),$$

where J_ν denotes the Bessel function of first kind and order ν . With standard asymptotic estimates on Bessel functions (see [25, Appendix B.6, Appendix B.8]), one can easily show $f \in L^p(\mathbb{R}^d)$ as long as it holds $(p-2)(d-1) > 2$.

The previous considerations immediately yield a phase retrieval result related to unions of hyperplanes.

Theorem 3.7. *Let $K \subseteq \mathbb{R}^d$ be a compact set. Suppose that $g : \mathbb{R}^d \rightarrow \mathbb{C}$ is a window function that satisfies the following two properties:*

- i) $g \in C(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$,
- ii) *for every $\omega \in K - K$ it holds that g_ω does not vanish identically.*

If Γ is a lattice such that $K - K$ is contained in a fundamental domain of the reciprocal lattice Γ^* , then every $f \in L^2(K)$ is determined up to a global phase by $|V_g f(\mathbb{R}^d \times \Gamma)|$.

Proof. Since $g \in C(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, it follows that for every $\omega \in \mathbb{R}^d$ we have $g_\omega \in C(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$. By assumption, $g_\omega \neq 0$ for every $\omega \in K - K$. In view of Proposition 3.6, $\mathfrak{T}(g_\omega, \mathbb{R}^d)$ is complete in $C(K)$ for every $\omega \in K - K$. Theorem 3.2 yields the statement. $\square$

Remark 3.8 (Non-uniqueness theory in $L^2(\mathbb{R}^d)$). It was recently shown in [31, Theorem 1.1] that for every lattice $\Gamma \subseteq \mathbb{R}^d$ and every window function $g \in L^2(\mathbb{R}^d)$ there exist $f, h \in L^2(\mathbb{R}^d)$ such that

$$|V_g f(\mathbb{R}^d \times \Gamma)| = |V_g h(\mathbb{R}^d \times \Gamma)| \quad \text{and} \quad f \approx h.$$

Hence, without a support condition on the underlying signal space, the corresponding phase retrieval problem always fails to be unique if one samples on $\mathbb{R}^d \times \Gamma$, irrespective of the choice of Γ and g . In contrast, Theorem 3.7 shows that mild conditions on g imply unique recovery of compactly supported signals from sampling sets of the form $\mathbb{R}^d \times \Gamma$.

Remark 3.9 (Landau's theorem). A theorem due to Landau [47] asserts that if $\varphi \in L^1(\mathbb{R})$ extends to an entire function and if $\Lambda \subseteq [-(\delta - \kappa), \delta - \kappa]$, $0 < \kappa < \delta < \infty$, is a sequence which converges to zero, then $\mathfrak{T}(\varphi, \Lambda)$ is complete in $C[-\kappa, \kappa]$. Now suppose that $g \in L^2(\mathbb{R})$ extends to an entire function. Then for every $\omega \in \mathbb{R}$ it holds that $g_\omega \in L^1(\mathbb{R})$. Further, for every $t \in \mathbb{R}$ we have $g_\omega(t) = g(t - \omega) \overline{g(\overline{t})}$, where the right-hand side of the previous equation extends to an entire function. Consequently, if Λ, κ, δ are given as in Landau's theorem and if $K = [-(\delta - \kappa), \delta - \kappa]$, then according to Theorem 3.2, every $f \in L^2(K)$ is determined up to a global phase by $|V_g f(\Lambda \times \Gamma)|$, provided that Γ is a uniqueness set for $PW_{K-K}^1(\mathbb{R})$. Note however that in this situation Λ , hence $\Lambda \times \Gamma$, is not separated.

3.3. Completeness via duality. In view of Theorem 3.2 and the search for separated sampling sets, we concern ourselves with criteria for completeness of discrete translates. One option is to use Stone-Weierstrass-type arguments, which will be pursued in Section 6 dedicated to the Gaussian window function. However, the algebra structure required in the Stone-Weierstrass

theorem is quite a strong one. We, therefore, consider a duality strategy. To that end, recall that for all compact sets $K \subseteq \mathbb{R}^d$ there is a one-to-one correspondence between the complex Radon measures $\mu \in M(K)$ and the dual space $C(K)^*$ via

$$L_\mu(f) = \int_K f d\mu, \quad \|L_\mu\|_{C(K)^*} = \|\mu\|_{TV}.$$

To relate this to the problem of completeness of discrete translates, we define an associated function space $\mathcal{O}_\varphi(K)$. For this, let $\varphi : \mathbb{C}^d \rightarrow \mathbb{C}$ be a continuous function and $K \subseteq \mathbb{R}^d$ compact. Then $\mathcal{O}_\varphi(K)$ is defined by

$$\mathcal{O}_\varphi(K) := \left\{ F : \mathbb{C}^d \rightarrow \mathbb{C} : F(z) = L_\mu(T_z\varphi) = \int_K T_z\varphi d\mu, \mu \in M(K) \right\}.$$

Proposition 3.10. *Let $K \subseteq \mathbb{R}^d$ be a compact set, $\Lambda \subseteq \mathbb{R}^d$ and $\varphi : \mathbb{C}^d \rightarrow \mathbb{C}$ be a continuous function which is not mean-periodic. Then the following are equivalent:*

- i) $\mathfrak{T}(\varphi, \Lambda)$ is complete in $C(K)$,
- ii) for every $\mu \in M(K)$ holds the implication

$$L_\mu(T_\lambda\varphi) = 0 \quad \forall \lambda \in \Lambda \implies L_\mu = 0,$$

- iii) Λ is a uniqueness set for $\mathcal{O}_\varphi(K)$.

Proof. Step 1: Equivalence of i) and ii). Suppose that $\mathfrak{T}(\varphi, \Lambda)$ is complete in $C(K)$. It follows that if $g \in C(K)$, then there exists a sequence $(g_k)_{k \in \mathbb{N}} \subseteq \text{span}_{\mathbb{C}} \mathfrak{T}(\varphi, \Lambda)$ such that $g_k \rightarrow g$ in $C(K)$. Hence, if $L_\mu(T_\lambda\varphi) = 0$ for all $\lambda \in \Lambda$, then by linearity, $L_\mu(g_k) = 0$ for all $k \in \mathbb{N}$. Since L_μ is a continuous linear functional, we have $L_\mu(g) = 0$. As g was arbitrary, it follows that $L_\mu = 0$. Suppose on the other hand that ii) holds, but $\mathfrak{T}(\varphi, \Lambda)$ is not complete. Then there exists a function $g \in C(K)$ such that

$$g \notin S := \overline{\text{span}_{\mathbb{C}} \mathfrak{T}(\varphi, \Lambda)}.$$

By Hahn-Banach theorem, it follows that there exists a continuous linear functional $\Psi \in C(K)^*$ such that $\|\Psi\| = 1$ and $\Psi|_S = 0$. Since $\Psi = \Psi_\mu$ for some $\mu \in M(K)$, we arrive at a contradiction.

Step 2: Equivalence of ii) and iii). Suppose that condition ii) holds. Let $F \in \mathcal{O}_\varphi(K)$ with

$$(4) \quad F(z) = \int_K T_z\varphi(t) d\mu(t)$$

for some $\mu \in M(K)$ and assume that $F(\lambda) = 0$ for every $\lambda \in \Lambda$. Since $F(\lambda) = L_\mu(T_\lambda\varphi)$, it follows that $L_\mu = 0$. Observing that $T_z\varphi \in C(K)$ for every $z \in \mathbb{C}^d$ shows that $L_\mu(T_z\varphi) = 0$ for every $z \in \mathbb{C}^d$. Therefore, the function F vanishes identically. Suppose on the other hand that Λ is a uniqueness set for $\mathcal{O}_\varphi(K)$ and let $\mu \in M(K)$ such that $L_\mu(T_\lambda\varphi) = 0$ for every $\lambda \in \Lambda$. If $F \in \mathcal{O}_\varphi(K)$ is given as in Equation (4), then the uniqueness property of Λ implies that F vanishes identically. Now observe that $F = \mathcal{R}\varphi * \mu$. Since the class of not mean-periodic functions is invariant under reflections, it follows that $\mathcal{R}\varphi$ is not mean-periodic. Therefore, $\mu = 0$, which in turn shows that $L_\mu = 0$ and the statement is proved. $\square$

Remark 3.11. We note that the equivalence (i) $\Leftrightarrow$ (ii) is independent of the mean-periodicity of the generator f . However, the valuable criterium for the following section is the third statement where the non-mean-periodicity is decisive for the equivalence. As a matter of fact, without assuming anything on the relationship between f and K , then f not being mean-periodic is both a necessary and sufficient condition for the equivalence chain.

4. THE FUNCTION SPACE $\mathcal{P}_{\alpha,\beta}(\mathbb{C}^d)$

This section builds up on the equivalences derived in Theorem 3.10. The main aim is to present the proofs of the results stated in Section 1.4.2.

4.1. Completeness. Recall that an entire function $F : \mathbb{C}^d \rightarrow \mathbb{C}$ is said to be of *exponential type* no greater than $\sigma < \infty$ if for every $\varepsilon > 0$ it holds that

$$|F(z)| \lesssim e^{(\sigma+\varepsilon)\|z\|_1}, \quad z \in \mathbb{C}^d.$$

The collection of all entire functions of exponential type no greater than σ is denoted by $\mathcal{E}_\sigma(\mathbb{C}^d)$. For $0 \leq \alpha, \beta < \infty$, we define the class of functions $\mathcal{P}_{\alpha,\beta}(\mathbb{C}^d)$ via

$$\begin{aligned} \mathcal{P}_{\alpha,\beta}(\mathbb{C}^d) := \{ & z \mapsto p(z)e^{-z^T A z - b^T z} : \\ & p \in \mathcal{E}_\alpha(\mathbb{C}^d), A \in \mathbb{C}^{d \times d}, b \in \mathbb{C}^d, \|A + A^T\|_1 \leq \beta \}. \end{aligned}$$

This class is intended to cover a spectrum of functions, ranging from band-limited functions $f(z)e^{-z^T \cdot 0 \cdot z}$, $f \in PW_K^2(\mathbb{R}^d)$, to exponential polynomials $p(z)e^{\omega^T z - z^T \cdot 0 \cdot z}$, $p : \mathbb{C}^d \rightarrow \mathbb{C}$ a polynomial, Hermite functions, and further. Notice that for all $z \in \mathbb{C}^d$ and $A \in \mathbb{C}^{d \times d}$ holds

$$z^T A z = \frac{1}{2} z^T A z + \frac{1}{2} (z^T A z)^T = \frac{1}{2} z^T (A + A^T) z,$$

so that we can without loss of generality assume that A is symmetric (with bound $2\|A\|_1 \leq \beta$). The last required parameter is the diameter of the domain of interest. For a compact set $K \subseteq \mathbb{R}^d$ denote its ℓ^∞ -diameter by

$$\Delta(K) := \max_{t, t' \in K} \|t - t'\|_\infty.$$

With the previous concepts in hand, we are equipped to prove one of the main theorems on completeness of discrete translates of the present paper.

Theorem 4.1. *Let $K \subseteq \mathbb{R}^d$ be a compact set, $\varphi \in \mathcal{P}_{\alpha,\beta}(\mathbb{C}^d)$ not mean-periodic, and*

$$\sigma \geq \alpha + \frac{\beta}{2} \Delta(K).$$

If $\Lambda \subseteq \mathbb{R}^d$ is a uniqueness set for $\mathcal{E}_\sigma(\mathbb{C}^d)$, then $\mathfrak{T}(\varphi, \Lambda)$ is complete in $C(K)$.

Proof. Let $\varphi(t) = p(t)e^{-t^T A t - b^T t}$ for some $p \in \mathcal{E}_\alpha(\mathbb{C}^d)$, $b \in \mathbb{C}^d$ and $A \in \mathbb{C}^{d \times d}$ a matrix which satisfies the upper bound $\|A + A^T\|_1 \leq \beta$. According to Theorem 3.10, it suffices to show that Λ is a uniqueness set for $\mathcal{O}_\varphi(K)$. To that end, let $F \in \mathcal{O}_\varphi(K)$ with

$$F(z) = \int_K T_z \varphi(t) d\mu(t)$$

for some $\mu \in M(K)$. We have to show that F vanishes identically, provided that F vanishes at Λ . Let $t_0 \in \mathbb{R}^d$ such that $K - t_0$ lies in the centered cube

$$\left[-\frac{\Delta(K)}{2}, \frac{\Delta(K)}{2} \right]^d.$$

Since the diameter is invariant under translations, this is the optimal centered cube that fits K . Observe that

$$\begin{aligned} T_z \varphi(t) &= p(t - z) e^{-(t-z)^T A(t-z) - b^T(t-z)} \\ &= p(t - z) e^{(t-t_0)^T (A+A^T)z} e^{-X(z)} e^{-t^T A t - b^T t}, \end{aligned}$$

where $X(z) = z^T A z - t_0^T (A + A^T) z - b^T z$ is an entire function independent of t . Hence, if G is defined by

$$(5) \quad G(z) = \int_K p(t - z) e^{(t-t_0)^T (A+A^T)z} d\tilde{\mu}(t)$$

with $\tilde{\mu}(t) = e^{-t^T A t - b^T t} \mu(t) \in M(K)$, then $e^X F = G$. For every $\varepsilon > 0$ the map G satisfies the estimate

$$\begin{aligned} |G(z)| &\leq \|\tilde{\mu}\|_{TV} \max_{t \in K} \left| p(t - z) e^{(t-t_0)^T (A+A^T)z} \right| \\ &\lesssim \|\tilde{\mu}\|_{TV} \max_{t \in K} \left(e^{(\alpha+\varepsilon)\|t-z\|_1} \cdot e^{\|t-t_0\|_\infty \|A+A^T\|_1 \|z\|_1} \right) \\ &\lesssim \|\tilde{\mu}\|_{TV} e^{(\alpha+\varepsilon)\|z\|_1} \cdot e^{\frac{\beta}{2} \Delta(K) \|z\|_1}. \end{aligned}$$

Since $\|\tilde{\mu}\|_{TV}$ is independent of z , it follows that G is an entire function of exponential type no greater than $\sigma = \alpha + \frac{\beta}{2} \Delta(K)$. We have therefore shown that $e^X F \in \mathcal{E}_\sigma(\mathbb{C}^d)$. Since $e^{X(z)} \neq 0$ for every z , it follows that F vanishes on Λ if and only if $e^X F$ vanishes on Λ . The assumption on Λ implies that $e^X F = 0$, whence $F = 0$, thereby proving the assertion. $\square$

At this stage, it is meaningful to reflect upon the proof approach presented in Theorem 4.1. To that end, we consider a minimal example with a univariate Gaussian $\varphi(t) = e^{-t^2}$, a centered interval $K = [-\kappa, \kappa]$, and a set $\Lambda \subseteq \mathbb{R}$. Since the Gaussian is integrable, hence not mean-periodic, by Theorem 3.10, the completeness is equivalent to showing

$$\forall \mu \in M([-\kappa, \kappa]) : L_\mu(T_\lambda \varphi) = 0 \ \forall \lambda \in \Lambda \implies L_\mu = 0.$$

A key step in the proof of Theorem 4.1 is the derivation of equation (5). In the setting of the univariate Gaussian, the functional equation of the exponential transforms the problem to considering for all $\lambda \in \Lambda$ the integral

$$0 = \sigma(\lambda) \int_{-\kappa}^{\kappa} e^{2\lambda t} \psi(t) d\mu(t) = \sigma(\lambda) \int_{-\kappa}^{\kappa} e^{2\lambda t} d\tilde{\mu}(t),$$

where $\sigma(\lambda) = e^{-\lambda^2}$, $\psi(t) = e^{-t^2}$, and $d\tilde{\mu}(t) = \psi(t) d\mu(t)$. The Gaussian is a bounded function on the subset $[-\kappa, \kappa] \subseteq \mathbb{R}$, so $\tilde{\mu} \in M([-\kappa, \kappa])$. The factor σ is not zero, so we are left with investigating the real zeroes of

$$F(z) = \int_{-\kappa}^{\kappa} e^{2zt} d\tilde{\mu}(t), \quad z \in \mathbb{C}.$$

This function is well-defined, holomorphic, and of exponential type. The problem, therefore, reduces to the search for uniqueness sets for functions of

exponential type. We point out that this method can only be used with the Gaussian due to its functional equation, as the following statement confirms.

Proposition 4.2. *Let $\varphi : \mathbb{C}^d \rightarrow \mathbb{C}$ be a non-zero entire function. There exist entire functions $\psi, \sigma : \mathbb{C}^d \rightarrow \mathbb{C}$ and a matrix $A \in \mathbb{C}^{d \times d}$ satisfying*

$$(6) \quad \varphi(z + \lambda) = \psi(z) \sigma(\lambda) e^{-2z^T A \lambda}, \quad z, \lambda \in \mathbb{C}^d$$

if and only if A is symmetric and there exist $b \in \mathbb{C}^d$ and $c \in \mathbb{C}$ such that φ is given by

$$(7) \quad \varphi(z) = e^{-z^T A z + z^T b + c}.$$

Proof. See Appendix. $\square$

4.2. Uniqueness sets for functions of exponential type. We present two distinct criteria for uniqueness sets due to Carlson [37] and Ronkin [57]. Carlson's theorem does assume a semigroup structure on the sampling set, while Ronkin's theorem holds for arbitrary separated sets. However, the former allows for a considerably better density condition on the point configuration.

Theorem 4.3 (Carlson [37]). *Suppose that $F : \mathbb{C} \rightarrow \mathbb{C}$ is an entire function which satisfies the growth condition $|F(z)| \lesssim e^{\beta|z|}$ where $\beta < \pi$. Then $F|_{\mathbb{N}} = 0$ implies that $F = 0$.*

Notice that the density assumption $\beta < \pi$ in Carlson's theorem is sharp as evidenced by $\sin(\pi z)$ which vanishes at $\mathbb{N}$ but does not vanish identically. Further observe that Carlson's theorem readily transfers to the multivariate setting, leading to the statement that $\alpha\mathbb{N}^d$ is a uniqueness set for $\mathcal{E}_\sigma(\mathbb{C}^d)$ provided that $\alpha\sigma < \pi$. The same holds for $\alpha\mathbb{Z}^d$ and this cannot be improved due to the same counterexample.

Now let $\Lambda \subseteq \mathbb{R}^d$ be an arbitrary countable set. Then it comes with two distinct quantities related to its separability and its density. Firstly, the minimum gap of Λ is defined via

$$\delta := \inf_{\substack{\lambda, \lambda' \in \Lambda \\ \lambda \neq \lambda'}} \|\lambda - \lambda'\|_2.$$

In view of this definition, Λ is called δ -separated if $\delta > 0$. On the other end, we denote by μ_Λ the number of points of Λ which lie in the cube $T_r = (-r, r)^d$. The lower resp. upper density of a set Λ is given by

$$D^-(\Lambda) := \liminf_{r \rightarrow \infty} \frac{\mu_\Lambda(r)}{(2r)^d}, \quad D^+(\Lambda) := \limsup_{r \rightarrow \infty} \frac{\mu_\Lambda(r)}{(2r)^d}.$$

For the prototypical example of a lattice $\Lambda = A\mathbb{Z}^d$, $A \in \text{GL}_d(\mathbb{R})$, we have

$$D^-(\Lambda) = D^+(\Lambda) = \text{vol}(\Lambda)^{-1}.$$

Theorem 4.4 (Ronkin [57]). *Let the countable set $\Lambda \subseteq \mathbb{R}^d$ be δ -separated. Furthermore, let the entire function $F : \mathbb{C}^d \rightarrow \mathbb{C}$, satisfy the condition $F(\lambda) = 0$ for all $\lambda \in \Lambda$. If $F(z)$ is of exponential type no greater than $\sigma < A_d \delta^{d-1} D^+(\Lambda)$, where*

$$A_d = \frac{2}{d!} \left(\frac{\pi}{2} \right)^{d-1} \left(\frac{d-1}{2} + e + \sum_{k=0}^{d-2} \left(\left(\frac{29}{3} \right)^{d-2-k} \left(\frac{13}{3} k + \frac{25}{3} e \right) \right) \right)^{-1},$$

then F vanishes identically.

4.3. Implications for sampled STFT phase retrieval. We are now in a position to cross over from complete systems of translates to phase retrieval results. We begin with a technical observation.

Lemma 4.5. *Let $\varphi \in \mathcal{P}_{\alpha,\beta}(\mathbb{C}^d)$, $\varphi \neq 0$, and let $\omega \in \mathbb{R}^d$. Then φ_ω is the restriction to $\mathbb{R}^d$ of a non-zero entire function which belongs to $\mathcal{P}_{2\alpha,2\beta}(\mathbb{C}^d)$.*

Proof. Let $\omega \in \mathbb{R}^d$ and let $\varphi \in \mathcal{P}_{\alpha,\beta}(\mathbb{C}^d)$, i.e., $\varphi(t) = p(t)e^{-t^T A t - b^T t}$ for a suitable non-zero $p \in \mathcal{E}_\alpha(\mathbb{C}^d)$, $b \in \mathbb{C}^d$ and $A \in \mathbb{C}^{d \times d}$. For every real $t \in \mathbb{R}^d$, we have

$$\begin{aligned} \varphi_\omega(t) &= p(t - \omega) \overline{p(\overline{t})} e^{-(t - \omega)^T A (t - \omega) - b^T (t - \omega)} e^{-t^T \overline{A} t - \overline{b}^T t} \\ &= p(t - \omega) \overline{p(\overline{t})} e^{-2t^T (\operatorname{Re} A) t} e^{(\omega^T (A + A^T) - (\operatorname{Re} b)^T) t} e^{-\omega^T A \omega + b^T \omega}. \end{aligned}$$

Replacing $t \in \mathbb{R}^d$ with $z \in \mathbb{C}^d$ we obtain for every $\varepsilon > 0$ the growth estimate

$$(8) \quad |e^{-\omega^T A \omega + b^T \omega} p(z - \omega) \overline{p(\overline{z})}| \lesssim e^{(\alpha + \varepsilon) \|z - \omega\|_1} e^{(\alpha + \varepsilon) \|z\|_1} \lesssim e^{(2\alpha + 2\varepsilon) \|z\|_1},$$

Note that the implicit constant in the estimate (8) depends only on ω . Since $p(z - \omega) \overline{p(\overline{z})} \in \mathcal{E}_{2\alpha}(\mathbb{C}^d)$, the last estimate shows that $\varphi_\omega \in \mathcal{P}_{2\alpha,2\beta}(\mathbb{C}^d)$. Considering the fact that p does not vanish identically, neither does the map $z \mapsto p(z - \omega) \overline{p(\overline{z})}$. $\square$

With Theorem 3.2 in hindsight, we arrive at the following statement on STFT phase retrieval with window functions belonging to the class $\mathcal{P}_{\alpha,\beta}(\mathbb{C}^d)$.

Theorem 4.6. *Let $K \subseteq \mathbb{R}^d$ be a compact set, $\varphi \in \mathcal{P}_{\alpha,\beta}(\mathbb{C}^d) \cap L^2(\mathbb{R}^d)$, $\varphi \neq 0$, and*

$$\sigma \geq 2\alpha + \beta \Delta(K).$$

If $\Lambda \subseteq \mathbb{R}^d$ is a uniqueness set for $\mathcal{E}_\sigma(\mathbb{C}^d)$ and if Γ is a uniqueness set for $PW_{K-K}^1(\mathbb{R}^d)$, then every $f \in L^2(K)$ is determined up to a global phase by $|V_\varphi f(\Lambda \times \Gamma)|$. In particular, this holds if $K \subseteq [-\kappa, \kappa]^d$ for some $\kappa > 0$ and if $\sigma \geq 2\alpha + 2\beta\kappa$.

Proof. Let $\omega \in \mathbb{R}^d$. According to Lemma 4.5, it holds that $\varphi_\omega \neq 0$. In addition, the assumption that $\varphi \in L^2(\mathbb{R}^d)$ implies that $\varphi_\omega \in L^1(\mathbb{R}^d)$. In particular, φ_ω is not mean-periodic. Now let $\omega \in K - K$. According to Lemma 4.5, the map φ_ω extends from $\mathbb{R}^d$ to an entire function which belongs to $\mathcal{P}_{2\alpha,2\beta}(\mathbb{C}^d)$. Applying Theorem 4.1 shows that $\mathfrak{T}(\varphi_\omega, \Lambda)$ is complete in $C(K)$. Since $\omega \in K - K$ was arbitrary, an application of Theorem 3.2 implies the assertion. $\square$

The Hermite functions on $\mathbb{R}$ are defined by

$$h_n(t) = \frac{2^{1/4}}{\sqrt{n!}} \left(-\frac{1}{\sqrt{2\pi}} \right)^n e^{\pi t^2} \left(\frac{d}{dt} \right)^n e^{-2\pi t^2}, \quad n \in \mathbb{N}_0.$$

The multivariate Hermite functions $\{\mathfrak{h}_k : k \in \mathbb{N}_0^d\} \subseteq L^2(\mathbb{R}^d)$ are defined via tensorisation,

$$\mathfrak{h}_k(t_1, \dots, t_d) := \prod_{j=1}^d h_{k_j}(t_j).$$

Every Hermite function $\mathfrak{h} = \mathfrak{h}_k$, $k \in \mathbb{N}_0^d$, arises as the product of a polynomial with the Gaussian $2^{d/4}e^{-\pi\|t\|_2^2}$. Since every polynomial is of exponential type zero it holds that $\mathfrak{h} \in \mathcal{P}_{0,2\pi}(\mathbb{C}^d)$. This implies the next corollary.

Corollary 4.7. *Let $K \subseteq \mathbb{R}^d$ be a compact set and $\mathfrak{h} = \mathfrak{h}_k$ be an arbitrary d -dimensional Hermite function. If $\Lambda \subseteq \mathbb{R}^d$ is a uniqueness set for $\mathcal{E}_\sigma(\mathbb{R}^d)$ with $\sigma \geq 2\pi\Delta(K)$ and $\Gamma \subseteq \mathbb{R}^d$ is a uniqueness set for $PW_{K-K}^1(\mathbb{R}^d)$, then every $f \in L^2(K)$ is determined up to a global phase by $|V_{\mathfrak{h}}f(\Lambda \times \Gamma)|$. In particular, this holds if $K \subseteq [-\kappa, \kappa]^d$ for some $\kappa > 0$ and if $\sigma \geq 4\pi\kappa$.*

5. BAND-LIMITED WINDOW FUNCTIONS

This section is dedicated to the proofs of the statements of Section 1.4.3. We start by showing that the discrete translates of a band-limited function form a complete set. If $d = 1$, then this is essentially known [55, Proposition 5.5]. A similar argument as in [55] yields a respective result for the multivariate case. For the convenience of the reader, we shall prove this statement in the following.

Proposition 5.1. *Let $K' \subseteq \mathbb{R}^d$ be a compact set and $\varphi \in PW_{K'}^2(\mathbb{R}^d)$, $\varphi \neq 0$. If $\Lambda \subseteq \mathbb{R}^d$ is a uniqueness set for $PW_{-K'}^2(\mathbb{R}^d)$, then $\mathfrak{T}(\varphi, \Lambda)$ is complete in $C(\mathbb{R}^d)$ with respect to the topology of uniform convergence on compact sets. In other words, $\mathfrak{T}(\varphi, \Lambda)$ is complete in $C(K)$ for every compact set $K \subseteq \mathbb{R}^d$.*

Proof. Let $K \subseteq \mathbb{R}^d$ be a compact set and let $\mu \in M(K)$. In view of Proposition 3.10, we have to show that the function

$$F(z) = \int_K T_z \varphi(t) d\mu(t)$$

vanishes identically provided that F vanishes at Λ . Since $\varphi \in PW_{K'}^2(\mathbb{R}^d)$, we have $\varphi = \mathcal{F}g$ for some $g \in L^2(-K') \subseteq L^1(-K')$. Using Fubini's theorem, we have

$$\begin{aligned} F(z) &= \int_K \mathcal{F}g(t-z) d\mu(t) = \int_K \int_{-K'} g(s) e^{-2\pi i s \cdot (t-z)} ds d\mu(t) \\ &= \int_{-K'} g(s) \mathcal{F}\mu(s) e^{2\pi i s \cdot z} ds = \mathcal{F}(g\mathcal{F}\mu)(-z) = \mathcal{F}^{-1}(g\mathcal{F}\mu)(z). \end{aligned}$$

Since $\mathcal{F}\mu$ is an entire function, it holds that $g\mathcal{F}\mu \in L^2(-K')$ and therefore $F \in PW_{-K'}^2(\mathbb{R}^d)$. The statement follows from the assumption of Λ being a uniqueness set for $PW_{-K'}^2(\mathbb{R}^d)$. $\square$

An immediate consequence of the previous statement is a phase retrieval result for band-limited window functions.

Corollary 5.2. *Let $K, K' \subseteq \mathbb{R}^d$ be compact sets and let $g \in PW_{K'}^2(\mathbb{R}^d)$, $g \neq 0$. If $\Gamma \subseteq \mathbb{R}^d$ is a uniqueness set for $PW_{K-K}^1(\mathbb{R}^d)$ and $\Lambda \subseteq \mathbb{R}^d$ is a uniqueness set for $PW_{K'-K}^2(\mathbb{R}^d)$, then every $f \in L^2(K)$ is determined up to a global phase by $|V_g f(\Lambda \times \Gamma)|$.*

Proof. If $g \in PW_{K'}^2(\mathbb{R}^d)$ does not vanish identically, then for every $\omega \in \mathbb{R}^d$, the map φ_ω does not vanish identically and extends from $\mathbb{R}^d$ to an entire

function on $\mathbb{C}^d$ which satisfies

$$g_\omega \in PW_{K'-K'}^1(\mathbb{R}^d) \subseteq PW_{-(K'-K')}^2(\mathbb{R}^d).$$

Combining Proposition 5.1 with Theorem 3.2 yields the statement. $\square$

We now consider function spaces of the form

$$\mathcal{S} = V_g L^2(K) = \{V_g f : f \in L^2(K)\},$$

where $K, K' \subseteq \mathbb{R}^d$ are compact sets and $g \in PW_{K'}^2(\mathbb{R}^d)$. According to the notion settled in Section 1.4.3, we call $\mathcal{S}$ an STFT-induced Paley-Wiener subspace. This notion is due to the following observation.

Lemma 5.3. *Let $K, K' \subseteq \mathbb{R}^d$ be compact sets and let $g \in PW_{K'}^2(\mathbb{R}^d)$, $g \neq 0$. If $J = K' \times (-K)$, then $V_g L^2(K) \subseteq PW_J^2(\mathbb{R}^d)$.*

Proof. First observe that $V_g f \in L^2(\mathbb{R}^{2d})$ since both f and g are square integrable. Since $V_g f(x, \omega) = \mathcal{F}(f \overline{T_x g})(\omega)$, we have

$$\mathcal{F}(V_g f(x, \cdot))(\omega) = \mathcal{R}(f \overline{T_x g})(\omega) = \overline{g(-\omega - x)} f(-\omega).$$

and the right-hand side vanishes provided that $\omega \in -K$. If $\psi(x) = \overline{g(-\omega - x)}$, then we further have that

$$\mathcal{F}\psi(y) = e^{2\pi i \omega \cdot y} \overline{\mathcal{F}g(y)}.$$

Therefore, the $2d$ -dimensional Fourier transform of $V_g f$ is given by

$$\begin{aligned} \mathcal{F}(V_g f)(y, \omega) &= \mathcal{F}(x \mapsto \mathcal{F}(V_g f(x, \cdot))(\omega))(y) \\ &= \mathcal{F}(x \mapsto \overline{g(-\omega - x)} f(-\omega))(y) \\ &= e^{2\pi i \omega \cdot y} f(-\omega) \overline{\mathcal{F}g(y)}. \end{aligned}$$

Since $g \in PW_{K'}^2(\mathbb{R}^d)$, this implies that the support of the $2d$ -dimensional Fourier transform of $V_g f$ is contained in $K' \times (-K)$, thereby proving the assertion. $\square$

The previous considerations imply uniqueness results for STFT phase retrieval with band-limited window functions, and equivalently, for STFT-induced Paley-Wiener subspaces.

Theorem 5.4. *Let $K, K' \subseteq \mathbb{R}^d$ be compact sets and let $g \in PW_{K'}^2(\mathbb{R}^d)$, $g \neq 0$. Further, let $U = (K' - K') \times (K - K)$ and $\mathcal{L} \subseteq \mathbb{R}^{2d}$ be a uniqueness set for $PW_U^1(\mathbb{R}^d)$. Then every $F \in \mathcal{S} = V_g L^2(K)$ is determined up to a global phase by $|F(\mathcal{L})|$.*

Proof. Let $F \in \mathcal{S}$ such that $F = V_g f$ for some $f \in L^2(K)$. According to Lemma 5.3 it holds that $F \in PW_J^2(\mathbb{R}^{2d})$ where $J = K' \times (-K)$. With a similar argument as in the previous sections, we deduce that $|F|^2 \in PW_{J-J}^1(\mathbb{R}^{2d})$. Since $J-J = U$ and since $\mathcal{L}$ is a uniqueness set for $PW_U^1(\mathbb{R}^{2d})$, it follows that if $H = V_g h \in \mathcal{S}$, $h \in L^2(K)$, is a second function such that $|F(\mathcal{L})| = |H(\mathcal{L})|$, then $|F| = |H|$ everywhere on $\mathbb{R}^{2d}$. Corollary 5.2 shows that $f \sim h$, and since V_g is a linear transformation, we obtain $F \sim H$. $\square$

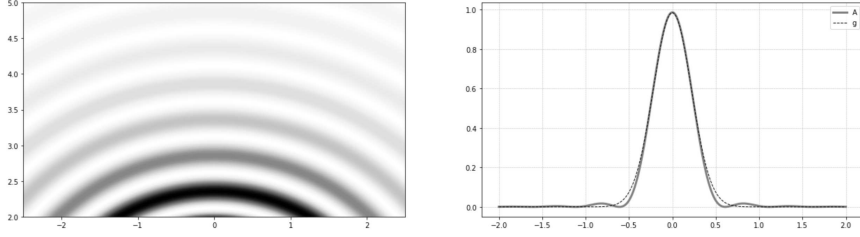


FIGURE 2. Left plot: contour plot of an Airy disk. Right plot: plot of a one-dimensional Gaussian g and an Airy disk function $r \mapsto \mathbf{A}_a(r) = \mathbf{A}_a(re_1) = \mathbf{A}_a(re_\theta)$, $e_\theta \in \partial B_1(0)$.

Example 5.5 (Airy disk). Let $D_a := \{x \in \mathbb{R}^2 : |x| \leq a\}$ be a circular disk of radius $a > 0$ and center 0 in $\mathbb{R}^2$. The square of the Fourier transform of the characteristic function of D_a is called the Airy disk of radius a . We have

$$\mathcal{F}(\mathbf{1}_{D_a})(\omega) = \int_0^a r \int_0^{2\pi} e^{-2\pi i r \|\omega\|_2 \cos(\theta)} d\theta dr = 2\pi \int_0^a r J_0(2\pi \|\omega\|_2 r) dr,$$

where we used polar coordinates as well as the definition of the Bessel function

$$J_n(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(x \sin \tau - n\tau)} d\tau.$$

The identity

$$\int_0^a x J_0(x) dx = a J_1(a)$$

implies that the Airy disk of radius a is then given by the radial function

$$\mathbf{A}_a(\omega) = (\mathcal{F}(\mathbf{1}_{D_a})(\omega))^2 = \left(\frac{a J_1(2\pi \|\omega\|_2 a)}{\|\omega\|_2} \right)^2.$$

As shown in Figure 2, this map behaves approximately like a Gaussian (except for the decay at infinity). The Airy disk frequently appears in diffraction imaging, namely in the situation where an incoming wavefront hits on a circular aperture (a pinhole) before getting diffracted at an object of interest. See, for instance, [16, Chapter 8.5.2]. The STFT phase retrieval problem with Airy window functions appears naturally in this context.

6. GAUSSIAN WINDOWS: GABOR PHASE RETRIEVAL

6.1. Completeness of discrete Gaussian translates. Any Gaussian belongs to the class $\mathcal{P}_{0,\beta}(\mathbb{C}^d)$ for a suitable $\beta > 0$, hence the completeness Theorem 4.1, and consequentially Theorem 3.7 on phase retrieval, apply with appropriate density assumptions on the generator set Λ . As already evidenced by the Bargmann transform [50, 61, 62], the uncertainty principles and the current progress on phase retrieval [4, 5, 8, 29, 32, 34, 35], the Gaussian is an exceptional window function. We underline this once more by showing that the completeness of $\mathfrak{T}(g, \Lambda)$ does not depend on the density of the lattice Λ when g is a Gaussian generator.

Theorem 6.1. *Let $g \in C(\mathbb{R}^d)$ be a multivariate complex-valued Gaussian*

$$g(x) = \exp\left(-(x - \nu)^T A(x - \nu)\right),$$

where $\nu \in \mathbb{C}^d$ and $A \in \mathbb{C}^{d \times d}$ is a Hermitian matrix mit invertible real part $\operatorname{Re} A$. If $K \subseteq \mathbb{R}^d$ is a compact set and Λ is a semigroup in $\mathbb{R}^d$ that contains a spanning set, then the system $\mathfrak{T}(g, \Lambda)$ is complete in $(C(K), \|\cdot\|_\infty)$. In particular, this holds when Λ is a lattice.

Proof. By expanding the term $-(x - \lambda - \nu)^T A(x - \lambda - \nu)$, we obtain the factorization

$$T_\lambda G(x) = c(\lambda) a(x) B_\lambda(x)$$

with

$$c(\lambda) = \exp\left(-(\lambda + \nu)^T A(\lambda + \nu)\right),$$

$$a(x) = \exp\left(-x^T A x + 2((\operatorname{Re} A) x)^T \nu\right),$$

$$B_\lambda(x) = \exp\left(2((\operatorname{Re} A) x)^T \lambda\right),$$

where we used that A is Hermitian. Notice that the completeness of $\mathfrak{T}(g, \Lambda)$ is unaffected if each $T_\lambda G$ is multiplied by a non-zero constant. Since $c(\lambda) \neq 0$, $\mathfrak{T}(g, \Lambda)$ is complete if and only if the set $\{a B_\lambda : \lambda \in \Lambda\}$ is complete. Moreover, a is a smooth, non-vanishing weighting factor independent of λ . Hence, $\{a B_\lambda : \lambda \in \Lambda\}$ is complete if and only if $\{B_\lambda : \lambda \in \Lambda\}$ is complete. Now set

$$S := \operatorname{span}_{\mathbb{C}}\{B_\lambda : \lambda \in \Lambda\}.$$

We show that S satisfies the assumptions of the Stone-Weierstrass theorem. Since B_λ is real-valued, the set S is invariant under complex conjugation. Further, for $\lambda, \lambda' \in \Lambda$ holds

$$B_\lambda(x) B_{\lambda'}(x) = B_{\lambda + \lambda'}(x),$$

which is an element of S since $\lambda + \lambda' \in \Lambda$. It remains to show that S separates points. To that end, let $x \neq y \in K$ and $\lambda \in \Lambda$. Then the equality $B_\lambda(x) = B_\lambda(y)$ holds if and only if

$$((\operatorname{Re} A)(x - y))^T \lambda = 0.$$

Because of the spanning property of Λ , if this equality holds for all $\lambda \in \Lambda$, then it holds for all $\lambda \in \mathbb{R}^d$, implying that

$$(\operatorname{Re} A)(x - y) = 0.$$

Since $\operatorname{Re} A$ is invertible, we have $x = y$. In conclusion, the Stone-Weierstrass theorem [23, Theorem 4.51] implies that the $\|\cdot\|_\infty$ -closure of S is either equal to $C(K)$ or equal to $\{f \in C(K) : f(x_0) = 0\}$ for a unique $x_0 \in K$. The later can be excluded since $B_\lambda(x_0) > 0$ for all $x_0 \in K$. $\square$

6.2. Implications for sampled Gabor phase retrieval. As an immediate consequence, we can retrieve compactly supported functions $f \in L^2(K)$ from the phaseless samples $|V_g f(\mathcal{L})|$ with respect to a Gaussian window g on sets $\mathcal{L} \subseteq \mathbb{R}^{2d}$ of arbitrary density.

Theorem 6.2. *Let $g \in C(\mathbb{R}^d)$ be a multivariate complex-valued Gaussian*

$$g(x) = \exp\left(-(x - \nu)^T A(x - \nu)\right),$$

where $\nu \in \mathbb{C}^d$ and where $A \in \mathbb{C}^{d \times d}$ is a positive definite Hermitian matrix. Suppose that $K \subseteq \mathbb{R}^d$ is a compact set. If Λ is a semigroup in $\mathbb{R}^d$ that contains a spanning set and Γ is a uniqueness set in $PW_{K-K}^1(\mathbb{R}^d)$, then every $f \in L^2(K)$ is determined up to a global phase by $|V_g f(\Lambda \times \Gamma)|$.

Proof. Let $\omega \in K - K$. Since $\operatorname{Re} A$ is positive definite, the window g is square-integrable. Following Theorem 3.2, we have to show that the system $\mathfrak{T}(g_\omega, \Lambda)$ is complete in $(C(K), \|\cdot\|_\infty)$. The map g_ω is given by

$$g_\omega(x) = C \cdot \exp\left(-2(x - (\operatorname{Re} \nu) - \frac{\omega}{2})^T (\operatorname{Re} A)(x - (\operatorname{Re} \nu) - \frac{\omega}{2})\right),$$

where $C = C(A, \omega, \nu)$ is a constant depending only on A, ω and ν . In view of Theorem 6.1, $\mathfrak{T}(g_\omega, \Lambda)$ is complete in $C(K)$. The statement follows from that. $\square$

Since every lattice is a semigroup that contains a spanning set, we obtain the following important consequence of Theorem 6.2.

Corollary 6.3. *Let $g \in C(\mathbb{R}^d)$ be a multivariate complex-valued Gaussian*

$$g(x) = \exp\left(-(x - \nu)^T A(x - \nu)\right),$$

where $\nu \in \mathbb{C}^d$ and where $A \in \mathbb{C}^{d \times d}$ is a positive definite Hermitian matrix. Suppose that $K \subseteq \mathbb{R}^d$ is a compact set. If $\Lambda, \Gamma \subseteq \mathbb{R}^d$ are lattices such that $K - K$ is contained in a fundamental domain of the reciprocal lattice Γ^* , then every $f \in L^2(K)$ is determined up to a global phase by $|V_g f(\Lambda \times \Gamma)|$.

Remark 6.4 (Zalik's theorem). Let $g(t) = e^{-at^2+bt+c}$, $a > 0$, $b, c \in \mathbb{R}$, be a univariate Gaussian, let $K \subseteq \mathbb{R}$ be a compact interval and let $\Lambda \subseteq \mathbb{R} \setminus \{0\}$ be a countable set. A theorem due to Zalik shows that $\mathfrak{T}(g, \Lambda)$ is complete in $L^2(K)$ if and only if

$$\sum_{\lambda \in \Lambda} \frac{1}{|\lambda|} = \infty,$$

see [72, Theorem 4]. For every $\omega \in \mathbb{R}$ it holds that g_ω is again a Gaussian of the form $t \mapsto e^{-\alpha t^2 + \beta t + \gamma}$ for some $\alpha > 0, \beta, \gamma \in \mathbb{R}$. By standard tensor arguments, one can show that a multivariate Gaussian of the form $g(t) = e^{-at \cdot t + b \cdot t + c}$, the system $\mathcal{T}(g, \Lambda^d)$ is complete in $L^2(K)$ for all Cartesian products of intervals K . Consequently, $\mathfrak{T}(g_\omega, \Lambda^d)$ is complete in $L^2(K)$ for every $\omega \in K - K \subseteq \mathbb{R}^d$. A combination of Zalik's theorem with Theorem 3.1 therefore implies a uniqueness statement for functions in $L^4(K)$. We refer to the articles [32, 70] for further references on the appearance of Zalik's theorem in STFT phase retrieval.

6.3. Uniqueness in dense subspaces. Note that Corollary 6.3 allows uniqueness when sampling the phaseless Gabor transform of a compactly supported $f \in L^2(K) \subseteq L^2(\mathbb{R}^d)$ on arbitrarily sparse lattices $\Lambda \times \Gamma$, since $D(\Lambda \times \Gamma) = \operatorname{vol}(\Lambda \times \Gamma)^{-1} = \operatorname{vol}(\Lambda)^{-1} \operatorname{vol}(\Gamma)^{-1}$, and we have complete freedom when choosing Λ . This is an improvement from band-limited windows or

Hermite windows, where both Λ and Γ need to be sufficiently dense. However, the sampling locations still depend on the compact set K through the choice of Γ . We now propose a trade-off for this dependency with a minor additional assumption on the compactly supported function $f \in L^2(\mathbb{R}^d)$. We recall the following theorem due to Kahane [44].

Theorem 6.5 (Kahane). *There exists a sequence $\mathcal{K} = \{\kappa_n : n \in \mathbb{Z}\} \subseteq \mathbb{R}$ such that the following three properties hold:*

- i) $\mathcal{K}$ is symmetric, that is, $\kappa_n = \kappa_{-n}$ for every $n \in \mathbb{Z}$,
- ii) $\lim_{n \rightarrow \infty} \frac{n}{\kappa_n} = 0$,
- iii) the system of complex exponentials $\{e^{2\pi i \kappa_n \cdot} : n \in \mathbb{Z}\}$ is complete in $C(I)$ for every compact interval $I \subseteq \mathbb{R}$.

The proof is constructive and the set $\mathcal{K}$ defined in [44] will from now on be referred to as *Kahane's sequence*. Notice that property ii) in Theorem 6.5 asserts that Kahane's sequence has density zero, where the density of a set $S \subseteq \mathbb{R}^d$ is defined via

$$D(S) := \lim_{r \rightarrow \infty} \frac{\#(S \cap B_r(0))}{m^d(B_r(0))},$$

assuming that the limit exists. Here m^d denotes the d -dimensional Lebesgue measure. By a tensor argument as with Zalik's theorem above, the system of multivariate complex exponentials $\{e^{2\pi i \kappa \cdot} : \kappa \in \mathcal{K}^d\}$ is complete in $C(K)$ for every compact set $K \subseteq \mathbb{R}^d$. To that, the density of the Cartesian product $\mathcal{K}^d$ is zero;

$$D(\mathcal{K}^d) = \lim_{r \rightarrow \infty} \frac{\#(\mathcal{K}^d \cap B_r(0))}{m^d(B_r(0))} \lesssim \lim_{r \rightarrow \infty} \prod_{j=1}^d \frac{\#(\mathcal{K} \cap [-r, r])}{2r} = 0.$$

We are now in a position to construct a sampling set of density zero which achieves phase retrieval in a dense subspace of $L^2(\mathbb{R}^d)$.

Theorem 6.6. *Let $g \in C(\mathbb{R}^d)$ be a real multivariate Gaussian window function. Let $\mathcal{K} \subseteq \mathbb{R}$ be Kahane's sequence and $\Lambda \subseteq \mathbb{R} \setminus \{0\}$ a sequence of distinct real numbers such that $\sum_{\lambda \in \Lambda} \frac{1}{|\lambda|} = \infty$ and such that $D(\Lambda) = 0$. Further, let*

$$\mathcal{V} := \bigcup_{\kappa > 0} L^4[-\kappa, \kappa]^d.$$

Then the following holds:

- i) $\mathcal{V}$ is dense in $L^2(\mathbb{R}^d)$,
- ii) every $f \in \mathcal{V}$ is determined up to a global phase by $|V_g f(\Lambda^d \times \mathcal{K}^d)|$,
- iii) $D(\Lambda^d \times \mathcal{K}^d) = 0$.

In particular, the theorem's conclusion holds true if Λ is the set of prime numbers.

Proof. It is evident that $\mathcal{V}$ is dense in $L^2(\mathbb{R}^d)$. Therefore, we continue with the proof of the second statement. To that end, let $K \subseteq \mathbb{R}^d$ be an arbitrary compact set. Since g is a Gaussian, so is g_ω for every $\omega \in \mathbb{R}^d$. In view

of Zalik's theorem (see Remark 6.4), the system $\mathfrak{T}(g_\omega, \Lambda^d)$ is complete in $L^2(K)$. In particular, the operator

$$\mathcal{C}_\omega : L^2(K) \rightarrow \mathbb{C}^{\Lambda^d}, \quad f \mapsto \left(\langle f, \overline{T_\lambda g_\omega} \rangle_{L^2(K)} \right)_{\lambda \in \Lambda^d}$$

is injective for every $\omega \in K - K$. According to Kahane's theorem, the system $\{e^{2\pi i \kappa \cdot} : \kappa \in \mathcal{K}^d\}$ is complete in $C(K - K)$. In particular, it is complete in $L^2(K - K)$, which in turn implies that $\mathcal{K}^d$ is a uniqueness set for $PW_{K-K}^2(\mathbb{R}^d) \supseteq PW_{K-K}^1(\mathbb{R}^d)$. The statement follows from the observation that $\Lambda^d, \mathcal{K}^d$ are independent of K and if $f, h \in \mathcal{V}$, then there exists a compact set $K \subseteq \mathbb{R}^d$ such that $\text{supp}(f), \text{supp}(h) \subseteq K$ and $f, h \in L^4(K)$. The property that $D(\Lambda^d \times \mathcal{K}^d) = 0$ follows from

$$\frac{\#((\Lambda^d \times \mathcal{K}^d) \cap B_r(0))}{m^{2d}(B_r(0))} \lesssim \frac{\#(\Lambda^d \cap B_r(0))\#(\mathcal{K}^d \cap B_r(0))}{r^{2d}} \rightarrow 0$$

as $r \rightarrow \infty$ (where on the left-hand side and right-hand side $B_r(0)$ denotes a ball in $\mathbb{R}^{2d}$ and in $\mathbb{R}^d$, respectively). $\square$

7. APPENDIX: PROOF OF PROPOSITION 4.2

Proof. We divide the proof into six steps. The main part considers the case $A = I$. From there, we first handle non-singular matrices and finish with the singular case. Observe that if $\varphi(z) = e^{-z^T A z + z^T b + c}$, then the statement holds true, so it suffices to show that the existence of a factorization of the form (6) for φ implies that φ has the form (7).

Step 1. We start by showing that φ is zero-free. To that end, assume the contrary, that is, the function φ has a zero at $z_0 \in \mathbb{C}^d$. Then z_0 is also a zero of ψ or σ . Without loss of generality, $\psi(z_0) = 0$. Then for all $z \in \mathbb{C}^d$ it holds that

$$\varphi(z) = \varphi(z_0 + z - z_0) = \psi(z_0)\sigma(z - z_0)e^{-2z_0^T A(z - z_0)} = 0,$$

contradicting the assumption that φ is not the zero function. Therefore, none of the functions φ, ψ, σ has a zero. Furthermore, observe that it holds $\varphi = \sigma(0)\psi = \psi(0)\sigma$. Hence, ψ and σ are entire functions. This also implies

$$\varphi(z)\varphi(\lambda)e^{-2z^T A \lambda} = \sigma(0)\psi(0)\varphi(z + \lambda) = \varphi(\lambda)\varphi(z)e^{-2\lambda^T A z},$$

which allows us to conclude that all $\lambda, z \in \mathbb{C}^d$ satisfy $1 = e^{2z^T(A^T - A)\lambda}$. Clearly, this can only be true if and only if $A = A^T$, i.e., A is symmetric.

Step 2: $A = \text{Id}$. Assume in the following that A is the identity matrix. Since

$$e^{-2z^T \lambda} \psi(z)\sigma(\lambda) = \varphi(z + \lambda) = \varphi(\lambda + z) = \psi(\lambda)\sigma(z)e^{-2\lambda^T z} = \psi(\lambda)\sigma(z)e^{-2z^T \lambda},$$

we conclude that there exists a constant $\eta \in \mathbb{C} \setminus \{0\}$ such that $\psi = \eta\sigma$. Therefore,

$$\begin{aligned} \varphi(z + \lambda) &= \psi(z)\sigma(\lambda)e^{-2z^T \lambda} = \frac{1}{\eta}\psi(z)\psi(\lambda)e^{-2z^T \lambda} \\ (9) \quad &= \frac{1}{\sqrt{\eta}}\psi(z)\frac{1}{\sqrt{\eta}}\psi(\lambda)e^{-2z^T \lambda}. \end{aligned}$$

Replacing ψ , σ by $\frac{1}{\sqrt{\eta}}\psi$ we can assume without loss of generality that $\eta = 1$. This implies that for each $z \in \mathbb{C}^d$ we have

$$\varphi(0) = \varphi(z - z) = \psi(z)\psi(-z)e^{2z^T z} = \Psi(z)\Psi(-z),$$

where $\Psi(z) := \psi(z)e^{z^T z}$.

Step 3: Projection on the coordinate axes. Let e_j be the j -th unit vector in $\mathbb{C}^d$. Further, let $\Psi_j(t) := \Psi(te_j)$ be the restriction of Ψ to the j -th coordinate axis. Since Ψ_j is zero-free and entire, there exists an entire function g_j such that $\Psi_j = e^{g_j}$. Equation (6) together with the previous considerations implies that for all $t, \xi \in \mathbb{C}$ holds

$$(10) \quad \varphi((t + \xi)e_j) = \Psi_j(t)\Psi_j(\xi)e^{-(t+\xi)^2} = e^{g_j(t)}e^{g_j(\xi)}e^{-(t+\xi)^2}.$$

On the other hand, it follows from the identity $\varphi = \psi(0)\sigma = \psi(0)\psi$ and the definition of Ψ_j that

$$(11) \quad \varphi((t + \xi)e_j) = \psi(0)\Psi_j(t + \xi)e^{-(t+\xi)^2} = \psi(0)e^{g_j(t+\xi)}e^{-(t+\xi)^2}.$$

Now fix $\xi \in \mathbb{C}$. Combining Equation (10) and (11) shows that there exists a constant $s \in \mathbb{C}$ such that

$$s = g_j(t) + g_j(\xi) - g_j(t + \xi)$$

for all $t \in \mathbb{C}$. Differentiating with respect to t yields

$$0 = g'_j(t) + 0 - g'_j(t + \xi).$$

Thus, g'_j is ξ -periodic, and since ξ was arbitrary, g'_j must be a constant function. Therefore, there exist constants $b_j, c_j \in \mathbb{C}$ satisfying

$$g_j(t) = b_j t + c_j, \quad t \in \mathbb{C}.$$

In particular, the map Ψ_j is given by

$$\Psi_j(t) = e^{b_j t + c_j}.$$

Since $\Psi(0) = \Psi_j(0) = e^{c_j}$, it holds $c_j - c_k \in 2\pi i\mathbb{Z}$ for all $j, k \in \{1, \dots, d\}$. Therefore, we can assume that $c := c_j$ for all $j \in \{1, \dots, d\}$.

Step 4: Derivation of φ for $A = \text{Id}$. Decomposing the vector $z = (z_1, \dots, z_d)^T \in \mathbb{C}^d$ with respect to the standard unit vectors $e_1, \dots, e_d \in \mathbb{C}^d$, and using the factorization derived in Equation (9) with $\eta = 1$ shows that

$$\varphi(z) = \varphi\left(z_1 e_1 + \sum_{j=2}^d z_j e_j\right) = \psi(z_1 e_1)\psi\left(\sum_{j=2}^d z_j e_j\right).$$

Since $\psi(z_j e_j) = e^{-z_j^2 + b_j z_j + c}$ and $\varphi = \psi(0)\psi = e^c \psi$,

$$\varphi(z) = e^{-z_1^2 + b_1 z_1} \varphi\left(\sum_{j=2}^d z_j e_j\right).$$

It follows via induction that

$$\begin{aligned} \varphi(z) &= \psi_d(z_d e_d) \prod_{j=1}^{d-1} e^{-z_j^2 + b_j z_j} = e^c \prod_{j=1}^d e^{-z_j^2 + b_j z_j} \\ &= \exp\left(-z^T z + z^T b + c\right). \end{aligned}$$

Step 5: Non-singular A . Now assume A is an arbitrary invertible symmetric matrix. By Takagi's factorization theorem, there exists an invertible matrix $B \in \mathbb{C}^{d \times d}$ satisfying $B^T B = A$. This transforms the problem to

$$\varphi(z + \lambda) = \psi(z) \sigma(\lambda) e^{-2z^T B^T B \lambda} = \psi(B^{-1} B z) \sigma(B^{-1} B \lambda) e^{-2(Bz)^T (B\lambda)}.$$

If we define the auxiliary functions φ_B , ψ_B , and σ_B via

$$\varphi_B := \varphi \circ B^{-1}, \quad \psi_B := \psi \circ B^{-1}, \quad \sigma_B := \sigma \circ B^{-1},$$

then it holds that

$$\varphi_B(z + \lambda) = \psi(B^{-1} z) \sigma(B^{-1} \lambda) e^{-2z \cdot \lambda} = \psi_B(z) \sigma_B(\lambda) e^{-2z^T \lambda}.$$

By Step 4 above, there is a vector $b' \in \mathbb{C}^d$ and a constant $c \in \mathbb{C}$ such that

$$\varphi_B(z) = \exp\left(-z^T z + z^T b' + c\right)$$

for all $z \in \mathbb{C}^d$. Therefore, denoting $b = B^T b'$, for all $z \in \mathbb{C}^d$ we have

$$\varphi(z) = \varphi_B(Bz) = \exp\left(-z^T A z + z^T b + c\right).$$

Step 6: General symmetric matrices. Finally, suppose that A is an arbitrary symmetric matrix. Let $A = U \Sigma U^T$ be the singular value decomposition of A with $U \in \mathbb{C}^{d \times d}$ unitary and $\Sigma \in \mathbb{R}^{d \times d}$ diagonal, $\Sigma_{i,i} \geq \Sigma_{i+1,i+1}$. Further, let Ω be the diagonal matrix defined by

$$\Omega_{i,i} := \begin{cases} 0, & \Sigma_{i,i} \neq 0, \\ 1, & \Sigma_{i,i} = 0. \end{cases}$$

Then the matrix

$$C := A + U \Omega U^T = U (\Sigma + \Omega) U^T$$

is invertible and it holds

$$\underbrace{\varphi(z + \lambda) e^{-(z+\lambda)^T U \Omega U^T (z+\lambda)}}_{:=\varphi_0(z+\lambda)} = \underbrace{\psi(z) e^{-z^T U \Omega U^T z}}_{:=\psi_0(z)} \underbrace{\sigma(\lambda) e^{-\lambda^T U \Omega U^T \lambda}}_{:=\sigma_0(\lambda)} e^{-2z^T C \lambda}.$$

By the previous step, there exist $b \in \mathbb{C}^d$ and $c \in \mathbb{C}$ such that

$$\varphi_0(z) = \exp\left(-z^T C z + z^T b + c\right).$$

By the definition of the map φ_0 and the matrix C , we obtain

$$\varphi(z) = e^{-z^T A z + z^T b + c}.$$

This completes the proof of the statement. $\square$

ACKNOWLEDGEMENTS

Irina Shafkulovska was supported by the Austrian Science Fund (FWF) project P33217 and thanks the Austrian Academy of Sciences for hosting an internship at the Acoustics Research Institute.

REFERENCES

- [1] ABDELMALEK, R., MNASRI, Z., AND BENZARTI, F. Audio signal reconstruction using phase retrieval: Implementation and evaluation. *Multimed. Tools Appl.* **81**, 11 (2022), 15919–15946. doi.org/10.1007/s11042-022-12421-1.
- [2] AKUTOWICZ, E. J. On the Determination of the Phase of a Fourier Integral, I. *Trans. Amer. Math. Soc.* **83**, 1 (1956), 179–192. doi.org/10.2307/1992910.
- [3] AKUTOWICZ, E. J. On the Determination of the Phase of a Fourier Integral, II. *Proc. Amer. Math. Soc.* **8**, 2 (1957), 234–238. doi.org/10.2307/2033718.
- [4] ALAIFARI, R., BARTOLUCCI, F., STEINERBERGER, S., AND WELLERSHOFF, M. On the connection between uniqueness from samples and stability in Gabor phase retrieval, 2022. arxiv.org/abs/2205.07013.
- [5] ALAIFARI, R., DAUBECHIES, I., GROHS, P., AND THAKUR, G. Reconstructing Real-Valued Functions from Unsigned Coefficients with Respect to Wavelet and Other Frames. *J. Fourier Anal. Appl.* **23** (2017), 1480–1494. doi.org/10.1007/s00041-016-9513-7.
- [6] ALAIFARI, R., DAUBECHIES, I., GROHS, P., AND YIN, R. Stable Phase Retrieval in Infinite Dimensions. *Found. Comput. Math.* **19** (2019), 869–900. doi.org/10.1007/s10208-018-9399-7.
- [7] ALAIFARI, R., AND GROHS, P. Phase Retrieval In The General Setting Of Continuous Frames For Banach Spaces. *SIAM J. Math. Anal.* **49** (2016). doi.org/10.1137/16M1071481.
- [8] ALAIFARI, R., AND WELLERSHOFF, M. Phase Retrieval from Sampled Gabor Transform Magnitudes: Counterexamples. *J. Fourier Anal. Appl.* **28**, 1 (2021), 9. doi.org/10.1016/10.1007/s00041-021-09901-7.
- [9] ALAIFARI, R., AND WELLERSHOFF, M. Uniqueness of STFT phase retrieval for bandlimited functions. *Appl. Comput. Harmon. Anal.* **50** (2021), 34–48. doi.org/10.1016/j.acha.2020.08.003.
- [10] ASCENSI, G., AND BRUNA, J. Model Space Results for the Gabor and Wavelet Transforms. *IEEE Trans. Inform. Theory* **55** (2009), 2250 – 2259. doi.org/10.1109/TIT.2009.2016028.
- [11] ATZMON, A., AND OLEVSKII, A. Completeness of Integer Translates in Function Spaces on $\mathbb{R}$. *J. Approx. Theory* **87**, 3 (1996), 291–327. doi.org/10.1006/jath.1996.0106.
- [12] AUSLANDER, L., AND TOLIMIERI, R. Radar Ambiguity Functions and Group Theory. *SIAM J. Math. Anal.* **16**, 3 (1985), 577–601. doi.org/10.1137/0516043.
- [13] BERENSTEIN, C. A., AND TAYLOR, B. A. Mean-periodic functions. *Internat. J. Math. Math. Sci.* **3** (1980), 199–235. doi.org/10.1155/S0161171280000154.
- [14] BHATTACHARYYA, P. K. *Distributions*. De Gruyter, Berlin, Boston, 2012. doi.org/10.1515/9783110269291.
- [15] BOCHE, H., POHL, V., AND YANG, F. Phaseless Signal Recovery in Infinite Dimensional Spaces Using Structured Modulations. *J. Fourier Anal. Appl.* **20**, 6 (2014), 1212–1233. doi.org/10.1007/s00041-014-9352-3.
- [16] BORN, M., AND WOLF, E. *Principles of Optics: 60th Anniversary Edition*, 7 ed. Cambridge University Press, 2020.
- [17] BRUNA, J., OLEVSKII, A., AND ULANOVSKII, A. Completeness in $L^1(\mathbb{R})$ of discrete translates. *Rev. Mat. Iberoam.* **22**, 1 (2006), 1 – 16.
- [18] CAHILL, J., CASAZZA, P. G., AND DAUBECHIES, I. Phase retrieval in infinite-dimensional Hilbert spaces. *Trans. Amer. Math. Soc. Ser. B* **3** (2016), 63–76. doi.org/10.1090/btran/12.
- [19] DA SILVA, J. C., AND MENZEL, A. Elementary signals in ptychography. *Opt. Express* **23**, 26 (2015), 33812–33821. doi.org/10.1364/OE.23.033812.
- [20] DELSARTE, J. Les fonctions moyenne-périodiques. *J. Math. Pures et Appl.* **14** (1935), 403–453.
- [21] DELSARTE, J. Theorie des Fonctions Moyenne-Periodiques de Deux Variables. *Ann. of Math* **72** (1960), 121–178.

- [22] FANNJIANG, A., AND STROHMER, T. The Numerics of Phase Retrieval. *Acta Numer.* 29 (2020), 125–228. doi.org/10.1017/S0962492920000069.
- [23] FOLLAND, G. B. *Real Analysis: Modern Techniques and Their Applications*. Wiley, 1984.
- [24] FORSTNER, A., KRAHMER, F., MELNYK, O., AND SISSOUNO, N. Well-conditioned ptychographic imaging via lost subspace completion. *Inverse Probl.* 36, 10 (2022). doi.org/10.1088/1361-6420/abaf3a.
- [25] GRAFAKOS, L. *Classical Fourier Analysis*. Springer New York, 2014. doi.org/10.1007/978-1-4939-1194-3.
- [26] GRÖCHENIG, K. The Mystery of Gabor Frames. *J. Fourier Anal. Appl.* 20 (2014), 865–895. doi.org/10.1007/s00041-014-9336-3.
- [27] GRÖCHENIG, K., AND KOPPENSTEINER, S. *Gabor Frames: Characterizations and Coarse Structure*. Springer International Publishing, Cham, 2019, pp. 93–120. doi.org/10.1007/978-3-030-32353-0_4.
- [28] GROHS, P., KOPPENSTEINER, S., AND RATHMAIR, M. Phase Retrieval: Uniqueness and Stability. *SIAM Rev.* 62, 2 (2020), 301–350. doi.org/10.1137/19M1256865.
- [29] GROHS, P., AND LIEHR, L. Stable Gabor phase retrieval in Gaussian shift-invariant spaces via biorthogonality. to appear in: *Constr. Approx.* (2021). arxiv.org/abs/2109.02494.
- [30] GROHS, P., AND LIEHR, L. Non-uniqueness theory in sampled STFT phase retrieval. *arXiv* (2022). doi.org/10.48550/arXiv.2207.05628.
- [31] GROHS, P., AND LIEHR, L. On Foundational Discretization Barriers in STFT Phase Retrieval. *J. Fourier Anal. Appl.* 28, 39 (2022). doi.org/10.1007/s00041-022-09935-5.
- [32] GROHS, P., AND LIEHR, L. Injectivity of Gabor phase retrieval from lattice measurements. *Appl. Comput. Harmon. Anal.* 62 (2023), 173–193. doi.org/10.1016/j.acha.2022.09.001.
- [33] GROHS, P., LIEHR, L., AND RATHMAIR, M. Multi-window STFT phase retrieval: lattice uniqueness, 2022. doi.org/10.48550/ARXIV.2207.10620.
- [34] GROHS, P., AND RATHMAIR, M. Stable Gabor Phase Retrieval and Spectral Clustering. *Comm. Pure Appl. Math.* 72, 5 (2019), 981–1043. doi.org/10.1002/cpa.21799.
- [35] GROHS, P., AND RATHMAIR, M. Stable Gabor phase retrieval for multivariate functions. *J. Eur. Math. Soc.* (2021). doi.org/10.4171/JEMS/1114.
- [36] GRÖCHENIG, K. *Foundations of Time-Frequency Analysis*. Birkhäuser Basel, 2001. doi.org/10.1007/978-1-4612-0003-1.
- [37] HARDY, G. H. On two theorems of F. Carlson and S. Wigert. *Acta Math.* 42 (1920), 327 – 339. doi.org/10.1007/BF02404414.
- [38] HIGGINS, J. R. *Sampling Theory in Fourier and Signal Analysis: Foundations*. Clarendon Press, 2001.
- [39] HOFSTETTER, E. Construction of time-limited functions with specified autocorrelation functions. *IEEE Trans. Inform. Theory* 10, 2 (1964), 119–126. doi.org/10.1109/TIT.1964.1053648.
- [40] HÖRMANDER, L. *The analysis of linear partial differential operators I*, 2 ed. Classics in mathematics. Springer, Berlin, Germany, 2003. doi.org/10.1007/978-3-642-61497-2.
- [41] JAMING, P., KELLAY, K., AND PEREZ, R. On the effect of zero-flipping on the stability of the phase retrieval problem in the Paley-Wiener class. *Monatsh. Math.* 198, 4 (2022), 757–776. doi.org/10.1007/s00605-022-01716-y.
- [42] JIANG, Y., CHEN, Z., HAN, Y., DEB, P., GAO, H., XIE, S., PUROHIT, P., TATE, M. W., PARK, J., GRUNER, S. M., ELSE, V., AND MULLER, D. A. Electron ptychography of 2d materials to deep sub-ångström resolution. *Nature* 559, 7714 (2018), 343–349. doi.org/10.1038/s41586-018-0298-5.
- [43] KAHANE, J. *Lectures on On Mean Periodic Functions*. Tata Institute of Fundamental Research, 1959.
- [44] KAHANE, J.-P. Sur la totalité des suites d’exponentielles imaginaires. *Ann. Inst. Fourier* 8 (1958), 273–275.

- [45] KALKER, T. On Multidimensional Sampling. In *Digital Signal Processing Fundamentals* (2009). doi.org/10.1201/9781420046076.
- [46] LAI, C.-K., LITTMANN, F., AND WEBER, E. S. Conjugate Phase Retrieval in Paley–Wiener Space. *J. Fourier Anal. Appl.* 27, 6 (2021), 89. doi.org/10.1007/s00041-021-09891-6.
- [47] LANDAU, H. On the completeness of a set of translates. *J. Approx. Theory* 5, 4 (1972), 438–440. doi.org/10.1016/0021-9045(72)90011-1.
- [48] LEV, N., AND OLEVSKII, A. Wiener’s ‘closure of translates’ problem and Piatetski-Shapiro’s uniqueness phenomenon. *Ann. of Math.* 174 (2011), 519–541. doi.org/10.4007/annals.2011.174.1.15.
- [49] LOGVINENKO, V. N., AND FAVOROV, S. Y. Cartwright-type theorems and real uniqueness sets for entire functions of exponential type. *Math. Notes* 51, 3 (1993), 294–299. doi.org/10.1007/bf01207716.
- [50] LYUBARSKII, Y. Frames in the Bargmann space of entire functions. In *Entire and Subharmonic Functions*. American Mathematical Society, Providence, RI, 1992, p. 167–180.
- [51] MELNYK, O. On connections between Amplitude Flow and Error Reduction for phase retrieval and ptychography. *Sampl. Theory Signal Process. Data Anal.* 20, 2 (2022), 16. doi.org/10.1007/s43670-022-00035-5.
- [52] OLEVSKII, A., AND ULANOVSKII, A. Almost Integer Translates. Do Nice Generators Exist? *J. Fourier Anal. Appl.* 10, 1 (2004), 93–104. doi.org/10.1007/s00041-004-8006-2.
- [53] OLEVSKII, A., AND ULANOVSKII, A. Discrete Translates in Function Spaces. *Anal. Math.* 44, 2 (2018), 251–261. doi.org/10.1007/s10476-018-0209-x.
- [54] OLEVSKII, A., AND ULANOVSKII, A. Discrete translates in $L^p(\mathbb{R})$. *Bull. Lond. Math. Soc.* 50 (2018), 561–568. doi.org/10.1112/blms.12159.
- [55] PINKUS, A. Density in Approximation Theory. *Surv. Approx. Theory* 1 (2005), 1–45. eudml.org/doc/51470.
- [56] RONKIN, L. I. The completeness of the system of functions $e^{i\langle\lambda,x\rangle}$ and the real sets of uniqueness of entire functions of several variables. *Funct. Anal. Its Appl.* 5 (1971), 343 – 344. doi.org/10.1007/BF01086755.
- [57] RONKIN, L. I. On real sets of uniqueness for entire functions of several variables and on the completeness of the system of functions $e^{i\langle\lambda,x\rangle}$. *Sib. Math. J.* 13 (1972), 439 – 443. doi.org/10.1007/BF00968121.
- [58] RONKIN, L. I. Some questions of completeness and uniqueness for functions of several variables. *Funct. Anal. Appl.* 7 (1973), 370. doi.org/10.1007/BF00969587-55. doi.org/10.1007/bf01075648.
- [59] RUDIN, W. *Real and Complex Analysis*, 3rd edition ed. McGraw-Hill, 1986.
- [60] SCHWARTZ, L. Theorie Generale des Fonctions Moyenne-Periodiques. *Ann. of Math.* 48, 4 (1947), 857–929. doi.org/10.2307/1969386.
- [61] SEIP, K. Density theorems for sampling and interpolation in the Bargmann – Fock space I. *J. für die Reine und Angew. Math.* 429 (1992), 91–106. doi.org/10.1515/crll.1992.429.107.
- [62] SEIP, K., AND WALLSTÉN, R. Density theorems for sampling and interpolation in the Bargmann – Fock space II. *J. für die Reine und Angew. Math.* 429 (1992), 107–114. doi.org/10.1090/S0273-0979-1992-00290-2.
- [63] STEINHAUS, H. Sur les distances des points dans les ensembles de mesure positive. *Fund. Math.* 1, 1 (1920), 93–104. doi.org/10.4064/fm-1-1-93-104.
- [64] STONE, M. H. Applications of the theory of Boolean rings to general topology. *Trans. Amer. Math. Soc.* 41, 3 (1937), 375–481. doi.org/10.1090/s0002-9947-1937-1501905-7.
- [65] STONE, M. H. The Generalized Weierstrass Approximation Theorem. *Math. Mag.* 21, 4 (1948), 167. doi.org/10.2307/3029750.
- [66] THAKUR, G. Reconstruction of Bandlimited Functions from Unsigned Samples. *J. Fourier Anal. Appl.* 17 (2011), 720–732. doi.org/10.1007/s00041-010-9144-3.
- [67] WALTHER, A. The Question of Phase Retrieval in Optics. *Optica Acta Int. J. Opt.* 10, 1 (1963), 41–49. doi.org/10.1080/713817747.

- [68] WEIERSTRASS, K. Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen. *Sitzungsberichte der Königl. Preußischen Akademie der Wissenschaften zu Berlin* (1885), 633–639.
- [69] WEIL, A. *L'intégration dans les groupes topologiques et ses applications*. Actualités scientifiques et industrielles. Hermann, 1941.
- [70] WELLERSHOFF, M. Injectivity of sampled Gabor phase retrieval in spaces with general integrability conditions, 2021. arxiv.org/abs/2112.10136.
- [71] WIENER, N. Tauberian Theorems. *Ann. of Math.* 33, 1 (1932), 1–100. doi.org/10.2307/1968102.
- [72] ZALIK, R. A. On approximations by shifts and a theorem of Wiener. *Trans. Amer. Math. Soc.* 243 (1978). doi.org/10.1090/S0002-9947-1978-0493077-1.
- [73] ZHANG, F., CHEN, B., MORRISON, G. R., VILA-COMAMALA, J., GUIZAR-SICAIROS, M., AND ROBINSON, I. K. Phase retrieval by coherent modulation imaging. *Nat. Commun.* 7, 1 (2016), 13367. doi.org/10.1038/ncomms13367.

3.6 Benedicks-type uncertainty principle for metaplectic time-frequency representations

Bibliographic Information

Gröchenig, K., Shafkulovska, I., Benedicks-type uncertainty principle for metaplectic time-frequency representations (2024)

Contribution

Karlheinz Gröchenig and Irina Shafkulovska contributed equally to the idea, concept, development, and writing of the paper.

BENEDICKS-TYPE UNCERTAINTY PRINCIPLE FOR METAPLECTIC TIME-FREQUENCY REPRESENTATIONS

KARLHEINZ GRÖCHENIG* AND IRINA SHAFKULOVSKA*

ABSTRACT. Metaplectic time-frequency representations are a special class of joint time-frequency representations which arise from metaplectic operators and are a generalization of the short-time Fourier transform. Given a metaplectic time-frequency representation, we investigate whether there exist non-zero functions whose metaplectic time-frequency representation is supported on a set of finite measure. We study this both in the sesquilinear and the quadratic version and provide a full characterization of the class of metaplectic time-frequency representations which exhibit an uncertainty principle of this type.

Keywords. uncertainty principle, metaplectic time-frequency representations, support of time-frequency representations, symplectic group, metaplectic group, oscillator representation

AMS subject classifications. 81S07, 81S30, 22E46

1. INTRODUCTION

The uncertainty principle refers to a huge collection of inequalities that assert that a function and its Fourier transform cannot be small simultaneously. In principle, one can formulate an uncertainty principle for every measure for the size of a function. Taking second moments $\Delta f = \int_{\mathbb{R}} x^2 |f(x)|^2 dx$ of a function f on the real line, the classical uncertainty principle, the Heisenberg-Pauli-Weyl inequality, asserts that

$$\Delta f \Delta \hat{f} \geq \frac{1}{16\pi^2} \|f\|_2^4$$

when we use the normalization $\hat{f}(\omega) = \mathcal{F}f(\omega) = \int_{\mathbb{R}} f(x) e^{-2\pi i x \omega} dx$ for the Fourier transform of f . If the size of f is measured by the support, one obtains a qualitative uncertainty principle: If both f and $\hat{f}$ have compact support, then f is the zero function. This is easy to see from the definition and analyticity properties of the Fourier transform. A less obvious version is Benedicks's uncertainty principle [1, 3]:

*If both f and $\hat{f}$ have a support of finite measure,
then f is the zero function.*

For excellent surveys of uncertainty principles that are formulated for the Fourier pair $(f, \hat{f})$, we recommend the surveys [6, 16, 30].

*FACULTY OF MATHEMATICS, UNIVERSITY OF VIENNA, OSKAR-MORGENSTERN-PLATZ 1, 1090 VIENNA, AUSTRIA

E-mail addresses: karlheinz.groechenig@univie.ac.at, irina.shafkulovska@univie.ac.at.

I. Shafkulovska was funded in whole or in part by the Austrian Science Fund (FWF) [10.55776/P33217]. For open access purposes, the authors have applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission.

For physicists and engineers, f and $\hat{f}$ are merely two representations of the same object, e.g., a quantum mechanical state or a signal, and they often prefer to study it in the time-frequency plane or in phase space via a joint time-frequency representation. The basic time-frequency representations are the short-time Fourier transform, the ambiguity function, and the Wigner distribution. The short-time Fourier transform of a function f on $\mathbb{R}^d$ with respect to a window g is defined as

$$V_g f(x, \omega) = \int_{\mathbb{R}^d} f(t) \bar{g}(t - x) e^{-2\pi i \omega \cdot t} dt,$$

the (cross-) ambiguity function of f and g is

$$A(f, g)(x, \omega) = \int_{\mathbb{R}^d} f(t + \frac{x}{2}) \overline{g(t - \frac{x}{2})} e^{-2\pi i \omega \cdot t} dt,$$

and the (cross-) Wigner distribution of f and g is

$$W(f, g)(x, \omega) = \int_{\mathbb{R}^d} f(x + \frac{t}{2}) \overline{g(x - \frac{t}{2})} e^{-2\pi i \omega \cdot t} dt.$$

All three time-frequency representations carry information simultaneously about f and $\hat{f}$, and therefore one expects versions of the uncertainty principle for these time-frequency representations as well. Indeed, Benedicks's uncertainty principle possesses the following formulation [20, 25, 27, 38].

If $V_g f$ is supported on a set of finite measure, then $f \equiv 0$ or $g \equiv 0$. Likewise, if $A(f, g)$ or $W(f, g)$ is supported on a set of finite measure, then $f \equiv 0$ or $g \equiv 0$.

There is an almost systematic technique to translate an uncertainty principle from the Fourier pair $(f, \hat{f})$ to an uncertainty principle of a joint time-frequency representation, e.g., $V_g f$, see the survey [19].

The short-time Fourier transform, the ambiguity function and the Wigner distribution are undoubtedly the most important and most frequently used time-frequency representations, as they are used for the definition of function spaces [13, 21] and in the calculus of pseudodifferential operators [4, 8, 15, 21], and of course in many applications in signal processing [7, 14, 18, 32, 35, 39] and in the phase space formulation of quantum mechanics [12]. Yet there is a whole zoo of equally interesting time-frequency representations available [23]. Further extensions were studied as linear perturbations of the Wigner distribution [2], as canonical linear transforms [41], free metaplectic transformations [40], and as symplectic Wigner distributions [9].

One expects many uncertainty principles to carry over to this huge class of time-frequency representations. This is the topic of our work.

To set up the results, we first recall the symplectic group and the associated metaplectic representation. Let

$$(1.1) \quad \mathcal{J} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \in \mathbb{R}^{2d \times 2d}$$

be the standard skew-symmetric matrix in $\mathrm{GL}(2d, \mathbb{R})$. Hereby, $0 = 0_{d \times d}$ and $I = I_d$ denote the zero and identity matrix in $\mathbb{R}^{d \times d}$ respectively¹. Then the symplectic group $\mathrm{Sp}(2d, \mathbb{R})$ consists of all $2d \times 2d$ -matrices $\mathcal{A}$ such that $\mathcal{A}^t \mathcal{J} \mathcal{A} = \mathcal{J}$. Associated to $\mathcal{A} \in \mathrm{Sp}(2d, \mathbb{R})$ is a unitary operator $\mu(\mathcal{A})$ which intertwines the symmetric phase-space shifts and acts on $L^2(\mathbb{R}^d)$ [12, 15, 21].

Now we double the dimension and assume that $\mathcal{A} \in \mathrm{Sp}(4d, \mathbb{R})$. Then $\mu(\mathcal{A})$ acts on $L^2(\mathbb{R}^{2d})$, and in particular on the elementary tensors $(f \otimes g)(x, y) = f(x)g(y)$ for $f, g \in L^2(\mathbb{R}^d)$. Motivated by a factorization of the short-time Fourier transform and the Wigner distribution [21], one can introduce a time-frequency representation as

$$W_{\mathcal{A}}(f, g) = \mu(\mathcal{A})(f \otimes \bar{g}) \quad f, g \in L^2(\mathbb{R}^d).$$

Following [9, 10], we call this time-frequency representation a metaplectic time-frequency representation or an $\mathcal{A}$ -Wigner distribution. All common time-frequency representations, in particular, the short-time Fourier transform, the ambiguity function, all τ -Wigner distributions, the canonical linear transforms, and their extension in [2] can be written as a symplectic time-frequency representation with a particular $4d \times 4d$ symplectic matrix $\mathcal{A}$. The question arises whether the usual uncertainty principles can be extended and be formulated for a general symplectic time-frequency representation. One would expect an affirmative answer, but after checking a few examples of symplectic matrices and their metaplectic operators, one begins to understand that this question leads to an interesting problem, whose answer is by far not obvious.

Examples. Since the intertwining operators $\mu(\mathcal{A}) \in \mathrm{Sp}(2d, \mathbb{R})$ are unique up to a multiplication with a global phase $\tau \in \mathbb{T}$ [12, 15, 21], we omit the global phase in the examples below.

(i) Let $\mathcal{A} = \mathcal{J}$ be the standard symplectic matrix in $\mathrm{Sp}(4d, \mathbb{R})$. Then $\mu(\mathcal{J})F = \hat{F}$ for $F \in L^2(\mathbb{R}^{2d})$ is the Fourier transform. Evidently, $W_{\mathcal{J}}(f \otimes \bar{g}) = \hat{f} \otimes \hat{g}$ has compact support (or support of finite measure) if and only if both $\hat{f}$ and $\hat{g}$ have compact support (support of finite measure) in $\mathbb{R}^d$. So there is no Benedicks-type uncertainty principle.

(ii) Let $L \in \mathrm{GL}(2d, \mathbb{R})$ and

$$\mathcal{A} = \begin{pmatrix} L & 0 \\ 0 & L^{-t} \end{pmatrix} \in \mathrm{Sp}(4d, \mathbb{R}).$$

The corresponding metaplectic operator is given by a unitary coordinate transform $\mu(\mathcal{A})f(\lambda) = |\det L|^{-1/2} F(L^{-1}\lambda)$ for $\lambda \in \mathbb{R}^{2d}$ and $F \in L^2(\mathbb{R}^{2d})$. Clearly, a coordinate transform changes the support and its measure, but the compactness and the finiteness of the measure are preserved. Thus, if we assume that the time-frequency representation $W_{\mathcal{A}}(f \otimes \bar{g})$ has compact support, then also $f \otimes g$ has compact support and every pair (f, g) of compactly supported functions in $L^2(\mathbb{R}^d)$ yields a counter-example to the analogous version of Benedicks's theorem.

¹We include the dimension in the index when necessary but will otherwise omit it if it is clear from the context.

(iii) Let $Q = Q^t \in \mathrm{GL}(2d, \mathbb{R})$ be symmetric and

$$\mathcal{A} = \begin{pmatrix} I & 0 \\ Q & I \end{pmatrix} \in \mathrm{Sp}(4d, \mathbb{R}).$$

Then $\mu(\mathcal{A})F(\lambda) = e^{i\pi\lambda \cdot Q\lambda}F(\lambda)$ for $F \in L^2(\mathbb{R}^{2d})$ amounts to the multiplication by a “chirp”. Thus, $\mathrm{supp} \, W_{\mathcal{A}}(f \otimes \bar{g}) = \mathrm{supp} (f \otimes \bar{g})$, and again every pair (f, g) of compactly supported functions in $L^2(\mathbb{R}^d)$ yields a counter-example to the analogous version of Benedicks’s theorem.

By contrast, we can represent the short-time Fourier transform, the ambiguity function as well as the Wigner distribution in the form $W_{\mathcal{A}}(f, g)$ for some symplectic matrix $\mathcal{A} \in \mathrm{Sp}(4d, \mathbb{R})$, and in this case, if $\mathrm{supp} \, W_{\mathcal{A}}(f, g)$ has finite measure, then $f \equiv 0$ or $g \equiv 0$.

These examples show that some symplectic time-frequency representations obey an uncertainty principle in the style of Benedicks, and others violate this uncertainty principle. So what is going on? The initial examples look rather puzzling and have struck our curiosity.

We investigate the Benedicks-type uncertainty principle for general symplectic time-frequency representations and will provide a complete description of those symplectic time-frequency representations that yield a Benedicks-type uncertainty principle. As this description requires some structure theory of the symplectic group and substantial set-up, we formulate the most important special case of Benedick-type uncertainty principle in order to offer a flavour of our general result.

Theorem 1.1. *Let $\mathcal{A} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \mathrm{Sp}(4d, \mathbb{R})$, and assume that B is invertible. Consider the symmetric matrix $P = B^{-1}A$.*

(i) *Assume P is not block-diagonal. Then $W_{\mathcal{A}}(f, g)$ satisfies the Benedicks-type uncertainty principle:*

*If $W_{\mathcal{A}}(f, g)$ is supported on a set of finite measure,
then $f \equiv 0$ or $g \equiv 0$.*

(ii) *Assume P is block-diagonal. Then there exist $f, g \in L^2(\mathbb{R}^d)$, $f \not\equiv 0, g \not\equiv 0$, such that the support of $W_{\mathcal{A}}(f, g)$ is compact.*

The matrices addressed in Theorem 1.1 are called *free symplectic matrices*. In the full treatment of a Benedicks-type uncertainty principle we will remove the condition that $\det B \neq 0$ with a trick that might be useful in other contexts as well. By multiplying $\mathcal{A}$ with an orthogonal symplectic matrix of the form

$$\mathcal{R}_{\tau I} = \begin{pmatrix} aI & bI \\ -bI & aI \end{pmatrix} \in \mathrm{Sp}(4d, \mathbb{R}) \cap \mathrm{O}(4d, \mathbb{R}),$$

we can guarantee that the assumption of Theorem 1.1 is satisfied for the resulting matrix $\mathcal{A}\mathcal{R}_{\tau}$ for all but finitely many $\tau = a + ib \in \mathbb{T}$.

The proof of Theorem 1.1 is based on the following key elements.

(i) We use a version of the Iwasawa decomposition of symplectic matrices that yields a factorization of a symplectic matrix into three matrices of the type that were discussed in the examples (i) — (iii). In view of the (counter-)examples above, the main problem is the maximal compact subgroup of $\mathrm{Sp}(4d, \mathbb{R})$, see Section 4. A

main role in the investigation will be played by a recent singular value decomposition of orthogonal symplectic matrices.

(ii) We introduce a new notion of a partial short-time Fourier transform by freezing some variables and taking the short-time Fourier transform only over a subset of variables.

Our final result will be an algorithm for deciding which symplectic matrices obey the Benedicks-type uncertainty principle. Although we obtain a complete description of all symplectic time-frequency representations for which a Benedicks-type uncertainty principle holds, the deeper geometric meaning of the conditions still remains unclear to us.

In the last part, we will investigate quadratic time-frequency representations of the form $W_{\mathcal{A}}(f, f) = \mu(\mathcal{A})(f \otimes \bar{f})$. It may seem that nothing new is gained by studying quadratic time-frequency representations. However, the example of a standard time-frequency representation points again to some mystery. Let

$$R(f, g)(x, \omega) = e^{-2\pi i x^t \omega} f(x) \overline{\hat{g}(\omega)}$$

be the so-called Rihaczek distribution. Again $R(f, g)$ can be written as a symplectic time-frequency representation. In view of the definition, all classical uncertainty principles for the Fourier pair $(f, \hat{f})$ are statements about the (quadratic) Rihaczek distribution $R(f, f)$. However, if we look at the sesquilinear Rihaczek distribution $R(f, g)$, then clearly there are $f, g \in L^2(\mathbb{R}^d)$, such that $R(f, g)$ has compact support, and there is no Benedicks-type uncertainty principle. By contrast, the uncertainty principle for the quadratic expression $R(f, f)$ is the classical uncertainty principle of Benedicks. Thus there is a subtle difference between quadratic and bilinear time-frequency representations. In Theorem 6.3, we will characterize all matrices $\mathcal{A}$ for which the quadratic symplectic time-frequency representation $W_{\mathcal{A}}(f, f)$ satisfies a Benedicks-type uncertainty principle.

The paper is organized as follows: we begin by introducing the symplectic group and the metaplectic group in Section 2. In Section 3, we introduce the partial short-time Fourier transform and use its properties to prove Theorem 1.1, which is a direct consequence of Theorem 3.3. Before we generalize this for arbitrary symplectic matrices, we address the structure of the symplectic group in Section 4. In Section 5, we deduce the main result (Theorem 5.5) from Theorem 3.3 through a series of reductions, and we formulate an algorithm for deciding which symplectic matrices obey the Benedicks-type uncertainty principle which avoids the technically difficult queries of the reduction process. We dedicate Section 6 to uncertainty principles for quadratic time-frequency representations. We illustrate the results by testing them on established results for the short-time Fourier transform, the ambiguity function and the (τ) -Wigner distribution in Section 7.

2. THE SYMPLECTIC GROUP AND THE METAPLECTIC GROUP

2.1. The symplectic group. The symplectic group $\mathrm{Sp}(2d, \mathbb{R})$ consists of the matrices $\mathcal{A} \in \mathbb{R}^{2d \times 2d}$ which preserve the standard skew-symmetric bilinear form, i.e.,

$$\mathrm{Sp}(2d, \mathbb{R}) = \left\{ \mathcal{A} \in \mathbb{R}^{2d \times 2d} : \mathcal{A}^t \mathcal{J} \mathcal{A} = \mathcal{J} \right\},$$

where $\mathcal{J}$ denotes the standard skew-symmetric matrix defined in (1.1). The symplectic group is a subgroup of the special linear group, and these two groups coincide only when $d = 1$. The simplest symplectic matrices are of the following type:²

$$(2.1) \quad \mathcal{V}_Q = \begin{pmatrix} I & 0 \\ Q & I \end{pmatrix}, \quad \mathcal{D}_L = \begin{pmatrix} L & 0 \\ 0 & L^{-T} \end{pmatrix}, \quad \mathcal{R}_{A+iB} = \begin{pmatrix} A & B \\ -B & A \end{pmatrix}.$$

The block structure of $\mathcal{J}$ provides four algebraic equations that characterize the blocks of symplectic matrices. With these equations, one can easily verify that $\mathcal{D}_L$ are the only block-diagonal symplectic matrices, $\mathcal{V}_Q$ is symplectic if and only if Q is symmetric, and $\mathcal{R}_{A+iB}$ is symplectic if and only if $U = A + iB$ is unitary on $\mathbb{C}^d$.

The symplectic group is a classical Lie group whose double cover is called the *metaplectic group*, denoted with $\text{Mp}(2d, \mathbb{R})$ [22, 28].

2.2. Metaplectic operators. The symmetric time-frequency shifts

$$\rho(x, \omega) = T_{x/2} M_\omega T_{x/2}$$

are a projective unitary representation of $\mathbb{R}^{2d}$, as they satisfy

$$(2.2) \quad \rho(\lambda + \gamma) = e^{\pi i \lambda \cdot \mathcal{J} \gamma} \rho(\lambda) \rho(\gamma), \quad \lambda, \gamma \in \mathbb{R}^{2d}.$$

This extends to a unitary representation of the Heisenberg group [12, 15, 21]. From this perspective, the symplectic group is precisely the subgroup of $\text{GL}(2d, \mathbb{R})$ which preserve the phase factor $e^{\pi i \lambda \cdot \mathcal{J} \gamma}$ in (2.2).

By results in representation theory on the Heisenberg group, specifically, the Stone-von-Neumann theorem [12, 15, 21], there exist unitary operators $\mu(\mathcal{A})$ with

$$(2.3) \quad \rho(\mathcal{A}\lambda) = \mu(\mathcal{A}) \rho(\lambda) \mu(\mathcal{A})^{-1},$$

i.e., $\mu(\mathcal{A})$ arises as an intertwining operator. These operators are called *metaplectic operators*. The operator is unique up to a global phase: it is obvious that any $\tau\mu(\mathcal{A})$ ($\tau \in \mathbb{T}$) also satisfies (2.3).

We already encountered the standard examples: the unitary dilations, the linear chirps, and the Fourier transform;

$$(2.4) \quad \begin{aligned} \mu(\mathcal{D}_L) f(x) &= |\det L|^{-1/2} f(L^{-1}x), \\ \mu(\mathcal{V}_Q) f(x) &= e^{\pi i x^t Q x} f(x), \\ \mu(\mathcal{J}) &= \mathcal{F}. \end{aligned}$$

In this work, we will work with this particular choice of $\mu(\mathcal{V}_Q)$, $\mu(\mathcal{D}_L)$ and $\mu(\mathcal{J})$, but the results are independent of the particular choice of the global phase. The metaplectic operators are a non-trivial projective representation of the symplectic group called the *oscillator representation* [36]. In particular, by selecting representatives $\mu(\mathcal{A}_0)$ for a set of generators $\mathcal{A}_0 \in \text{Sp}(2d, \mathbb{R})$, we can decompose any intertwining operator $\mu(\mathcal{A})$, $\mathcal{A} \in \text{Sp}(2d, \mathbb{R})$ up to a global phase.

The name *metaplectic* stems from the Weil representation [36], which states that the double cover of $\text{Sp}(2d, \mathbb{R})$ is isomorphic to a group of metaplectic operators,

²We use the convention $A + iB \in \mathbb{C}^{d \times d}$ with $A = \text{Re}(A + iB)$, $B = \text{Im}(A + iB)$.

i.e., we can identify the abstract cover $\text{Mp}(2d, \mathbb{R})$ with a group of operators. With this designation, the projection

$$\pi^{\text{Mp}} : \text{Mp}(2d, \mathbb{R}) \rightarrow \text{Sp}(2d, \mathbb{R})$$

is a surjective group homomorphism, with kernel $\ker(\pi^{\text{Mp}}) = \{\text{id}, -\text{id}\}$. The fact that many well-known unitary operators are, in fact, metaplectic operators, allows for a unified understanding of their interactions and their action on $L^2(\mathbb{R}^d)$.

2.3. Free symplectic matrices and quadratic Fourier transforms. Besides the examples above, an important subset of $\text{Sp}(2d, \mathbb{R})$ that stands out is the set of *free symplectic matrices*, i.e., the symplectic matrices with an invertible upper-right $d \times d$ block. The following factorization result provides a simplified case to examine for the uncertainty principles.

Proposition 2.1. *Let $\mathcal{A} \in \text{Sp}(2d, \mathbb{R})$ be a free symplectic matrix, i.e., have a block structure*

$$\mathcal{A} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

with $A, B, C, D \in \mathbb{R}^{d \times d}$, and $B \in \text{GL}(d, \mathbb{R})$. Then

$$\mathcal{A} = \mathcal{V}_{DB^{-1}} \mathcal{D}_B \mathcal{J} \mathcal{V}_{B^{-1}A}.$$

The proof can be found in [11, Prop. 2.39].

With the exception of the lower-triangular symplectic matrices, all metaplectic operators have an integral representation [34]. This holds in particular for the free symplectic matrices. The corresponding metaplectic operators of free symplectic matrices are called *quadratic Fourier transforms* [12].

Remark 2.2. We distinguish matrices in $\text{Sp}(2d, \mathbb{R})$ and $\text{Sp}(4d, \mathbb{R})$ by the following notation:

$$\begin{aligned} \mathcal{A}, \quad \mathcal{V}_Q, \mathcal{D}_L, \mathcal{R}_U &\in \text{Sp}(2d, \mathbb{R}), \quad Q, L \in \mathbb{R}^{d \times d}, U \in \mathbb{C}^{d \times d}, \\ \mathbf{A}, \quad \mathbf{V}_Q, \mathbf{D}_L, \mathbf{R}_U &\in \text{Sp}(4d, \mathbb{R}), \quad Q, L \in \mathbb{R}^{2d \times 2d}, U \in \mathbb{C}^{2d \times 2d}. \end{aligned}$$

The blocks of a symplectic matrix are in general denoted with A, B, C, D , but we rely on the following convention whenever possible: L is used for invertible matrices, W, V denote orthogonal matrices, P, Q denote symmetric matrices, U is a unitary matrix and Σ stands for a unitary diagonal matrix. Note that only U and Σ are *complex-valued*.

Remark 2.3. Let $A, B \in \mathbb{C}^{d \times d}$. We denote with $\text{diag}(A, B) = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ block-diagonal matrices with $d \times d$ blocks. One can easily verify that $\text{Sp}(2d, \mathbb{R}) \times \text{Sp}(2d, \mathbb{R})$ can be embedded in $\text{Sp}(4d, \mathbb{R})$ by embedding the generators, cf. Section 4, in the following way:

$$\begin{aligned} (\mathcal{D}_{L_1}, \mathcal{D}_{L_2}) &\mapsto \mathbf{D}_{\text{diag}(L_1, L_2)}, \quad L_1, L_2 \in \text{GL}(d, \mathbb{R}), \\ (\mathcal{V}_{Q_1}, \mathcal{V}_{Q_2}) &\mapsto \mathbf{V}_{\text{diag}(Q_1, Q_2)}, \quad Q_1 = Q_1^t, Q_2 = Q_2^t \in \mathbb{R}^{d \times d}, \\ (\mathcal{R}_{U_1}, \mathcal{R}_{U_2}) &\mapsto \mathbf{R}_{\text{diag}(U_1, U_2)}, \quad U_1, U_2 \in \text{U}(d, \mathbb{C}). \end{aligned}$$

On the operator level, the action on the elementary tensors is given by

$$\begin{aligned}\mu(\mathcal{D}_{\text{diag}(L_1, L_2)})(f \otimes g) &= \mu(\mathcal{D}_{L_1})f \otimes \mu(\mathcal{D}_{L_2})g, \\ \mu(\mathcal{V}_{\text{diag}(Q_1, Q_2)})(f \otimes g) &= \mu(\mathcal{V}_{Q_1})f \otimes \mu(\mathcal{V}_{Q_2})g, \\ \mu(\mathcal{R}_{\text{diag}(U_1, U_2)})(f \otimes g) &= \mu(\mathcal{R}_{U_1})f \otimes \mu(\mathcal{R}_{U_2})g.\end{aligned}$$

We will also rely heavily on the following relations between the matrices $\mathcal{V}_Q$, $\mathcal{D}_L$, $\mathcal{R}_U$, $Q = Q^t \in \mathbb{R}^{d \times d}$, $L \in \text{GL}(d, \mathbb{R})$, $U = A + iB \in \text{U}(d, \mathbb{C})$.

Remark 2.4. Shearing matrices $\mathcal{V}_Q$, dilation matrices $\mathcal{D}_L$ and orthogonal matrices $\mathcal{R}_U$ satisfy the following relations.

- (i) It holds $\mathcal{D}_L = \mathcal{R}_L$ if and only if L is orthogonal. Vice versa, $\mathcal{R}_U = \mathcal{D}_U$ if and only if U is real-valued, i.e., orthogonal.
- (ii) For all $L \in \text{GL}(d, \mathbb{R})$ holds $\mathcal{J}\mathcal{D}_L = \mathcal{D}_{L^{-t}}\mathcal{J}$ (recall the interaction of the Fourier transform with dilations).
- (iii) For all $L_1, L_2 \in \text{GL}(d, \mathbb{R})$ and $Q = Q^t \in \mathbb{R}^{d \times d}$ holds

$$\mathcal{D}_{L_1}\mathcal{V}_Q\mathcal{D}_{L_2} = \mathcal{D}_{L_1L_2}\mathcal{V}_{L_2^tQL_2} = \mathcal{V}_{L_1^{-t}QL_1^{-1}}\mathcal{D}_{L_1L_2}.$$

By setting either of L_1, L_2, Q to I , one can derive simpler relations.

- (iv) For all $Q_1 = Q_1^t, Q_2 = Q_2^t \in \mathbb{R}^{d \times d}$, $L_1, L_2 \in \text{GL}(d, \mathbb{R})$, and $U_1, U_2 \in \text{U}(d, \mathbb{C})$ holds

$$\mathcal{V}_{Q_1}\mathcal{V}_{Q_2} = \mathcal{V}_{Q_1+Q_2}, \quad \mathcal{D}_{L_1}\mathcal{D}_{L_2} = \mathcal{D}_{L_1L_2}, \quad \mathcal{R}_{U_1}\mathcal{R}_{U_2} = \mathcal{R}_{U_1U_2}.$$

3. UNCERTAINTY PRINCIPLES FOR FREE SYMPLECTIC MATRICES

In this section, we will prove a Benedicks-type uncertainty principle for the $\mathcal{A}$ -distribution of free symplectic matrices $\mathcal{A} \in \text{Sp}(4d, \mathbb{R})$. We reduce an uncertainty principle for arbitrary $\mathcal{A}$ -Wigner distributions to this special case.

Given a function $f \in L^2(\mathbb{R}^d)$ and writing $x = (x_1, x_2)$ for $x_1 \in \mathbb{R}^k$ and $x_2 \in \mathbb{R}^{d-k}$ for some $1 \leq k \leq d$, we denote with f_{x_2} the restriction³

$$f_{x_2} : \mathbb{R}^k \rightarrow \mathbb{C}, \quad x_1 \mapsto f(x_1, x_2).$$

By Fubini's theorem, for almost all $x_2 \in \mathbb{R}^{d-k}$, the function f_{x_2} is square-integrable. This allows us to define the partial short-time Fourier transform in k variables almost everywhere.

Definition 3.1. Let $f, g \in L^2(\mathbb{R}^d)$, $g \neq 0$, and let $k \in \{1, \dots, d\}$. We define the partial short-time Fourier transform in the first k variables of the function f with respect to g as the almost everywhere defined function

$$\begin{aligned}(3.1) \quad V_g^k f(x_1, x_2, \omega_1, \omega_2) &= \int_{\mathbb{R}^k} f(t, x_2) \overline{g(t - x_1, -\omega_2)} e^{-2\pi i t \cdot \omega_1} dt \\ &= \langle f_{x_2}, M_{\omega_2} T_{x_1} g_{-\omega_2} \rangle \\ &= V_{g_{-\omega_2}} f_{x_2}(x_1, \omega_1),\end{aligned}$$

for $(x_1, x_2, \omega_1, \omega_2) \in \mathbb{R}^k \times \mathbb{R}^{d-k} \times \mathbb{R}^k \times \mathbb{R}^{d-k}$.

³If $k = d$, then we interpret f_{x_2} as f .

This is a special case of a metaplectic time-frequency representation, i.e., there exists a matrix $\mathcal{A}_k \in \text{Sp}(4d, \mathbb{R})$ such that $V_g^k f = W_{\mathcal{A}_k}(f, g)$.

If $k = d$, $V_g^d = V_g$ is the classical short-time Fourier transform. Otherwise, the integral in (3.1) is well-defined almost everywhere because f_{x_2} and g_{ω_2} are square-integrable for almost all $x_2, \omega_2 \in \mathbb{R}^{d-k}$. Furthermore, it is continuous in $x_1, \omega_1 \in \mathbb{R}^k$.

Recall the notation introduced in (2.1): $\mathcal{D}_L = \begin{pmatrix} L & 0 \\ 0 & L^{-t} \end{pmatrix}$ and $\mathcal{R}_{A+iB} = \begin{pmatrix} A & B \\ -B & A \end{pmatrix}$, where $L \in \text{GL}(d, \mathbb{R})$ and $U = A + iB \in \text{U}(d, \mathbb{C})$. Consider the partial Fourier transform which projects to the matrix

$$\mathcal{J}_k = \mathcal{R}_{\text{diag}(I_k, I_{d-k}, iI_k, I_{d-k})} \in \text{Sp}(4d, \mathbb{R}).$$

By choosing an appropriate dilation matrix $L_k \in \text{GL}(2d, \mathbb{R})$, we see that the partial short-time Fourier transform $V_g^k f$ is the metaplectic time-frequency representation $W_{\mathcal{A}_k}(f, g)$ with $\mathcal{A}_k = \mathcal{J}_k \mathcal{D}_{L_k}$.

Lemma 3.2. *Let $f, g \in L^2(\mathbb{R}^d)$ and let $k \in \{1, \dots, d\}$. If $V_g^k f$ is supported on a set of finite measure, then $f \equiv 0$ or $g \equiv 0$.*

Proof. If $k = d$, then this is the Benedicks-type uncertainty principle for the short-time Fourier transform [20, 25, 27, 38]. Therefore, we make the additional assumption that $1 \leq k < d$.

By Fubini's theorem, there exists a set $N_1 \subseteq \mathbb{R}^{2d}$ of measure zero such that for all $(x_1, x_2, \omega_1, \omega_2) \in \mathbb{R}^{2d} \setminus N_1$, the functions f_{x_2}, g_{ω_2} are square-integrable and the representation (3.1) holds pointwise on $\mathbb{R}^{2d} \setminus N_1$.

Let $\mathcal{S} \subseteq \mathbb{R}^{2d} \setminus N_1$ be the support of $V_g^k f$, where we have already omitted the null set of restrictions $f_{x_2}, g_{\omega_2}, x_2, \omega_2 \in \mathbb{R}^{d-k}$ that are meaningless or not in $L^2(\mathbb{R}^k)$. By Fubini's theorem, there exists a further set $N_2 \subseteq \mathbb{R}^{d-k} \times \mathbb{R}^{d-k}$ of measure zero such that for fixed $(x_2, \omega_2) \in (\mathbb{R}^{d-k} \times \mathbb{R}^{d-k}) \setminus N_2$ every section

$$\mathcal{S}_{x_2, \omega_2} := \{(x_1, \omega_1) \in \mathbb{R}^k \times \mathbb{R}^k : (x_1, x_2, \omega_1, \omega_2) \in \mathcal{S}\}$$

- (i) is measurable and has finite measure in $\mathbb{R}^k \times \mathbb{R}^k$,
- (ii) the measure of $\mathcal{S}$ is given by

$$(3.2) \quad |\mathcal{S}| = \int_{(\mathbb{R}^{d-k} \times \mathbb{R}^{d-k}) \setminus N_2} |\mathcal{S}_{x_2, \omega_2}| d(x_2, \omega_2),$$

- (iii) it holds $\mathcal{S}_{x_2, \omega_2} = \text{supp } V_{g-\omega_2} f_{x_2}$.

Since $\mathcal{S}_{x_2, \omega_2}$ has finite measure in $\mathbb{R}^k \times \mathbb{R}^k$, the a Benedicks-type uncertainty principle for the short-time Fourier transform [20, 25, 27, 38] implies that $f_{x_2} \equiv 0$ or $g_{\omega_2} \equiv 0$, and as a consequence, $|\mathcal{S}_{x_2, \omega_2}| = 0$.

Going back to (3.2), this implies that $|\mathcal{S}| = 0$, hence $V_g^k f \equiv 0$. As already noted in Definition 3.1, $V_g^k f = \mu(\mathcal{A}_k)(f \otimes \bar{g})$ for a suitable symplectic matrix $\mathcal{A}_k \in \text{Sp}(4d, \mathbb{R})$. Since $\mu(\mathcal{A}_k)$ is unitary, it follows that $f \otimes \bar{g} \equiv 0$, or equivalently, $f \equiv 0$ or $g \equiv 0$. This proves the claim. $\blacksquare$

We now prove an uncertainty principle for the $\mathcal{A}$ -Wigner distribution of free symplectic matrices $\mathcal{A} \in \text{Sp}(4d, \mathbb{R})$.

Theorem 3.3. *Let $\mathcal{A} = \mathcal{J}\mathcal{V}_P = \begin{pmatrix} P & I \\ -I & 0 \end{pmatrix} \in \text{Sp}(4d, \mathbb{R})$ be a free symplectic matrix with*

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \in \mathbb{R}^{2d \times 2d},$$

such that the $d \times d$ blocks on the side diagonal $P_{12} = P_{21}^t \neq 0$ are not zero. Then a Benedicks-type uncertainty principle holds:

If $W_{\mathcal{A}}(f, g)$ is supported on a set of finite measure for some $f, g \in L^2(\mathbb{R}^d)$, then $f \equiv 0$ or $g \equiv 0$.

Proof. Let $f, g \in L^2(\mathbb{R}^d)$ be two functions with $W_{\mathcal{A}}(f, g)$ supported on a set of finite measure. We want to show that $f \equiv 0$ or $g \equiv 0$. We begin with a reduction to metaplectic operators with a simple action on elementary tensors.

Since $P_{12} = P_{21}^t \neq 0$, it has a singular value decomposition

$$P_{12} = W_1 \cdot \Gamma \cdot W_2^t,$$

where $W_1, W_2 \in \text{O}(d, \mathbb{R})$ and Γ is diagonal, with singular values sorted in descending order on the main diagonal, i.e.,

$$\Gamma = \begin{pmatrix} \Gamma_1 & 0_{k \times k} \\ 0_{d-k \times k} & 0_{d-k \times d-k} \end{pmatrix}, \quad \Gamma_1 \in \text{GL}(k, \mathbb{R}), \quad k = \text{rank}(P_{12}) \geq 1.$$

Due to the relations described in Remark 2.4, we can write $\mathcal{A}$ as

$$\begin{aligned} \mathcal{A} &= \mathcal{J}\mathcal{V}_P = \mathcal{J}\mathcal{V} \begin{pmatrix} 0 & P_{12} \\ P_{12}^t & 0 \end{pmatrix} + \text{diag}(P_{11}, P_{22}) = \mathcal{J}\mathcal{V} \begin{pmatrix} 0 & W_1 \Gamma W_2^t \\ W_2 \Gamma W_1^t & 0 \end{pmatrix} \mathcal{V}^{\text{diag}(P_{11}, P_{22})} \\ &= \mathcal{J}\mathcal{D}^{\text{diag}(W_1, W_2)} \mathcal{V} \begin{pmatrix} 0 & \Gamma \\ \Gamma & 0 \end{pmatrix} \mathcal{D}^{\text{diag}(W_1, W_2)^t} \mathcal{V}^{\text{diag}(P_{11}, P_{22})} \\ &= \mathcal{D}^{\text{diag}(W_1, W_2)} \mathcal{J}\mathcal{V} \begin{pmatrix} 0 & \Gamma \\ \Gamma & 0 \end{pmatrix} \mathcal{D}^{\text{diag}(W_1^t, W_2^t)} \mathcal{V}^{\text{diag}(P_{11}, P_{22})}. \end{aligned}$$

We set $f_1 := \mu(\mathcal{D}_{W_1^t} \mathcal{V}_{P_{11}})f$, $g_1 := \mu(\mathcal{D}_{W_2^t} \mathcal{V}_{-P_{22}})g \in L^2(\mathbb{R}^d)^4$. By Remark 2.3, we can rewrite $W_{\mathcal{A}}(f, g)$ as

$$W_{\mathcal{A}}(f, g) = \mu(\mathcal{D}^{\text{diag}(W_1, W_2)}) \mathcal{F}\mu \left(\mathcal{V} \begin{pmatrix} 0 & \Gamma \\ \Gamma & 0 \end{pmatrix} \right) (f_1 \otimes \overline{g_1}).$$

As discussed in the introduction, the dilation operator $\mu(\mathcal{D}^{\text{diag}(W_1, W_2)})$ only rescales the support of $\mathcal{F}\mu(\mathcal{V} \begin{pmatrix} 0 & \Gamma \\ \Gamma & 0 \end{pmatrix})(f_1 \otimes \overline{g_1})$. Thus, $\mathcal{F}\mu(\mathcal{V} \begin{pmatrix} 0 & \Gamma \\ \Gamma & 0 \end{pmatrix})(f_1 \otimes \overline{g_1})$ is supported on a set of finite measure.

By (2.4),

$$\mathcal{F}\mu(\mathcal{V} \begin{pmatrix} 0 & \Gamma \\ \Gamma & 0 \end{pmatrix})(f_1 \otimes \overline{g_1})(x, y) = \mu(\mathcal{V} \begin{pmatrix} 0 & \Gamma \\ \Gamma & 0 \end{pmatrix})(f_1 \otimes \overline{g_1})(x, y) = e^{2\pi i x \cdot \Gamma y} f_1(x) \overline{g_1}(y).$$

⁴Note that $\mu(\mathcal{D}_{W_1^t} \mathcal{V}_{P_{11}}), \mu(\mathcal{D}_{W_2^t} \mathcal{V}_{-P_{22}})$ belong to $\text{Mp}(2d, \mathbb{R})$, not to $\text{Mp}(4d, \mathbb{R})$.

By Plancherel's theorem and by Fubini's theorem, for almost all $(\omega, \xi) \in \mathbb{R}^{2d}$, the integrals to follow are well-defined;⁵

$$\begin{aligned}
 (3.3) \quad & \mathcal{F}\mu(\mathcal{V}_{\begin{pmatrix} 0 & \Gamma \\ \Gamma & 0 \end{pmatrix}})(f_1 \otimes \overline{g_1})(\omega, \xi) \\
 &= \int_{\mathbb{R}^d} f_1(x) e^{-2\pi i \omega \cdot x} \int_{\mathbb{R}^d} \overline{g_1}(y) e^{-2\pi i \xi \cdot y} \cdot e^{2\pi i x \cdot \Gamma y} dy dx \\
 &= \int_{\mathbb{R}^d} f_1(x) e^{-2\pi i \omega \cdot x} (\widehat{g_1})(\Gamma x - \xi) dx.
 \end{aligned}$$

If Γ were invertible, then we could identify the above expression, after a coordinate change, as a short-time Fourier transform with a support of finite measure, and we could finalize the proof. However, in general, the off-diagonal blocks P_{12} , $P_{21} = P_{12}^t$, and thus Γ , are *not* invertible. This is the reason why we have to use the partial short-time Fourier transform in (3.1). To process the integral in (3.3) further, we evaluate $W_{\mathcal{A}}(f, g)$ at the points

$$(\text{diag}(\Gamma_1, I)\omega, \xi) = (\Gamma_1\omega_1, \omega_2, \xi)$$

instead of at (ω, ξ) . In setting this coordinate change, the final expression in (3.3) becomes⁶

$$\begin{aligned}
 (3.4) \quad & \mathcal{F}\mu(\mathcal{V}_{\begin{pmatrix} 0 & \Gamma \\ \Gamma & 0 \end{pmatrix}})(f_1 \otimes \overline{g_1})(\text{diag}(\Gamma_1, I)\omega, \xi) \\
 &= \int_{\mathbb{R}^d} f_1(x) e^{-2\pi i \omega \cdot \text{diag}(\Gamma_1, I)x} \widehat{g_1}(\Gamma x - \xi) dx \\
 &= \int_{\mathbb{R}^k} \widehat{g_1}(\Gamma_1 x_1 - \xi_1, -\xi_2) e^{-2\pi i \omega_1 \cdot \Gamma_1 x_1} \int_{\mathbb{R}^{d-k}} f_1(x_1, x_2) e^{-2\pi i \omega_2 \cdot x_2} dx_2 dx_1.
 \end{aligned}$$

We set

$$\begin{aligned}
 f_2(\Gamma_1 x_1, \omega_2) &:= \int_{\mathbb{R}^{d-k}} f_1(x_1, x_2) e^{-2\pi i \omega_2 \cdot x_2} dx_2, \\
 g_2(\omega) &:= \mathcal{F}g_1(\omega) = \int_{\mathbb{R}^d} g_1(x) e^{-2\pi i x \cdot \omega} dx.
 \end{aligned}$$

Both f_2 and g_2 are (rescaled) partial Fourier transforms of f_1 and g_1 , therefore $f_2, g_2 \in L^2(\mathbb{R}^d)$ (and we note that the integrals make sense only for $f_1, g_1 \in L^1(\mathbb{R}^d)$ and almost everywhere in x_1 , otherwise, we would use Plancherel's theorem). With this definition and the substitution $s = \Gamma_1 \eta_1$, (3.4) can be turned into the final form

$$\begin{aligned}
 (3.5) \quad & \mathcal{F}\mu(\mathcal{V}_{\begin{pmatrix} 0 & \Gamma \\ \Gamma & 0 \end{pmatrix}})(f_1 \otimes \overline{g_1})(\text{diag}(\Gamma_1, I)\omega, \xi) \\
 &= \int_{\mathbb{R}^k} f_2(\Gamma_1 x_1, \omega_2) \overline{g_2}(\Gamma_1 x_1 - \xi_1, -\xi_2) e^{-2\pi i \omega_1 \cdot \Gamma_1 x_1} dx_1 \\
 &\stackrel{s=\Gamma_1 x_1}{=} |\det \Gamma_1|^{-1} \int_{\mathbb{R}^k} f_2(t, \omega_2) \overline{g_2}(t - \xi_1, -\xi_2) e^{-2\pi i \omega_1 \cdot t} ds \\
 &= |\det \Gamma_1|^{-1} V_{g_2} f_2(\xi_1, \omega_2, \omega_1, \xi_2).
 \end{aligned}$$

Since $W_{\mathcal{A}}(f, g)$ has a support of finite measure, the transformed expression (3.5) also has a support of finite measure. By Lemma 3.2, this implies that $f_2 \equiv 0$ or

⁵For ease of reading, we use $\mathcal{F}f = \hat{f}$ interchangeably.

⁶If $k = d$, we set the inner integral over $\mathbb{R}^{d-k}$ to be the point evaluation of the integrand.

$g_2 \equiv 0$. Since f_2 and g_2 are the image of f and g under unitary operators, it follows that $f \equiv 0$ or $g \equiv 0$. This concludes the proof. $\blacksquare$

Remark 3.4. Theorem 3.3 directly Theorem 1.1. Let $\mathcal{A} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \text{Sp}(4d, \mathbb{R})$ be a free symplectic matrix. We set $Q = DB^{-1}$ and $P = B^{-1}A$. By Proposition 2.1,

$$\mathcal{A} = \mathcal{V}_Q \mathcal{D}_B \mathcal{J} \mathcal{V}_P.$$

(i) Assume P is not block-diagonal. By (2.4), for all $\lambda \in \mathbb{R}^{2d}$ holds

$$|W_{\mathcal{A}}(f, g)(\lambda)| = |\mu(\mathcal{D}_B) W_{\mathcal{J} \mathcal{V}_P}(f, g)(\lambda)| = |\det B|^{-1/2} |W_{\mathcal{J} \mathcal{V}_P}(f, g)(B^{-1}\lambda)|.$$

Thus, $W_{\mathcal{A}}(f, g)$ is supported on a set of finite measure for some $f, g \in L^2(\mathbb{R}^d)$ if and only if $W_{\mathcal{J} \mathcal{V}_P}(f, g)$ is supported on a set of finite measure. By Theorem 3.3, this implies that $f \equiv 0$ or $g \equiv 0$, as claimed.

(ii) If $P = \begin{pmatrix} P_{11} & 0 \\ 0 & P_{22} \end{pmatrix}$ is block-diagonal, then by Remark 2.3

$$\begin{aligned} |W_{\mathcal{A}}(f, g)(\lambda)| &= |\mu(\mathcal{D}_B) \mathcal{F}((\mu(\mathcal{V}_{P_{11}})f \otimes \overline{\mu(\mathcal{V}_{-P_{22}})g})(\lambda))| \\ &= |\det B|^{-1/2} \left| (\mu(\mathcal{J} \mathcal{V}_{P_{11}})f \otimes \overline{\mu(\mathcal{J}^t \mathcal{V}_{-P_{22}})g})(B^{-1}\lambda) \right|. \end{aligned}$$

For all compactly supported $f_0, g_0 \in L^2(\mathbb{R}^d)$, $f_0 \not\equiv 0$, $g_0 \not\equiv 0$, the functions

$$f = (\mu(\mathcal{J} \mathcal{V}_{P_{11}}))^{-1} f_0, \quad g = \mu(\mathcal{J}^t \mathcal{V}_{-P_{22}})^{-1} g_0$$

satisfy

$$|W_{\mathcal{A}}(f, g)(\lambda)| = |\det B|^{-1/2} |f_0 \otimes \overline{g_0}(B^{-1}\lambda)|.$$

In particular, $W_{\mathcal{A}}(f, g)$ is supported on a compact set. This concludes the proof.

4. THE STRUCTURE OF THE SYMPLECTIC GROUP

Before treating uncertainty principles for arbitrary symplectic matrices, we need a deeper grasp of the structure of the symplectic group and its factorizations.

The symplectic group admits an Iwasawa decomposition KAN [24, 28], where K is a maximal compact Lie subgroup, A is a suitable abelian Lie subgroup and N is a nilpotent Lie subgroup. Due to the group structure of the decomposition, $\text{Sp}(2d, \mathbb{R})$ possesses also the Iwasawa decomposition $\text{Sp}(2d, \mathbb{R}) = NAK$, which is better suited for our needs. The maximal compact subgroup can be chosen to be $K = \text{Sp}(2d, \mathbb{R}) \cap \text{O}(2d, \mathbb{R})$ and is isomorphic to the unitary group $\text{U}(d, \mathbb{C})$ via the Lie group isomorphism [12, Sec. 2.3], [15, Sec. 2.1]

$$\text{U}(d, \mathbb{C}) \rightarrow \text{Sp}(2d, \mathbb{R}) \cap \text{O}(2d, \mathbb{R}), \quad A + iB \mapsto \mathcal{R}_{A+iB} := \begin{pmatrix} A & B \\ -B & A \end{pmatrix}.$$

The abelian Lie subgroup consists the all positive definite diagonal matrices

$$D = \text{diag}(a_1, \dots, a_d, a_1^{-1}, \dots, a_d^{-1}) = \mathcal{D}_{\text{diag}(a_1, \dots, a_d)} \in \text{Sp}(2d, \mathbb{R}), \quad a_1, \dots, a_d > 0.$$

By simple arguments [11, Ex. 2.28.], the nilpotent part can be written in the form of lower-triangular matrices

$$N = \begin{pmatrix} L_0 & 0 \\ QL_0 & L_0^{-t} \end{pmatrix} = \begin{pmatrix} I & 0 \\ Q & I \end{pmatrix} \cdot \begin{pmatrix} L_0 & 0 \\ 0 & L_0^{-t} \end{pmatrix} = \mathcal{V}_Q \mathcal{D}_{L_0},$$

where $L_0 \in \mathbb{R}^{d \times d}$ is a lower-triangular matrix with ones on the main diagonal and $Q = Q^t \in \mathbb{R}^{d \times d}$. By combining N and A , and recalling Remark 2.4, in particular, $\mathcal{D}_{L_0} \mathcal{D}_{\text{diag}(a_1, \dots, a_d)} = \mathcal{D}_L$, we can then write every symplectic matrix $\mathcal{A} \in \text{Sp}(2d, \mathbb{R})$ uniquely in the form

$$(4.1) \quad \mathcal{A} = \mathcal{V}_Q \mathcal{D}_L \mathcal{R}_U$$

with $L \in \text{GL}(d, \mathbb{R})$, $Q = Q^t \in \mathbb{R}^{d \times d}$, and $U \in \text{U}(d, \mathbb{C})$ such that $\mathcal{V}_Q \mathcal{D}_L \in NA$. Since the particular choice of Q , L , U is determined by the Iwasawa decomposition, they are unique in (4.1). We mention that there are other matrices Q' , L' , U' which satisfy $\mathcal{A} = \mathcal{V}_{Q'} \mathcal{D}_{L'} \mathcal{R}_{U'}$, and such a decomposition is referred to as a *pre-Iwasawa decomposition* of $\mathcal{A}$ [12]. The lack of ambiguity is mainly due to $\mathcal{R}_U = \mathcal{D}_U$ for all orthogonal matrices $U \in \text{O}(d, \mathbb{R})$ as noted in Remark 2.4. In the following, we will refer to the pre-Iwasawa decomposition (4.1) determined by the Iwasawa decomposition as *the pre-Iwasawa decomposition* of $\mathcal{A} \in \text{Sp}(2d, \mathbb{R})$. The main results are stated in terms of this pre-Iwasawa decomposition but hold for any pre-Iwasawa decomposition of $\mathcal{A} \in \text{Sp}(2d, \mathbb{R})$.

The advantage of (4.1) is the fact that the operators $\mu(\mathcal{V}_Q)$, $\mu(\mathcal{D}_L)$ do not change the nature of the support of a function, so that a compact support or a support of finite measure are preserved, see the examples in the introduction and (2.4). Thus the deciding factor for the validity of an uncertainty principle is played by the symplectic rotations $\mathcal{R}_U \in \text{Sp}(2d, \mathbb{R})$. Since we have already established an uncertainty principle for free symplectic matrices, we aim to reduce any case to that special one.

We make use of the following result [17, Thm. 5.1.], which is a version of a joint singular value decomposition over $\mathbb{R}$ of the real and imaginary part of $U \in \text{U}(d, \mathbb{C})$.

Lemma 4.1. *Let $d \in \mathbb{N}$. For every $U = A + iB \in \text{U}(d, \mathbb{C})$ there exist orthogonal matrices $W, V \in \text{O}(d, \mathbb{R})$ and a diagonal unitary matrix $\Sigma = \Sigma_r + i\Sigma_i \in \text{U}(d, \mathbb{C})$ such that*

$$U = W \Sigma V^t, \quad \text{i.e.,} \quad A = W \Sigma_r V^t, \quad B = W \Sigma_i V^t.$$

To deal with general symplectic matrices, we use the following trick.

Lemma 4.2. *Let $U = A + iB \in \text{U}(d, \mathbb{C})$ be given. Then there exists a $\tau \in \mathbb{T}$ such that the orthogonal symplectic matrix*

$$\mathcal{R}_{\tau U} = \mathcal{R}_{\tau I} \mathcal{R}_U = \mathcal{R}_U \mathcal{R}_{\tau I}$$

is free. Equivalently, $\text{Im}(\tau U)$ is invertible. Furthermore, for a fixed U , $\text{Im}(\tau U)$ fails to be invertible for at most $2d$ choices of $\tau \in \mathbb{T}$.

Proof. By Lemma 4.1, there exist $W, V \in \text{O}(d, \mathbb{R})$ and a diagonal unitary matrix $\Sigma = \Sigma_r + i\Sigma_i \in \text{U}(d, \mathbb{C})$ with

$$A = W \Sigma_r V^t, \quad B = W \Sigma_i V^t.$$

For all $\tau \in \mathbb{T}$,

$$\text{Im}(\tau U) = W \cdot \text{Im}(\tau \Sigma) \cdot V^t.$$

This shows that the diagonal matrix $\text{Im}(\tau\Sigma)$ has zeros on the diagonal if and only if $\tau = \overline{\Sigma_{kk}}$ for some $1 \leq k \leq d$. Thus $\text{Im}(\tau U)$ is singular if and only if $\tau = \pm \overline{\Sigma_{kk}}$ for some $1 \leq k \leq d$.

This proves both the existence of $\tau \in \mathbb{T}$ as desired as well as the upper bound on the number of exceptions.

Note that for $U = A + iB \in \text{U}(d, \mathbb{C})$, the corresponding orthogonal symplectic matrix $\mathcal{R}_U = \begin{pmatrix} A & B \\ -B & A \end{pmatrix}$ is free if and only if $B = \text{Im}(U)$ is invertible. Thus, $\mathcal{R}_{\tau U}$ is free if and only if $\text{Im}(\tau U)$ is invertible. $\blacksquare$

4.1. Fractional Fourier transforms. Before proceeding with the reductions, we take a closer look at the metaplectic operators which project to $\mathcal{R}_\Sigma$ for some *diagonal* unitary matrix $\Sigma \in \text{U}(d, \mathbb{C})$.

In the one-dimensional case, the unitary group is isomorphic to the torus $\mathbb{T}$, so any $\sigma = e^{i\alpha} \in \mathbb{T}$ is trivially diagonal. Hence, the maximal compact subgroup of $\text{Sp}(2, \mathbb{R})$ is abelian. The corresponding metaplectic operator is (up to global phase) the fractional Fourier transform $\mathcal{F}_\sigma$, and in the literature, it is often identified with the corresponding angle $\alpha \in [0, 2\pi)$. In particular, $\mathcal{F}_1 f = f$, $\mathcal{F}_{-1} f = f(-\cdot)$, and $\mathcal{F}_i f = \mathcal{F} f$ is the classical Fourier transform [29, 37]. For all $\sigma \in \mathbb{T} \setminus \{-1, 1\}$, the matrix $\mathcal{R}_\sigma = \mathcal{R}_{e^{i\alpha}} = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix}$ is free, thus the corresponding operator is a quadratic Fourier transform of the form

$$(4.2) \quad \mathcal{F}_\sigma f(\omega) = k_\sigma \int_{\mathbb{R}} f(t) e^{\pi i \cot(\alpha) \omega^2 - 2\pi i \csc(\alpha) \omega t + \pi i \cot(\alpha) t^2} dt,$$

where $k_\sigma \in \mathbb{C}$ is a normalizing constant.

In higher dimensions, the correspondence of diagonal unitary matrices to fractional Fourier transforms generalizes via partial fractional Fourier transforms. Given a diagonal matrix $\Sigma = \text{diag}(\Sigma_{jj})_{1 \leq j \leq d} \in \text{U}(d, \mathbb{C})$, let Σ_j be the diagonal matrix with Σ_{jj} on the j -th diagonal entry and otherwise 1. Then $\mu(\mathcal{R}_{\Sigma_j}) = \mathcal{F}_{\Sigma_j}$ acts as the partial fractional Fourier transform in the j -th coordinate, and

$$\Sigma = \prod_{j=1}^d \Sigma_j, \quad \mathcal{R}_\Sigma = \prod_{j=1}^d \mathcal{R}_{\Sigma_j}, \quad \mu(\mathcal{R}_\Sigma) = \mathcal{F}_\Sigma = \prod_{j=1}^d \mathcal{F}_{\Sigma_j}.$$

In particular, $\mathcal{F}_\Sigma$ acts on elementary tensors as

$$(4.3) \quad \mathcal{F}_\Sigma(f_1 \otimes \cdots \otimes f_d) = \mathcal{F}_{\Sigma_{11}} f_1 \otimes \cdots \otimes \mathcal{F}_{\Sigma_{dd}} f_d.$$

All $\mathcal{F}_\Sigma$ commute. Furthermore, from the integral representation (4.2) follows that for all diagonal unitary matrices $\Sigma \in \text{U}(d, \mathbb{C})$ exists a constant $c_\Sigma \in \mathbb{T}$ such that

$$(4.4) \quad \mathcal{F}_\Sigma \bar{f} = c_\Sigma \overline{\mathcal{F}_{\bar{\Sigma}} f}, \quad \Sigma \in \text{U}(d, \mathbb{C}) \text{ diagonal},$$

which will play a role in the existence of uncertainty principles.

We conclude the section with an uncertainty principle for pairs $(f, \mathcal{F}_\Sigma f)$ for diagonal $\Sigma \in \text{U}(d, \mathbb{C})$. The proof idea is motivated by the one-dimensional version [26], though it is not a trivial generalization.

Theorem 4.3. *Let $\Sigma \in \text{U}(d, \mathbb{C})$ be a diagonal unitary matrix, with a non-zero imaginary part. Then a Benedicks-type uncertainty principle holds:*

If both f and $\mathcal{F}_\Sigma f$ are supported on a set of finite measure, then f is the zero function.

Proof. Assume f and $\mathcal{F}_\Sigma f$ are supported on a set of finite measure. We want to show that f is the zero function.

Let $\Sigma = A + iB$. If B is invertible, then $\mathcal{R}_\Sigma$ is free, with

$$\mathcal{R}_\Sigma = \mathcal{V}_{AB^{-1}} \mathcal{D}_B \mathcal{F} \mathcal{V}_{B^{-1}A} = \mathcal{V}_{AB^{-1}} \mathcal{F} \mathcal{D}_{B^{-1}} \mathcal{V}_{B^{-1}A}.$$

In this case, the approach in [26] is applicable. By (2.4), $\mu(V_Q)$ and $\mu(D_L)$ have no substantial effect on the support, in particular, both $f_1 = \mu(\mathcal{D}_{B^{-1}} \mathcal{V}_{B^{-1}A})f$ and

$$\mu(\mathcal{V}_{AB^{-1}}) \mathcal{F}_\Sigma f = \mathcal{F} \mathcal{D}_{B^{-1}} \mathcal{V}_{B^{-1}A} f = \mathcal{F} f_1$$

are supported on a set of finite measure. By the classical Benedicks's uncertainty principle [1, 3], f_1 is the zero function, thus $f = \mu(\mathcal{D}_{B^{-1}} \mathcal{V}_{B^{-1}A})^{-1} f_1 \equiv 0$.

The general case follows by considering restrictions f_x to the range of B analogously to the proof of Theorem 3.3. ■

5. REDUCTIONS

In this section, we provide a criterion for the validity of an uncertainty principle for $W_{\mathcal{A}}(f, g)$ which can be easily checked from the blocks of the associated symplectic matrix. Technically, we will reduce the general case to the special case of free symplectic matrices.

Lemma 5.1. *Let $\mathcal{A} = \mathcal{V}_Q \mathcal{D}_L \mathcal{R}_U \in \text{Sp}(4d, \mathbb{R})$ be the pre-Iwasawa decomposition (4.1) of $\mathcal{A} \in \text{Sp}(4d, \mathbb{R})$. Then the following are equivalent.*

(i) *There exist $f, g \in L^2(\mathbb{R}^d)$, $f \not\equiv 0$, $g \not\equiv 0$, with*

$$|\text{supp } W_{\mathcal{A}}(f, g)| < \infty.$$

(ii) *There exist $f, g \in L^2(\mathbb{R}^d)$, $f \not\equiv 0$, $g \not\equiv 0$, with*

$$|\text{supp } W_{\mathcal{R}_U}(f, g)| < \infty.$$

Proof. By (2.4), for all $f, g \in L^2(\mathbb{R}^d)$, $\lambda \in \mathbb{R}^{2d}$, holds

$$\begin{aligned} W_{\mathcal{A}}(f, g)(\lambda) &= \mu(\mathcal{V}_Q) \mu(\mathcal{D}_L) \mu(\mathcal{R}_U)(f \otimes \bar{g})(\lambda) \\ &= |\det L|^{-1/2} e^{\pi i \lambda \cdot Q \lambda} \mu(\mathcal{R}_U)(f \otimes \bar{g})(L^{-1} \lambda). \end{aligned}$$

Therefore,

$$|W_{\mathcal{A}}(f, g)(\lambda)| = |\det L|^{-1/2} |(W_{\mathcal{R}_U}(f, g))(L^{-1} \lambda)|,$$

and thus,

$$\text{supp } W_{\mathcal{A}}(f, g) = L \cdot \text{supp } W_{\mathcal{R}_U}(f, g).$$

Now it is clear that $\text{supp } W_{\mathcal{A}}(f, g)$ has finite measure if and only if $\text{supp } W_{\mathcal{R}_U}(f, g)$ has finite measure. ■

Lemma 5.2. *Let $\mathcal{R}_U \in \text{Sp}(4d, \mathbb{R}) \cap \text{O}(4d, \mathbb{R})$ for some $U \in \text{U}(2d, \mathbb{C})$. Then the following are equivalent.*

(i) There exist $f, g \in L^2(\mathbb{R}^d)$, $f \not\equiv 0$, $g \not\equiv 0$, with

$$|\text{supp } W_{\mathcal{R}_U}(f, g)| < \infty.$$

(ii) For all $\tau \in \mathbb{T}$ there exist $f_\tau, g_\tau \in L^2(\mathbb{R}^d)$, $f_\tau \not\equiv 0$, $g_\tau \not\equiv 0$, with

$$|\text{supp } W_{\mathcal{R}_{\tau U}}(f_\tau, g_\tau)| < \infty.$$

Proof. Let $\tau \in \mathbb{T}$. Then for all $f, g \in L^2(\mathbb{R}^d)$ holds

$$\begin{aligned} |W_{\mathcal{R}_U}(f, g)| &= |\mu(\mathcal{R}_{\tau U})\mu(\mathcal{R}_{\bar{\tau}I})(f \otimes \bar{g})| \\ &\stackrel{(4.3)}{=} |\mu(\mathcal{R}_{\tau U})(\mu(\mathcal{R}_{\bar{\tau}I})f \otimes \overline{\mu(\mathcal{R}_{\tau I}g)})|. \end{aligned}$$

The claim follows by choosing

$$f_\tau = \mu(\mathcal{R}_{\bar{\tau}I})f, \quad g_\tau = \mu(\mathcal{R}_{\tau I})g.$$

■

The characterization of $W_{\mathcal{A}}$, $\mathcal{A} \in \text{Sp}(4d, \mathbb{R})$, which obey a Benedicks-type uncertainty principle is a direct consequence of Theorem 3.3, Lemma 5.1, and Lemma 5.2.

Theorem 5.3. *Let $\mathcal{A} \in \text{Sp}(4, \mathbb{R})$ be an arbitrary symplectic matrix with pre-Iwasawa decomposition*

$$\mathcal{A} = \mathcal{V}_Q \mathcal{D}_L \mathcal{R}_U.$$

and $\tau \in \mathbb{T}$ such that $\text{Im}(\tau U)$ is invertible. Assume $P = \text{Im}(\tau U)^{-1} \text{Re}(\tau U)$ is not block-diagonal. Then a Benedicks-type uncertainty principle holds:

If $W_{\mathcal{A}}(f, g)$ is supported on a set of finite measure for some $f, g \in L^2(\mathbb{R}^d)$, then $f \equiv 0$ or $g \equiv 0$.

Proof. By Lemma 5.1, it suffices to show that there exist no $f, g \in L^2(\mathbb{R}^d)$, $f \not\equiv 0$, $g \not\equiv 0$, such that $W_{\mathcal{R}_U}(f, g)$ is supported on a set of finite measure. Furthermore, by Lemma 5.2, it suffices to show that $W_{\mathcal{R}_{\tau U}}(f, g)$ does not have support of finite measure unless f or g are the zero function.

By Proposition 2.1,

$$\mathcal{R}_{\tau U} = \mathcal{V}_{\text{Re}(\tau U)\text{Im}(\tau U)^{-1}} \mathcal{D}_{\text{Im}(\tau U)} \mathcal{J} \mathcal{V}_P.$$

By Lemma 5.1, $W_{\mathcal{R}_{\tau U}}(f, g)$ is supported on a set of finite measure if and only if $W_{\mathcal{J} \mathcal{V}_P}(f, g)$ is supported on a set of finite measure. By Theorem 3.3, if the support of $W_{\mathcal{J} \mathcal{V}_P}(f, g)$ has finite measure, then $f \equiv 0$ or $g \equiv 0$, as claimed. ■

By Lemma 4.2, there always exists a $\tau \in \mathbb{T}$ such that $\text{Im}(\tau U)$ is invertible, so Theorem 5.3 can be applied to any symplectic matrix. However, given an arbitrary orthogonal symplectic matrix $\mathcal{A} \in \text{Sp}(4d, \mathbb{R})$, it is not obvious whether it can be related through the reductions above to a free symplectic matrix of the form $\mathcal{J} \mathcal{V}_P$ with a block-diagonal symmetric matrix $P \in \mathbb{R}^{2d \times 2d}$ or not. Thus we revisit the decomposition of free orthogonal symplectic matrices in Proposition 2.1, i.e.,

$$\mathcal{A} = \mathcal{R}_U = \mathcal{V}_{\text{Re}(U)\text{Im}(U)^{-1}} \mathcal{D}_{\text{Im}(U)} \mathcal{J} \mathcal{V}_P, \quad P = \text{Im}(U)^{-1} \text{Re}(U).$$

We first characterize when P is block-diagonal in terms of the joint SVD decomposition of the real and the imaginary part of U , or equivalently (Lemma 5.4). The derived characterization is invariant under multiplication with τI . By Lemma 4.2, this provides a characterization of the orthogonal symplectic matrices $\mathcal{A} \in \text{Sp}(2d, \mathbb{R})$ such that

$$\mathcal{A} = \mathcal{V}_{\text{Re}(\tau U) \text{Im}(\tau U)^{-1}} \mathcal{D}_{\text{Im}(\tau U)} \mathcal{J} \mathcal{V}_P \mathcal{R}_{\bar{\tau} I}, \quad P = \text{Im}(\tau U)^{-1} \text{Re}(\tau U),$$

with a block-diagonal symmetric $P \in \mathbb{R}^{d \times d}$ for some $\tau \in \mathbb{T}$.

Lemma 5.4. *Let $U = A + iB \in \text{U}(2d, \mathbb{C})$ be a unitary matrix with an invertible imaginary part $B \in \text{GL}(2d, \mathbb{R})$ and let $P = B^{-1}A$. Then the following are equivalent.*

(i) *There exist symmetric matrices $P_1, P_2 \in \mathbb{R}^{d \times d}$ such that*

$$P = \text{diag}(P_1, P_2).$$

(ii) *There exist orthogonal matrices $W \in \text{O}(2d, \mathbb{R})$, $V_1, V_2 \in \text{O}(d, \mathbb{R})$ and diagonal unitary matrices $\Sigma_1, \Sigma_2 \in \text{U}(d, \mathbb{C})$ with*

$$U = W \cdot \text{diag}(\Sigma_1, \Sigma_2) \cdot \text{diag}(V_1, V_2)^t = W \cdot \text{diag}(\Sigma_1 V_1^t, \Sigma_2 V_2^t).$$

(iii) *The unitary matrix $U^t U$ is block-diagonal.*

Proof. We will show the chain of implications (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i) $\Rightarrow$ (ii).

A simple computation shows that (ii) implies (iii).

We now show that (iii) implies (i). Since $U = A + iB$ is unitary, its real and its imaginary part satisfy

$$I = \text{Re}(U^* U) = A^t A + B^t B \quad \text{and} \quad 0 = \text{Im}(U^* U) = A^t B - B^t A.$$

Then

$$U^t U = A^t A - B^t B + i(A^t B + B^t A) = I - 2B^t B + 2iB^t A.$$

By assumption, $U^t U$ is block-diagonal, thus its real part $I - 2B^t B$ and its imaginary part $2B^t A$ are both block-diagonal. Therefore, $B^t B$ is block-diagonal, and invertible by assumption on B . From this, we conclude that

$$P = B^{-1}A = B^{-1}B^{-t}B^t A = (B^t B)^{-1}(B^t A)$$

is block-diagonal as a product of block-diagonal matrices.

We conclude with the implication (i) $\Rightarrow$ (ii). Since the matrix $P = B^{-1}A$ is symmetric, the matrix $I + P^2$ is positive definite, and the following matrix is well-defined:

$$U' = (I + P^2)^{-1/2}(P + iI) = A' + iB'.$$

The matrix U' is also unitary, which can be easily verified by computing $U' U'^*$. Furthermore,

$$A' = (I + P^2)^{-1/2}P \quad \text{and} \quad B' = (I + P^2)^{-1/2},$$

thus $B'^{-1}A = P$. Since B, B' are invertible, $W' = BB'^{-1} \in \text{GL}(d, \mathbb{R})$, and consequently,

$$A = BP = W'B'P = W'A'.$$

Thus $U = W'U'$. By assumption, P is block-diagonal, hence U' is block-diagonal as well. By Lemma 4.1, there exists matrices $W_1, W_2, V_1, V_2 \in \text{O}(d, \mathbb{R})$ and diagonal matrices $\Sigma_1, \Sigma_2 \in \text{U}(d, \mathbb{C})$ such that

$$U' = \begin{pmatrix} W_1 & 0 \\ 0 & W_2 \end{pmatrix} \cdot \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \cdot \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}$$

We set

$$W = W' \cdot \text{diag}(W_1, W_2)$$

to obtain the claimed factorization. ■

We now formulate the main result.

Theorem 5.5. *Let $\mathcal{A} \in \text{Sp}(4, \mathbb{R})$ be a symplectic matrix with pre-Iwasawa decomposition*

$$\mathcal{A} = \mathbf{v}_Q \mathbf{D}_L \mathbf{R}_U.$$

Then the following statements hold.

(i) *Assume $U^t U$ is not block-diagonal. Then a Benedicks-type uncertainty principle holds:*

If $W_{\mathcal{A}}(f, g)$ is supported on a set of finite measure for some $f, g \in L^2(\mathbb{R}^d)$, then $f \equiv 0$ or $g \equiv 0$.

(ii) *Assume $U^t U$ is block-diagonal. Then there exist $f, g \in L^2(\mathbb{R}^d)$, $f \not\equiv 0$, $g \not\equiv 0$, such that $W_{\mathcal{A}}(f, g)$ is supported on a compact set, in particular, on a set of finite measure.*

Proof. The uncertainty principle is a direct consequence of Theorem 5.3 and Lemma 5.4. To see this, let $f, g \in L^2(\mathbb{R}^d)$ such that $W_{\mathcal{A}}(f, g)$ is supported on a set of finite measure. We aim to show that $f \equiv 0$ or $g \equiv 0$.

By Lemma 4.2, there exists a $\tau \in \mathbb{T}$ such that $\text{Im}(\tau U)$ is invertible. By assumption, the matrix

$$(\tau U)^t (\tau U) = \tau^2 U^t U$$

is not block-diagonal. By Lemma 5.4, the symmetric matrix $P = \text{Im}(\tau U)^{-1} \text{Re}(\tau U)$ is not block-diagonal. Theorem 5.3 implies that $f \equiv 0$ or $g \equiv 0$.

We now prove the second statement. If $U^t U$ is block-diagonal, so is the matrix $(\tau U)^t (\tau U) = \tau^2 U^t U$ for any $\tau \in \mathbb{T}$. By Lemma 4.2, there exists a $\tau \in \mathbb{T}$ such that $\text{Im}(\tau U)$ is invertible and Lemma 5.4 applies. Thus, there exist $W \in \text{O}(2d, \mathbb{R})$, $V_1, V_2 \in \text{O}(d, \mathbb{R})$ and diagonal unitary matrices $\Sigma_{1,\tau}, \Sigma_{2,\tau} \in \text{U}(d, \mathbb{C})$ with

$$\tau U = W \cdot \text{diag}(\Sigma_{1,\tau}, \Sigma_{2,\tau}) \cdot \text{diag}(V_1, V_2)^t = W \cdot \text{diag}(\Sigma_{1,\tau} V_1^t, \Sigma_{2,\tau} V_2^t).$$

Set $\Sigma_1 = \bar{\tau} \Sigma_{1,\tau}$ and $\Sigma_2 = \bar{\tau} \Sigma_{2,\tau}$. Then

$$U = W \cdot \text{diag}(\Sigma_1, \Sigma_2) \cdot \text{diag}(V_1, V_2)^t = W \cdot \text{diag}(\Sigma_1 V_1^t, \Sigma_2 V_2^t).$$

By Remark 2.3, for all $f, g \in L^2(\mathbb{R}^d)$

$$\begin{aligned} |W_{\mathcal{A}}(f, g)| &= |\mu(\mathcal{V}_Q)\mu(\mathcal{D}_L)\mu(\mathcal{D}_W)\mu(\mathcal{R}_{\text{diag}(\Sigma_1 V_1^t, \Sigma_2 V_2^t)})(f \otimes \bar{g})| \\ &= |\det L|^{-1/2} |(\mu(\mathcal{R}_{\Sigma_1 V_1^t})f \otimes \overline{\mu(\mathcal{R}_{\Sigma_2 V_2^t}g)})(W^t L^{-1} \cdot)|. \end{aligned}$$

Since $\mu(\mathcal{R}_{\Sigma_1 V_1^t}), \mu(\mathcal{R}_{\Sigma_2 V_2^t}) \in \text{Mp}(2d, \mathbb{R})$ are unitary operators, by choosing compactly supported $f_0, g_0 \in L^2(\mathbb{R}^d)$, $f \neq 0, g \neq 0$, the functions

$$f = \mu(\mathcal{R}_{\Sigma_1 V_1^t})^{-1} f_0 \neq 0, \quad g = \mu(\mathcal{R}_{\Sigma_2 V_2^t})^{-1} g_0 \neq 0$$

are square-integrable and have a compactly supported $\mathcal{A}$ -Wigner distribution

$$|W_{\mathcal{A}}(f, g)| = |\det L|^{-1/2} |(f_0 \otimes \bar{g}_0)(W^t L^{-1} \cdot)|.$$

■

We summarize Theorem 5.5 in an algorithm.

Input: $\mathcal{A} \in \text{Sp}(4d, \mathbb{R})$.

- (1) Compute the pre-Iwasawa decomposition $\mathcal{A} = \mathcal{V}_Q \mathcal{D}_L \mathcal{R}_U$.
 - (2) Compute $U^t U$.
 - (3) If $U^t U$ is block-diagonal, then there exist $f, g \in L^2(\mathbb{R}^d)$, $f \neq 0, g \neq 0$, such that $W_{\mathcal{A}}(f, g)$ is supported on a compact set.
 - (4) Otherwise, whenever $W_{\mathcal{A}}(f, g)$ is supported on a set of finite measure, then $f \equiv 0$ or $g \equiv 0$.
-

We refer to [5, 31, 33] for the numerical treatment of the Iwasawa decomposition of the symplectic group. Note that the Iwasawa composition is typically in the literature, in particular, in [5, 31, 33], in the form

$$\begin{aligned} \mathcal{A} &= \begin{pmatrix} A & B \\ -B & A \end{pmatrix} \cdot \begin{pmatrix} \text{diag}(a_1, \dots, a_d) & 0 \\ 0 & \text{diag}(a_1^{-1}, \dots, a_d^{-1}) \end{pmatrix} \cdot \begin{pmatrix} L & LQ \\ 0 & L^{-t} \end{pmatrix} \\ &= \mathcal{R}_{A+iB} \cdot \mathcal{D}_{\text{diag}(a_1, \dots, a_d)} \cdot \mathcal{D}_L \mathcal{V}_{Q^t}^t. \end{aligned}$$

In order to work with our desired order, one can, for instance, compute instead the Iwasawa decomposition of $\mathcal{A}^t$. The desired NAK-Iwasawa decomposition is obtained by transposing this factorization. Alternatively, one can use a pre-Iwasawa decomposition [12], not necessarily the one derived from the Iwasawa decomposition, to conclude whether $W_{\mathcal{A}}$ obeys a Benedicks-type of an uncertainty principle.

6. QUADRATIC UNCERTAINTY PRINCIPLES

This section is dedicated to uncertainty principles with an additional restriction on the pair (f, g) , namely, $g = f$. The motivation stems from the Rihaczek distribution

$$R(f, g)(x, \omega) = e^{-2\pi i x^t \omega} f(x) \overline{\mathcal{F}g(\omega)},$$

in which case we can clearly find functions $f, g \in L^2(\mathbb{R}^d)$, $f \neq 0, g \neq 0$, such that $R(f, g)$ is compactly supported. However, with the additional assumption $g = f$, $R(f, f)$ cannot be supported on a set of finite measure unless f is the zero function. This is precisely a reformulation of Benedicks's uncertainty principle [1, 3].

Before we state the main result, we prove a technical lemma analogous to Lemma 5.4 which is better suited for this setting.

Lemma 6.1. *Let $U = A + iB \in \mathrm{U}(2d, \mathbb{C})$ be a unitary matrix. Then the following statements are equivalent.*

(i) *There exist an orthogonal matrix $W \in \mathrm{O}(2d, \mathbb{R})$ and a unitary matrix $\Delta \in \mathrm{U}(d, \mathbb{C})$ such that*

$$U = W \cdot \mathrm{diag}(\Delta, \overline{\Delta}).$$

(ii) *There exist an orthogonal matrix $W \in \mathrm{O}(2d, \mathbb{R})$ and two unitary matrices $\Delta_1, \Delta_2 \in \mathrm{U}(d, \mathbb{C})$ such that $\Delta_1 \Delta_2^t \in \mathbb{R}^{d \times d}$ and*

$$U = W \cdot \mathrm{diag}(\Delta_1, \Delta_2).$$

(iii) *There exists a unitary matrix $\Delta \in \mathrm{U}(2d, \mathbb{C})$ such that*

$$U^t U = \mathrm{diag}(\Delta, \overline{\Delta}).$$

Proof. We show the following chain of implications

$$(i) \Rightarrow (iii) \Rightarrow (ii) \Rightarrow (i).$$

A simple computation shows that (i) implies (iii).

We now assume that (iii) holds and we show (ii). Let $\tau \in \mathbb{T}$ such that $\mathrm{Im}(\tau U)$ is invertible as per Lemma 4.2. By Lemma 5.4, there exist $W \in \mathrm{O}(2d, \mathbb{R})$, $V_1, V_2 \in \mathrm{O}(d, \mathbb{R})$ and diagonal unitary matrices $\Sigma_{1,\tau}, \Sigma_{2,\tau} \in \mathrm{U}(d, \mathbb{C})$ with

$$\tau U = W \cdot \mathrm{diag}(\Sigma_{1,\tau}, \Sigma_{2,\tau}) \cdot \mathrm{diag}(V_1, V_2)^t.$$

Set $\Sigma_1 = \bar{\tau} \Sigma_{1,\tau}$, $\Sigma_2 = \bar{\tau} \Sigma_{2,\tau} \in \mathrm{U}(d, \mathbb{C})$ and $\Delta_1 = \Sigma_1 V_1^t$, $\Delta_2 = \Sigma_2 V_2^t \in \mathrm{U}(d, \mathbb{C})$. Then

$$U = W \cdot \mathrm{diag}(\Sigma_1, \Sigma_2) \cdot \mathrm{diag}(V_1, V_2)^t = W' \cdot \mathrm{diag}(\Delta_1, \Delta_2).$$

By assumption,

$$\mathrm{diag}(\Delta, \overline{\Delta}) = U^t U = \begin{pmatrix} \Delta_1^t \Delta_1 & 0 \\ 0 & \Delta_2^t \Delta_2 \end{pmatrix},$$

hence

$$(6.1) \quad \overline{\Delta_1^t \Delta_1} = \Delta_2^t \Delta_2.$$

We need to show that $\Delta_1 \Delta_2^t$ is real-valued. Indeed, by multiplying (6.1) with Δ_1 from the left and $\Delta_2^{-1} = \Delta_2^* = \overline{\Delta_2^t}$ from the right,

$$\Delta_1 \Delta_1^* \overline{\Delta_1^t \Delta_1} \Delta_2^* = \Delta_1 \Delta_2^t \Delta_2 \Delta_2^* \quad \Leftrightarrow \quad \overline{\Delta_1 \Delta_2^t} = \Delta_1 \Delta_2^t,$$

as claimed.

If (ii) holds, i.e., if there exist $W' \in \mathrm{O}(2d, \mathbb{R})$ and $\Delta_1, \Delta_2 \in \mathrm{U}(d, \mathbb{C})$ with $W_0 = \Delta_1 \Delta_2^t \in \mathbb{R}^{d \times d}$ such that

$$U = W' \cdot \mathrm{diag}(\Delta_1, \Delta_2),$$

then W_0 is orthogonal. Using $\Delta_1^{-1} = \overline{\Delta_1^t}$,

$$\Delta_2 = (\Delta_1^{-1} W_0)^t = W_0^t \overline{\Delta_1}.$$

We set $W = W' \cdot \text{diag}(I, W_0^t) \in \text{O}(2d, \mathbb{R})$ and $\Delta = \Delta_1 \in \text{U}(d, \mathbb{C})$. Then

$$U = W \cdot \text{diag}(\Delta, \bar{\Delta})$$

and (i) follows. ■

Before we formulate the main result for quadratic metaplectic time-frequency representations, we prove a Benedicks-type uncertainty principle for a pair of metaplectic operators $(\mathcal{R}_{U_1}, \mathcal{R}_{U_2}) \in \text{Sp}(2d, \mathbb{R})^2$.

Theorem 6.2. *Let $\mathcal{R}_{U_1}, \mathcal{R}_{U_2} \in \text{Sp}(2d, \mathbb{R})$ be a orthogonal symplectic matrices such that the matrix $U_1 U_2^*$ is not real-valued. Then a Benedicks-type uncertainty principle holds for the pair $(\mu(\mathcal{R}_{U_1})f, \mu(\mathcal{R}_{U_2})f)$:*

If both $\mu(\mathcal{R}_{U_1})f$ and $\mu(\mathcal{R}_{U_2})f$ are supported on a set of finite measure, then $f \equiv 0$.

Proof. The proof idea relies on Theorem 4.3, but it also incorporates the more sophisticated matrix decompositions used so far.

Assume $f \in L^2(\mathbb{R}^d)$ such that both $\mu(\mathcal{R}_{U_1})f$ and $\mu(\mathcal{R}_{U_2})f$ are supported on a set of finite measure. We aim to show that f is the zero function.

We set $U = U_1 U_2^*$ and $f_1 = \mu(\mathcal{R}_{U_2})f$. Then by Remark 2.3

$$\mu(\mathcal{R}_{U_1})f = \mu(\mathcal{R}_{U_1})\mu(\mathcal{R}_{U_2})^{-1}f_1 = \mu(\mathcal{R}_{U_1}\mathcal{R}_{U_2}^{-1})f_1 = \mu(\mathcal{R}_{U_1 U_2^*})f_1 = \mu(\mathcal{R}_U)f_1.$$

By construction, f_1 and $\mu(\mathcal{R}_U)f_1$ are supported on a set of finite measure.

By Lemma 4.1, there exist $W, V \in \text{O}(d, \mathbb{R})$ and a diagonal $\Sigma \in \text{U}(d, \mathbb{C})$ such that $U = W\Sigma V^t$. As per Remark 2.4, $\mathcal{R}_W = \mathcal{D}_W$ and $\mathcal{R}_{V^t} = \mathcal{D}_{V^t}$, and $\mu(\mathcal{D}_W), \mu(\mathcal{D}_{V^t})$ act as dilation operators, both $f_2 = \mu(\mathcal{D}_{V^t})f_1$ and

$$\mu(\mathcal{R}_\Sigma \mathcal{R}_{V^t})f_1 = \mu(\mathcal{R}_\Sigma \mathcal{D}_{V^t})f_1 = \mu(\mathcal{R}_\Sigma)f_2 = \mathcal{F}_\Sigma f_2$$

are supported on a set of finite measure. Furthermore, since W, V are real-valued but U is not, Σ is not real-valued. By Theorem 4.3 (Benedicks-type uncertainty principle for partial Fourier transforms), f_2 is the zero function, hence

$$f = \mu(\mathcal{R}_{U_2})^{-1}f_1 = \mu(\mathcal{R}_{U_2})^{-1}\mu(\mathcal{D}_{V^t})^{-1}f_2 \equiv 0,$$

which concludes the proof. ■

The previous Theorem has direct consequences for quadratic metaplectic time-frequency representations.

Theorem 6.3. *Let $\mathcal{A} \in \text{Sp}(4d, \mathbb{R})$ be a symplectic matrix with pre-Iwasawa decomposition*

$$\mathcal{A} = \mathcal{V}_Q \mathcal{D}_L \mathcal{R}_U.$$

Then the following statements hold.

(i) *Assume $U^t U \neq \text{diag}(\Delta, \bar{\Delta})$ for all unitary matrices $\Delta \in \text{U}(d, \mathbb{C})$. Then a Benedicks-type uncertainty principle holds:*

If $W_{\mathcal{A}}(f, f)$ is supported on a set of finite measure for some $f \in L^2(\mathbb{R}^d)$, then $f \equiv 0$.

(ii) Assume $U^t U = \text{diag}(\Delta, \overline{\Delta})$ for some unitary matrix $\Delta \in \text{U}(d, \mathbb{C})$. Then there exists a function $f \in L^2(\mathbb{R}^d)$, $f \not\equiv 0$, such that the support of $W_{\mathcal{A}}(f, f)$ is compact, in particular, has finite measure.

Proof. If $U^t U$ is not block-diagonal, then (i) is a special case of Theorem 5.5 with $g = f$. Thus, from now on, we can assume that $U^t U$ is block-diagonal. By Lemma 5.1, we may assume that $Q = 0$ and $L = I$. Further let $f \in L^2(\mathbb{R}^d)$ be a function whose metaplectic time-frequency representation $W_{\mathcal{A}}(f, f)$ is supported on a set of finite measure. We aim to show that f has to be the zero function.

By Lemma 6.1, there exists a matrix $W \in \text{O}(2d, \mathbb{R})$ and unitary matrices $\Delta_1, \Delta_2 \in \text{U}(d, \mathbb{C})$ with $\Delta_1 \Delta_2^t \notin \mathbb{R}^{d \times d}$ such that

$$U = W \cdot \text{diag}(\Delta_1, \Delta_2).$$

Recall that by Remark 2.3 the action of $\mu(\mathcal{R}_{\text{diag}(\Delta_1, \Delta_2)})$ tensorizes, and by Remark 2.4 $\mu(\mathcal{R}_W) = \mu(\mathcal{D}_W)$. By Lemma 5.1, the support of the function F_1 below has finite measure.

$$\begin{aligned} F_1 &= |\mu(\mathcal{R}_{\text{diag}(\Delta_1, \Delta_2)})(f \otimes \overline{f})| \\ (6.2) \quad &= |\mu(\mathcal{R}_{\Delta_1})f| \otimes |\mu(\mathcal{R}_{\Delta_2})\overline{f}| \\ &= |\mu(\mathcal{R}_{\Delta_1})f| \otimes |\mu(\mathcal{R}_{\overline{\Delta_2}})f|, \end{aligned}$$

where the last equality is due to (4.4), Lemma 4.1, and the obvious formula

$$\mu(\mathcal{R}_V)\overline{f} = \mu(\mathcal{D}_V)\overline{f} = \overline{\mu(\mathcal{D}_V)f}, \quad V \in \text{O}(d, \mathbb{R}).$$

Let $f_1 = \mu(\mathcal{R}_{\overline{\Delta_2}})f$. Then we can rewrite F_1 as

$$F_1 = |\mu(\mathcal{R}_{\Delta_1} \mathcal{R}_{\overline{\Delta_2}}^{-1})f_1| \otimes |f_1| = |\mu(\mathcal{R}_{\Delta_1 \Delta_2^t})f_1| \otimes |f_1|.$$

If F_1 is supported on a finite measure, then both f_1 and $\mu(\mathcal{R}_{\Delta_1 \Delta_2^t})f_1$ are supported on a set of finite measure. By Theorem 6.2, $f_1 \equiv 0$, thus $f = \mu(\mathcal{R}_{\overline{\Delta_2}})^{-1}f_1 \equiv 0$.

We conclude with the proof of (ii). By Lemma 6.1, there exists an orthogonal matrix $W \in \text{O}(2d, \mathbb{R})$ and a unitary matrix $\Delta \in \text{U}(d, \mathbb{C})$ such that

$$U = W \cdot \text{diag}(\Delta, \overline{\Delta}).$$

Let $f_0 \in L^2(\mathbb{R}^d)$, $f_0 \not\equiv 0$, be supported on a compact set. We set $f = \mu(\mathcal{R}_{\Delta})^{-1}f_0$. Then $f \in L^2(\mathbb{R}^d)$ and $f \not\equiv 0$ because $\mu(\mathcal{R}_{\Delta})^{-1}$ is a unitary operator.

We recall that by Remark 2.4 $\mathcal{R}_W = \mathcal{D}_W$. By Lemma 5.1, we can assume without loss of generality that $W = I$. Then $W_{\mathcal{A}}(f, f)$ is supported on a compact set because, analogously to (6.2),

$$|W_{\mathcal{A}}(f, f)| = |\mu(\mathcal{R}_{\Delta \Delta^{-1}})\mu(\mathcal{R}_{\Delta})f| \otimes |\mu(\mathcal{R}_{\Delta})f| = |f_0| \otimes |f_0|,$$

which is clearly supported on a compact set. ■

Remark 6.4. We remark that the same proof can clearly be easily modified for $g = \iota f$, where ι is an involution ι on $L^2(\mathbb{R}^d)$ which has a simple interaction with the fractional Fourier transforms, e.g., $f \mapsto \overline{f}$ or $f \mapsto \overline{f}(-\cdot)$, but the condition $\Delta_1 \Delta_2^t \in \mathbb{R}^{d \times d}$ needs to be replaced with $\Delta_1 \Delta_2^{-1} = \Delta_1 \overline{\Delta_2^t} \in \mathbb{R}^{d \times d}$.

7. EXAMPLES

In this section, we show how to apply Theorem 5.5 and Theorem 6.3 and to recover the known uncertainty principles for the short-time Fourier transform, the ambiguity function, and the τ -Wigner distribution. In the following, $I = I_d$ and $0 = 0_{d \times d}$ are the identity and the zero matrix in $\mathbb{R}^{d \times d}$ respectively.

The short-time Fourier transform is given by

$$V_g f(x, \omega) = \int_{\mathbb{R}^d} f(t) \overline{g(t-x)} e^{2\pi i \omega^t t} dt.$$

Let

$$L_{\text{STFT}} = \begin{pmatrix} I & -I \\ I & 0 \end{pmatrix} \in \text{GL}(2d, \mathbb{R}) \quad \text{with inverse} \quad L_{\text{STFT}}^{-1} = \begin{pmatrix} 0 & I \\ -I & I \end{pmatrix}$$

(the matrix L_{STFT}^{-1} corresponds to the linear map $(x, t) \mapsto (t, t-x)$). Then $V_g f$ can be written as the $\mathcal{A}$ -Wigner distribution

$$V_g f = \mu(\mathcal{A}_{\text{STFT}})(f \otimes \bar{g}) = \mu(\mathcal{R}_{\text{diag}(I, iI)} \mathcal{D}_{L_{\text{STFT}}})(f \otimes \bar{g}) \quad \text{with the matrix}$$

$$\mathcal{A}_{\text{STFT}} = \begin{pmatrix} I & -I & 0 & 0 \\ 0 & 0 & I & I \\ 0 & 0 & 0 & -I \\ -I & 0 & 0 & 0 \end{pmatrix}.$$

The pre-Iwasawa decomposition of $\mathcal{A}_{\text{STFT}}$ is given by

$$(7.1) \quad \mathcal{A}_{\text{STFT}} = \mathcal{V}_{-\frac{1}{2}} \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \cdot \mathcal{D}_{2^{1/2}I} \cdot \mathcal{R}_{U_{\text{STFT}}}, \quad U_{\text{STFT}} = \frac{1}{\sqrt{2}} \begin{pmatrix} I & -I \\ iI & iI \end{pmatrix} \in \text{U}(2d, \mathbb{C}).$$

The deciding product is

$$U_{\text{STFT}}^t U_{\text{STFT}} = \begin{pmatrix} 0 & -I \\ -I & 0 \end{pmatrix},$$

which is not block-diagonal, thus a Benedicks-type uncertainty principle holds for the short-time Fourier transform [20, 25, 27, 38].

The ambiguity function satisfies

$$A(f, g)(x, \omega) = e^{\pi i x \cdot \omega} V_g f(x, \omega) = \mu(\mathcal{V}_{\frac{1}{2}} \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}) V_g f(x, \omega), \quad x, \omega \in \mathbb{R}^d,$$

thus is given by $A(f, g) = \mu(\mathcal{A}_A)(f, g)$, where

$$\mathcal{A}_A = \mathcal{V}_{\frac{1}{2}} \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \cdot \mathcal{A}_{\text{STFT}} \begin{pmatrix} I & -I & 0 & 0 \\ 0 & 0 & I & I \\ 0 & 0 & \frac{1}{2}I & -\frac{1}{2}I \\ -\frac{1}{2}I & -\frac{1}{2}I & 0 & 0 \end{pmatrix}.$$

By (7.1), the pre-Iwasawa decomposition of $\mathcal{A}_A$ is given by

$$\mathcal{A}_A = \mathcal{V}_0 \cdot \mathcal{D}_{\sqrt{2}I} \cdot \mathcal{R}_{U_A}, \quad U_A = \frac{1}{\sqrt{2}} \begin{pmatrix} I & -I \\ iI & iI \end{pmatrix} \in \text{U}(2d, \mathbb{C}),$$

and

$$U_A^t U_A = \begin{pmatrix} 0 & -I \\ -I & 0 \end{pmatrix},$$

hence it obeys a Benedicks-type uncertainty principle [20, 25, 27, 38]. **The τ -Wigner distribution** is given by

$$W_\tau(f, g)(x, \omega) = \int_{\mathbb{R}^d} f(x + \tau t) \overline{g(x - (1 - \tau)t)} e^{-2\pi i \omega t} dt.$$

This family of time-frequency representations includes the Wigner distribution for $\tau = \frac{1}{2}$ and the Rihaczek distribution for $\tau = 0$. By choosing

$$L_\tau = \begin{pmatrix} I - \tau I & \tau I \\ I & -I \end{pmatrix} \in \text{GL}(2d, \mathbb{R}) \quad \text{with inverse} \quad L_\tau^{-1} = \begin{pmatrix} I & \tau I \\ I & -(1 - \tau)I \end{pmatrix},$$

$W_\tau(f, g)$ can be written as

$$W_\tau(f, g) = \mu(\mathcal{A}_\tau)(f \otimes \bar{g}) = \mu(\mathcal{R}_{\text{diag}(I, iI)} \mathcal{D}_{L_\tau})(f \otimes \bar{g}) \quad \text{with the matrix}$$

$$\mathcal{A}_\tau = \begin{pmatrix} (1 - \tau)I & \tau I & 0 & 0 \\ 0 & 0 & \tau & -(1 - \tau)I \\ 0 & 0 & I & I \\ -I & I & 0 & 0 \end{pmatrix}.$$

We set

$$\alpha_\tau := (1 - 2\tau + 2\tau^2)^{1/2} = ((1 - \tau)^2 + \tau^2)^{1/2}.$$

The pre-Iwasawa decomposition of $\mathcal{A}_\tau$ is given by

$$\mathcal{A}_\tau = \mathcal{V}_{-\frac{(1-2\tau)}{\alpha_\tau^2} \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}} \cdot \mathcal{D}_{\alpha_\tau I} \cdot \mathcal{R}_{U_\tau}, \quad U_\tau = \frac{1}{\alpha_\tau} \begin{pmatrix} 1 - \tau & \tau \\ i\tau & -i(1 - \tau) \end{pmatrix} \in \text{U}(2d, \mathbb{C}).$$

The deciding product is given by

$$U_\tau^t U_\tau = \frac{1}{\alpha_\tau^2} \begin{pmatrix} (1 - 2\tau)I & 2(1 - \tau)\tau I \\ 2(1 - \tau)\tau I & -(1 - 2\tau)I \end{pmatrix}.$$

The matrix is block-diagonal if and only if $\tau = 0$ or $\tau = 1$, as expected for the sesquilinear form (Theorem 5.5). Concerning the uncertainty principle for quadratic time-frequency representations in Theorem 6.3, if $\tau \in \{0, 1\}$, then

$$U_0^t U_0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \quad U_1^t U_1 = \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix}.$$

Therefore, for all $\tau \in \mathbb{T}$, there exists no unitary matrix $\Delta \in \text{U}(d, \mathbb{C})$ such that $U_\tau^t U_\tau = \text{diag}(\Delta, \overline{\Delta})$. Hence, for all $\tau \in \mathbb{T}$, the only function $f \in L^2(\mathbb{R}^d)$ such that $W_\tau(f, f)$ is supported on a set of finite measure is the zero function.

8. DISCUSSION AND OUTLOOK

This approach provides an idea in the spirit of [26] for the generalization of existing uncertainty principles for the short-time Fourier transform or the Rihaczek distribution to uncertainty principles for metaplectic time-frequency representation. However, this seems to be highly dependent on the nature of the investigated uncertainty principle and we are working on a more encompassing method to cover a broad variety of uncertainty principles.

REFERENCES

- [1] AMREIN, W. O., AND BERTHIER, A. M. On support properties of L^p -functions and their Fourier transforms. *J. Functional Analysis* 24, 3 (1977), 258–267.
- [2] BAYER, D. *Bilinear Time-Frequency Distributions and Pseudodifferential Operators*. PhD thesis, University of Vienna, Vienna, Austria, 2010.
- [3] BENEDICKS, M. On Fourier transforms of functions supported on sets of finite Lebesgue measure. *J. Math. Anal. Appl.* 106, 1 (1985), 180–183.
- [4] BÉNYI, A., AND OKOUDJOU, K. A. *Modulation spaces—with applications to pseudodifferential operators and nonlinear Schrödinger equations*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2020.
- [5] BENZI, M., AND RAZOUK, N. On the Iwasawa decomposition of a symplectic matrix. *Appl. Math. Lett.* 20, 3 (2007), 260–265.
- [6] BONAMI, A., AND DEMANGE, B. A survey on uncertainty principles related to quadratic forms. *Collect. Math.*, Vol. Extra (2006), 1–36.
- [7] CARMONA, R., HWANG, W.-L., AND TORRÉSANI, B. *Practical time-frequency analysis*, vol. 9 of *Wavelet Analysis and its Applications*. Academic Press, Inc., San Diego, CA, 1998. Gabor and wavelet transforms with an implementation in S, With a preface by Ingrid Daubechies.
- [8] CORDERO, E., AND RODINO, L. *Time-frequency analysis of operators*, vol. 75 of *De Gruyter Studies in Mathematics*. De Gruyter, Berlin, 2020.
- [9] CORDERO, E., AND RODINO, L. Wigner analysis of operators. Part I: Pseudodifferential operators and wave fronts. *Appl. Comput. Harmon. Anal.* 58 (2022), 85–123.
- [10] CORDERO, E., AND RODINO, L. Characterization of modulation spaces by symplectic representations and applications to Schrödinger equations. *J. Funct. Anal.* 284, 9 (2023), Paper No. 109892, 40.
- [11] DE GOSSON, M. *Symplectic Geometry and Quantum Mechanics*. Birkhäuser Basel, 2006.
- [12] DE GOSSON, M. A. *Symplectic Methods in Harmonic Analysis and in Mathematical Physics*. Springer Basel, 2011.
- [13] FEICHTINGER, H. G. Modulation Spaces: Looking Back and Ahead. *Sampl. Theory Signal Process. Data Anal.* 5, 2 (2006), 109–140.
- [14] FLANDRIN, P. *Time-frequency/time-scale analysis*, vol. 10 of *Wavelet Analysis and its Applications*. Academic Press, Inc., San Diego, CA, 1999. With a preface by Yves Meyer, Translated from the French by Joachim Stöckler.
- [15] FOLLAND, G. B. *Harmonic Analysis in Phase Space. (AM-122)*. Princeton University Press, 1989.
- [16] FOLLAND, G. B., AND SITARAM, A. The uncertainty principle: a mathematical survey. *J. Fourier Anal. Appl.* 3, 3 (1997), 207–238.
- [17] FÜHR, H., AND RZESZOTNIK, Z. A note on factoring unitary matrices. *Linear Algebra Appl.* 547 (2018), 32–44.
- [18] GABOR, D. Theory of communication. Part 1: The analysis of information. *J. IEE (London)*, 93(III) 93, 26 (1946), 429–441.
- [19] GRÖCHENIG, K. Uncertainty principles for time-frequency representations. In *Advances in Gabor analysis*, Appl. Numer. Harmon. Anal. Birkhäuser Boston, Boston, MA, 2003, pp. 11–30.
- [20] GRÖCHENIG, K., AND ZIMMERMANN, G. Hardy’s theorem and the short-time Fourier transform of Schwartz functions. *J. London Math. Soc. (2)* 63, 1 (2001), 205–214.
- [21] GRÖCHENIG, K. *Foundations of Time-Frequency Analysis*. Birkhäuser Boston, 2001.
- [22] HALL, B. *Lie groups, Lie algebras, and representations*, second ed., vol. 222 of *Graduate Texts in Mathematics*. Springer International Publishing, 2015.
- [23] HLAWATSCH, F., AND BOUDREAUX-BARTELS, G. Linear and quadratic time-frequency signal representations. *IEEE Signal Process. Mag.* 9, 2 (Apr. 1992), 21–67.

- [24] IWASAWA, K. On some types of topological groups. *Ann. Math.* 50, 3 (1949), 507–558.
- [25] JAMING, P. Principe d’incertitude qualitatif et reconstruction de phase pour la transformée de Wigner. *C. R. Acad. Sci. Paris Sér. I Math.* 327, 3 (1998), 249–254.
- [26] JAMING, P. A simple observation on the uncertainty principle for the fractional Fourier transform. *J. Fourier Anal. Appl.* 28, 3 (2022), Paper No. 51, 8.
- [27] JANSSEN, A. J. E. M. Proof of a conjecture on the supports of Wigner distributions. *J. Fourier Anal. Appl.* 4, 6 (1998), 723–726.
- [28] KNAPP, A. W. *Lie groups beyond an introduction*, vol. 140 of *Progress in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, 1996.
- [29] NAMIAS, V. The fractional order Fourier transform and its application to quantum mechanics. *J. Inst. Math. Appl.* 25, 3 (1980), 241–265.
- [30] RICAUD, B., AND TORRÉSANI, B. A survey of uncertainty principles and some signal processing applications. *Adv. Comput. Math.* 40, 3 (2014), 629–650.
- [31] SAWYER, P. Computing the Iwasawa decomposition of the classical Lie groups of noncompact type using the QR decomposition. *Linear Algebra Appl.* 493 (2016), 573–579.
- [32] SCHEMPP, W. Radar ambiguity functions, the Heisenberg group, and holomorphic theta series. *Proc. Amer. Math. Soc.* 92, 1 (1984), 103–110.
- [33] TAM, T.-Y. Computing the Iwasawa decomposition of a symplectic matrix by Cholesky factorization. *Appl. Math. Lett.* 19, 12 (2006), 1421–1424.
- [34] TER MORSCHÉ, H. G., AND OONINCX, P. J. On the Integral Representations for Metaplectic Operators. *J. Fourier Anal. Appl.* 8 (2002), 245–258.
- [35] TORRÉSANI, B. Time-frequency analysis, from geometry to signal processing. In *Contemporary problems in mathematical physics (Cotonou, 1999)*. World Sci. Publ., River Edge, NJ, 2000, pp. 74–96.
- [36] WEIL, A. Sur certains groupes d’opérateurs unitaires. *Acta Math.*, 111 (1964), 143–211.
- [37] WIENER, N. Hermitian polynomials and Fourier analysis. *J. Math. and Phys.* 8, 1–4 (Apr. 1929), 70–73.
- [38] WILCZOK, E. New uncertainty principles for the continuous Gabor transform and the continuous wavelet transform. *Doc. Math.* 5 (2000), 201–226.
- [39] WOODWARD, P. M. *Probability and information theory, with applications to radar*, second ed. Pergamon Press, Oxford-Edinburgh-New York-Paris-Frankfurt, 1964.
- [40] ZHANG, Z. Uncertainty principle for free metaplectic transformation. *J. Fourier Anal. Appl.* 29, 6 (2023), Paper No. 71, 33.
- [41] ZHANG, Z.-C. Uncertainty principle for linear canonical transform using matrix decomposition of absolute spread matrix. *Digit. Signal Process.* 89 (2019), 145–154.