# The Framework of $\alpha$-Molecules
## Theory and Applications

vorgelegt von
Diplom-Mathematiker Martin Schäfer
geb. in Köln

Von der Fakultät II - Mathematik und Naturwissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
Dr. rer. nat.

**genehmigte** Dissertation

Promotionsausschuss:

| | |
|---|---|
| Vorsitzender: | Prof. Dr. Wolfgang König |
| Gutachterin: | Prof. Dr. Gitta Kutyniok |
| Gutachterin: | Prof. Dr. Gerlind Plonka-Hoch |
| Gutachter: | Prof. Dr. Martin Ehler |

Tag der wissenschaftlichen Aussprache: 17. September 2018

Berlin 2018

# Deutsche Zusammenfassung

Waveletsysteme sind heutzutage ein integraler Bestandteil der harmonischen Analysis und dienen zum Beispiel als effizientes Werkzeug zur Darstellung und Approximation von Signalen. Ihr großer Erfolg beruht dabei unter anderem auf der Fähigkeit, glatte Signale mit lokalen Singularitäten besser zu approximieren als es traditionelle Fouriersysteme können. Bei isotropen Daten, welche insbesondere univariate Signale miteinschließen, ist ihre Performanz bei entsprechender Regularität sogar quasi-optimal.

Für die Approximation multivariater Daten hingegen sind Wavelets im allgemeinen nicht optimal geeignet. Der Grund hierfür liegt in ihrer isotropen Skalierung, die keine optimale Auflösung anisotroper Strukturen erlaubt. Da solche Strukturen für multivariate Daten jedoch sehr typisch sind – man denke nur an Kanten in Bilddaten zum Beispiel – sind in den letzten Jahre viele Anstrengungen unternommen worden, um diese Unzulänglichkeit zu überwinden. Insbesondere wurden viele neuartige sogenannte direktionale Repräsentationssysteme eingeführt, von denen wir als einige der bekanntesten Ridgelets, Curvelets und Shearlets nennen wollen.

Solche direktionalen Systeme lassen sich anhand der ihnen zugrundeliegenden Skalierung kategorisieren. Wavelets zum Beispiel sind isotroper Natur, eine rein direktionale Skalierung findet bei Ridgelets Verwendung, die Konstruktion klassischer Curvelets und Shearlets basiert auf einer parabolischen Skalierung. Eine Vielzahl unterschiedlicher Skalierungstypen wird durch das Konzept der $\alpha$-Skalierung abgedeckt, wo mit Hilfe eines Parameters $\alpha \in [0, 1]$ zwischen dem isotropen und dem direktionalen Fall interpoliert wird. Die vorgenannten Systeme zum Beispiel sind $\alpha$-skaliert mit zugehörigen Parametern $\alpha = 1$, $\alpha = 0$ und $\alpha = \frac{1}{2}$.

Das Hauptziel dieser Dissertation besteht darin, eine einheitliche Theorie für derartige $\alpha$-skalierte Repräsentationssysteme zu entwickeln. Den grundlegenden Begriff bilden dabei sogenannte $\alpha$-Moleküle, die eine Weiterentwicklung des Konzepts der parabolischen Moleküle darstellen. Letztere wurden eingeführt, um eine simultane Behandlung parabolisch skalierter Systeme zu ermöglichen.

Per Definition entstehen sie durch parabolische Skalierung sowie durch Rotation und Translation aus einer Menge generierender Funktionen, für die lediglich eine gemeinsame Zeit-Frequenz-Lokalisierung gefordert wird. Die Bezeichnung „Molekül“ rührt dabei von der möglichen Variabilität der Generatoren her. Zusammen mit der Verwendung sogenannter Parametrisierungen, welche eine generische Indizierung ermöglichen, bringt diese die nötige Flexibilität in die Konstruktion, um verschiedenartige parabolisch skalierter Systeme einheitlich zu beschreiben. Tatsächlich ist das Konzept allgemein genug, um sowohl rotations-basierte als auch scherungs-basierte Systeme wie die klassischen Curvelets und die klassischen Shearlets zu umfassen.

Nach dem Vorbild parabolischer Molekülsysteme werden auch $\alpha$-Molekülsysteme mittels Dilatation, Rotation und Translation aus einer zugrundeliegenden Generatormenge erzeugt, wobei die Generatoren wieder einer gemeinsamen Zeit-Frequenz-Lokalisierung unterliegen müssen. Statt einer parabolischen Skalierung wird jedoch eine allgemeinere $\alpha$-Skalierung verwendet. Aufgrund dieses Konstruktionsprinzips ist jedem $\alpha$-Molekül eine bestimmte Skalierung, eine bestimmte Orientierung und ein bestimmter Ort zugeordnet, und damit ein Punkt im sogenannten Parameterraum, welcher per Definition alle möglichen Tripel solcher Parameter umfasst.

Ein zentraler Baustein der Theorie der $\alpha$-Moleküle ist die Tatsache, dass dieser Parameterraum mit einem Distanzbegriff ausgestattet werden kann, so dass ein großer Abstand zwischen den Parametern einer kleinen Kreuzkorrelation entsprechender $\alpha$-Moleküle entspricht. Wie wir zeigen können, induziert diese auch Indexabstand genannte Distanz sogar eine quasi-metrische Struktur auf dem Parameterraum. Auf ihrer Grundlage kann bewiesen werden, dass $\alpha$-Molekülsysteme fast orthogonal zueinander stehen, wenn gewisse Konsistenzbedingungen erfüllt sind.

Dieses Resultat wiederum führt zu einem anderen Stützpfeiler der Theorie, dem sogenannten Transferprinzip, das besagt, dass $\alpha$-Molekülframes ein gleichartiges Approximationsverhalten haben, falls ihre Ordnung genügend groß ist und gewisse Konsistenzbedingungen erfüllt sind. Damit wird ein Transfer von Approximationsresultaten innerhalb des Framework ermöglicht und damit eine systematische Untersuchung sparser Approximationseigenschaften von $\alpha$-Molekülen. Da dabei auch die Frameeigenschaft der Systeme eine Rolle spielt, beweisen wir zudem ein Daubechies-artiges Framekriterium, das frühere Kriterien für Shearlets und Wavelets verallgemeinert.

Als Anwendung des Transferprinzips interessieren wir uns für das Approximationsverhalten von $\alpha$-skalierten Systemen im Falle cartoon-artiger Daten. Als konkretes Datenmodell verwenden wir dabei $C^\beta$-Cartoons, also Funktionen welche mit Ausnahme von $C^\beta$-Unstetigkeitskurven $C^\beta$-glatt sind. Es ist bekannt, dass für solche Daten die maximal erreichbare $N$-Term Approximationsrate von der Ordnung $N^{-\beta}$ ist. Desweiteren ist bekannt, dass $C^2$-Cartoons von parabolisch skalierten Systemen, wie zum Beispiel den klassischen Curvelets und Shearlets, mit einer Rate der Ordnung $N^{-2}$ quasi-optimal approximiert werden können.

Dieses Resultat wird in dieser Arbeit auf allgemeinere $\alpha$-skalierte Systeme erweitert. Dafür untersuchen wir zuerst einen Parsevalframe aus $\alpha$-Curvelets, der als prototypisches Referenzsystem fungiert. Als negatives Resultat zeigen wir, dass eine Cartoonapproximationsrate besser als $N^{-1/(1-\alpha)}$ von diesem System nicht erreicht werden kann. Die durch einfaches Thresholding der Curveletkoeffizienten erreichbare Rate ist sogar durch $N^{-1/\max\{\alpha,1-\alpha\}}$ begrenzt. Mit $\alpha$-Curvelets ist eine optimale Approximation von $C^\beta$-Cartoons also nicht möglich, wenn $\beta > 2$ gilt. Demgegenüber steht das positive Resultat, dass für die Wahl $\alpha = \beta^{-1}$ im Bereich $\beta \in (1, 2]$ quasi-optimale Approximation mit einer Rate der Ordnung $N^{-\beta}$ von $\alpha$-Curvelets erreicht wird. Über das Transferprinzip können wir schließlich schlussfolgern, dass diese für $\alpha$-Curvelets erzielten Ergebnisse auch für eine größere Klasse von $\alpha$-Molekülframes Gültigkeit besitzen.

Als weitere Anwendung verwenden wir das Konzept der $\alpha$-Moleküle in der Theorie der Funktionenräume, wo es eine einheitliche Behandlung von Curvelet- und Shearleträumen ermöglicht. Dazu führen wir mit Hilfe einer kontinuierlichen $\alpha$-Molekültransformation Besov-artige Coorbiträume ein, die von gewissen gemischt-normierten Lebesgueräumen auf dem Transformationsbereich erzeugt werden. Ein Hauptresultat, das als eine kontinuierliche Variante des Transferprinzips gedeutet werden kann, zeigt, dass diese Coorbiträume übereinstimmen, falls die Molekülordnung ausreichend hoch ist. Aus allgemeinen Prinzipien der Coorbittheorie erhalten wir zudem diskrete Charakterisierungen für diese Räume. Insbesondere können wir sie so mit bereits bekannten Curvelet- und Shearleträumen identifizieren.

Am Ende der Arbeit wenden wir uns noch einer Erweiterung der Theorie auf höhere Dimensionen zu. Dabei beschränken wir uns auf einige ausgewählte Aspekte, insbesondere werden die Definition der Indexdistanz und das Transferprinzip verallgemeinert. Als Anwendung untersuchen wir die Approximation von Video-Daten, welche als $3D$-cartoon-artige Funktionen modelliert werden können.

# Abstract


The theory of wavelets constitutes an integral part of modern harmonic analysis with many theoretical and practical applications. In engineering for example, wavelets are nowadays a popular tool for the efficient representation and approximation of functions. Much of their success thereby relies on the fact that they are more suited to represent smooth signals with singularities than traditional Fourier systems. In fact, for smooth signals with point singularities wavelet systems perform quasi-optimally with respect to sparse approximation purposes. This makes them particularly useful for the approximation of 1-dimensional data.

When approximating multivariate data, however, wavelets only show a suboptimal performance if anisotropic features are involved. The reason for this is that wavelets are inherently isotropic objects and thus not optimally suited for this task. Since in practice such anisotropic structures are very common, think of images with edges for example, over the recent years much effort has been invested to deal with this shortcoming. In particular, this led to the invention of various novel so-called directional representation systems, some of the most well-known of which are ridgelets, curvelets, and shearlets, to name just a few.

Such directional systems can conveniently be categorized according to the type of scaling involved in their construction. Wavelets for example are isotropically scaled, the scaling of ridgelets is purely directional, and the construction of the classic curvelets and shearlets is based on parabolic scaling. A great variety of different scalings is covered by the concept of $\alpha$-scaling, where a parameter $\alpha \in [0,1]$ is used to interpolate between the isotropic case and the purely directional case. The former systems, for example, are special instances of $\alpha$-scaled systems corresponding to the parameters $\alpha = 1$, $\alpha = 0$, and $\alpha = \frac{1}{2}$.

The main endeavour of this thesis is to develop a common framework for such $\alpha$-scaled representation systems. The basic notion are so-called $\alpha$-molecules which generalize the earlier concept of parabolic molecules. Those were introduced to enable a unified treatment of parabolically scaled systems. By definition, they are obtained via parabolic dilations, rotations, and translations from a set of generating functions, whereby the generators are allowed to vary as long as they obey a certain time-frequency localization. This concept of variable generators explains the terminology 'molecules'. Together with the utilization of so-called parametrizations to enable generic indexing, it provides the flexibility to cast different parabolically scaled systems as instances of one unifying construction principle.

Indeed, the framework of parabolic molecules is general enough to unite rotation-based and shear-based constructions such as the classic curvelets and the classic shearlets under one common roof. Recently, this framework has been further generalized to also include continuous systems. The limitation to parabolic scaling however still excludes systems like ridgelets and wavelets, as well as hybrid constructions such as $\alpha$-curvelets and $\alpha$-shearlets. This is the motivation behind the generalization to $\alpha$-molecules.

Like parabolic molecules, systems of $\alpha$-molecules consist of dilated, rotated, and translated versions of a set of generators which are merely required to fulfill a common time-frequency localization. However, instead of parabolic scaling, more general $\alpha$-scaling is used. Due to this construction, each $\alpha$-molecule is associated with a certain scale, a certain

location, and a certain orientation, and thus determines a point in the parameter space, which is defined as the space comprising all possible triples of such parameters.

A central building block of the theory of $\alpha$-molecules is the observation that this parameter space can be equipped with a notion of distance such that a high distance between indices corresponds to a low cross-correlation of the respective $\alpha$-molecules. This so-called index distance even induces a quasi-metric structure on the parameter space. Based on this distance, it can be proven that two systems of $\alpha$-molecules are almost orthogonal, provided that certain consistency and time-frequency localization conditions are satisfied.

This, in turn, leads to another central result of the theory, the so-called transfer principle, which states that any two frames of $\alpha$-molecules, which are consistent in a certain sense and have sufficiently high order, exhibit the same approximation behavior. It enables the transfer of approximation results within the framework and thus provides a systematic way to prove results on sparse approximation for certain model data. Thereby, also the frame property of the systems comes into play, wherefore we prove a Daubechies-type frame criterion for $\alpha$-molecules generalizing earlier criteria for shearlets and wavelets.

As an application of the transfer principle, we explore the approximation performance of $\alpha$-scaled representation systems with respect to cartoon-like data. More concretely, as data classes we consider $C^\beta$-cartoons which are $C^\beta$-smooth functions apart from $C^\beta$-discontinuity curves. It is known that the best $N$-term approximation rate achievable for such classes is of order $N^{-\beta}$. It is further known that for $C^2$-cartoons parabolically scaled systems such as the classic curvelets and shearlets achieve a quasi-optimal rate of order $N^{-2}$.

In this thesis, we extend this result to more general $\alpha$-scaled systems. For this, we first analyze the approximation properties of a prototypical anchor system, where we choose a discrete Parseval frame of $\alpha$-curvelets. As a negative result, we will find that a cartoon approximation rate exceeding $N^{-1/(1-\alpha)}$ is not possible with this system. The maximal rate obtainable by simply thresholding the curvelet coefficients is even limited to $N^{-1/\max\{\alpha,1-\alpha\}}$. Consequently, an optimal approximation of $C^\beta$-cartoons cannot be achieved if $\beta > 2$. On the positive side, however, we will see that in the range $\beta \in (1,2]$ and for the choice $\alpha = \beta^{-1}$, which in particular includes the parabolic case, the anchor frame of $\alpha$-curvelets indeed provides quasi-optimal approximation with a rate of order $N^{-\beta}$. Via the transfer principle, we finally conclude that these findings for $\alpha$-curvelets apply to a larger class of $\alpha$-molecule frames.

As another application of the concept of $\alpha$-molecules, we use it in the theory of function spaces for a unified treatment of curvelet and shearlet smoothness spaces. To this end, we introduce a continuous $\alpha$-molecule transform and associated Besov-type coorbit spaces corresponding to certain mixed-norm Lebesgue spaces on the transform domain. A main result, which can be interpreted as another manifestation of the transfer principle, shows that these $\alpha$-molecule coorbit spaces coincide if the order of the $\alpha$-molecules is sufficiently high. Moreover, the abstract machinery of coorbit theory yields discrete characterizations which allow to identify them with known scales of curvelet and shearlet smoothness spaces.

At the end of the thesis, we turn to an extension of the theory to higher dimensions. Thereby we focus on some main aspects, in particular the index distance and the transfer principle are generalized. As an application of the extension, we investigate the approximation of video data, which can be modelled as 3D-cartoon-like functions.

# Acknowledgements

First of all, I would like to thank my family for their great support during my promotional studies. Without you the writing of this thesis would not have been possible. Especially, I would like to thank Steffi for her perpetual backing and patient trust and my mother Marlis for her invaluable help in solving the NP-hard problem of organizing the childcare for our son Joel. At this point, also a big kiss to you, Joel, for your ability to always cheer me up with your deep-felt joy of life and laughter.

Concerning the research and working environment, my first thanks go to my supervisor Gitta Kutyniok for enabling this thesis and suggesting the topic. Being part of your team, the Applied Functional Analysis Group, has been a great experience and inspiration. From your knowledge, I could learn a lot, and it was always very helpful to get your valuable advice. I am also grateful for the several opportunities to participate in international conferences, which you encouraged. Here I would also like to mention the Berlin Mathematical School for their generous travel support.

I would further like to thank all members of our working group. Your were great colleagues and friends, and I enjoyed the time with you very much. Namely, I would like to thank Jan Vybíral, Friedrich Philipp, Martin Genzel, Sandra Keiper, Axel Flinth, Maximilian März, Mones Raslan, Rafael Reisenhofer, Emily King, Victoria Paternostro, Sadegh Jokar, Jan Macdonald, Stephan Wäldchen, Maximilian Leitheiser, Hector Andrade, Ali Hashemi, Alexander Paprotny, Wang-Q Lim, Mijail Guillemard, Philipp Petersen, Jackie Ma, Felix Voigtlaender, Tiep Dovan, Yizhi Sun, Katharina Eller, Ansgar Freyer, Ingo Gühring, Annika Preuß, Anja Peter, and Kerstin Eckstein.

In particular, I would like to thank my long-time office mates Anton Kolleck and Irena Bojarovska, who made the office feel like a second home. With you the work was much more fun and I will always fondly recall our time together. I also address a special thanks to Anja Hedrich for her great help and oversight in organizational and administrational issues. I always enjoyed coming to your office and having a nice conversation with you.

Finally, I would like to thank my co-authors for their fruitful collaboration and contributions to the theory. It was a pleasure to work with you. Apart from those already mentioned, I thereby thank Philipp Grohs for his cooperation in the theory of $\alpha$-molecules as well as Tino Ullrich and Henning Kempka concerning our joint work in coorbit theory and the theory of function spaces.

# Contents

# Chapter 1

# Introduction

Due to the great progress in sensor, computer, and network technology, one is nowadays able to acquire, collect, store, and process more data than ever before. In many areas of science and engineering the efficient handling of data and the question of how to extract useful information from the acquired data have thus become central topics of major importance.

In principle, a larger data pool offers the prospect of capturing more relevant information leading for example to a better understanding of observed phenomena or an improved modelling of underlying processes. The collection of large amounts of data thus promises a great potential for applications. However, in order to realize this potential, the ability to adequately process the acquired data is essential. Over the recent decades, research in this direction has therefore attracted much attention.

One area of mathematics which has greatly benefited from this development is the area of applied harmonic analysis. Rooted in classical Fourier analysis, this field provides many useful tools for the analysis and the processing of signals. In particular, its great variety of different representation systems is a great resource.

## 1.1 Multiscale Analysis

Historically, the development of applied harmonic analysis and in particular the subfield of multiscale analysis was triggered by the invention of the classic Fourier transform and related Fourier systems. Those enable a decomposition of a signal into plane wave functions and thus allow to represent a function in terms of its frequency information (see e.g. [51]). From a modern viewpoint, this can already be considered as a multiscale approach since information about higher frequencies can be interpreted as belonging to a higher scale.

A disadvantage of the Fourier transform is the fact that it only provides global information on the frequencies occurring in a signal. In order to enable a more localized query of frequency information, two other classic systems of applied harmonic analysis were developed, namely Gabor systems (see e.g. [56, 20]) and wavelet systems (see e.g. [31, 97, 114]).

Whereas Gabor systems use a fixed size window for the localization, wavelets use dilations across different scales. As a consequence, the spatial resolution of Gabor systems remains fixed. Wavelets on the other hand have the ability to zoom in on points with rising scale, at the cost of a deteriorating frequency resolution.

Both systems have had a tremendous impact on the further development of applied harmonic analysis and are still active areas of research. Due to their distinct characteristics, Gabor systems are more inclined for the use as a tool in applications where frequencies play the primary role, as for example in audio analysis, whereas wavelets have had great success in imaging science or the field of PDEs. Our focus will subsequently be on the wavelet side, mainly motivated by applications in imaging science.

### 1.1.1 Wavelets

Wavelet systems are nowadays one of the most widely used systems in applied harmonic analysis. Some real-world applications are for example the task of image compression (e.g. JPEG2000 [21]) or the restoration of corrupted image data [7]. In the field of PDEs they play a central role in solving elliptic equations [22].

The construction of a system of wavelets $\{\psi_\lambda\}_{\lambda\in\Lambda}$ in $L_2(\mathbb{R}^2)$ (see e.g. [31, 97, 114]) is based on isotropic dilations and translations of a set of generating functions $\{g_e \in L_2(\mathbb{R}^2)\}_{e\in E}$, where $E$ is some finite index set. With the isotropic scaling matrix

$$A_{1,t} := \begin{pmatrix} t & 0 \\ 0 & t \end{pmatrix}, \qquad t > 0,$$

every wavelet $\psi_\lambda \in L_2(\mathbb{R}^2)$ can be written in the form

$$\psi_\lambda = t_\lambda g_{e_\lambda}(A_{1,t_\lambda} \cdot - x_\lambda)$$

with associated parameters $x_\lambda \in \mathbb{R}^2$, $t_\lambda \in \mathbb{R}_+$, and $e_\lambda \in E$. Thereby, the prefactor $t_\lambda$ merely serves as an $L_2$-normalization constant.

By carefully choosing the generators and the parameters, usually cast in the form of appropriate admissibility and feasibility conditions, the resulting systems constitute frames or even orthonormal bases. Depending on the desired application, it is further possible to realize additional properties such as for example smoothness or compact support conditions.

A primary application of wavelets is the utilization as dictionaries for the representation and approximation of functions. In fact, their great success – besides the elegant construction principle and available fast numerical implementations – rests upon their ability to provide efficient multiscale representations for data that is subject to certain smoothness assumptions.

For example, there exist wavelet frames in $L_2(\mathbb{R}^2)$ with a quasi-optimal performance concerning the sparse approximation of functions that are smooth apart from a finite number of point singularities. In concrete terms, this means that there exist wavelet-based approximation schemes that deliver for each such signal $f \in L_2(\mathbb{R}^2)$ a sequence of $N$-term approximants $(f_N)_{N\in\mathbb{N}}$ such that the order of the decay of the $L_2$-approximation error $\|f - f_N|L_2\|$ is quasi-optimal, in an asymptotic sense. Remarkably, these $N$-term approximants can even be obtained by a simple nonadaptive thresholding scheme of the wavelet coefficients.

### 1.1.2 Cartoon-like Functions

General image data usually do not fulfill as rigid smoothness conditions as assumed in the previous example. Let us subsequently consider the continuum setting, where an image is commonly represented as a function in $L_2(\mathbb{R}^2)$ with compact support and values containing pixel information for the respective positions. Using such a representation, every edge in the image corresponds to a curvilinear discontinuity in the data. In contrast to point singularities, the approximation performance of wavelets with respect to such line singularities is not quasi-optimal any more. The isotropy of their scaling prohibits an optimal resolution, an observation which motivated the search for more efficient ways to approximate image data.

For such an endeavour it is helpful to, beforehand, precisely specify the type of data under consideration in the form of an appropriate model. With the desire to specifically model the occurrence of edges in an image, the concept of cartoon-like functions emerged. These are piecewise smooth functions featuring discontinuities along lower-dimensional manifolds. Based on such functions, different model classes for image data have been defined, typically characterized by the regularity of the smooth regions and the separating edges. As examples, let us mention the classic $C^2$-cartoons [38, 15] featuring $C^2$-regularity of the regions and the discontinuity curves, or the horizon classes considered in [35, 18, 87].

### 1.1.3 Cartoon Approximation

With the model of cartoon-like functions at hand, the question of efficient image approximation can be formulated as the task of sparsely approximating cartoon-like functions $f \in L_2(\mathbb{R}^2)$. The aim are approximation schemes with a best possible speed of convergence of the $N$-term approximants $f_N$ quantified by the asymptotic decay of the $L_2$-approximation error $\|f - f_N|L_2\|$.

The achievable approximation rate thereby depends on the regularity of the considered cartoons. Typically, this regularity is determined by the smoothness of both the edge curves and the regions in between. It was shown in [87, 86] that $C^\beta$-regularity with $\beta > 0$ allows for an asymptotic rate of order $N^{-\beta}$. By information theoretic arguments, it is further known that this rate cannot be surpassed [38], at least in a class-wise sense. Hence, the rate $N^{-\beta}$ provides an optimality benchmark for the approximation of $C^\beta$-cartoons. Interestingly, this benchmark remains the same for the subclass of binary cartoons, where the regions are assumed to be constant, and it also does not change if one restricts to $C^\beta$-smooth functions without any edges.

After the realization that wavelet-based approximation methods only provide a suboptimal performance for cartoon-like functions, a great amount of energy was devoted to the effort of constructing dictionaries better-suited for this task. Thereby, the developed methods can be divided into two categories: adaptive and nonadaptive methods.

Adaptive methods are inherently more flexible than nonadaptive methods and have the advantage of being more adjustable to the given data. On the downside, their higher flexibility typically comes at the cost of an increased computational complexity.

Some prominent examples of adaptive methods for cartoon approximation are based on wedgelet dictionaries [35] and their higher-order relatives, so-called surflets [19, 18]. Those have been shown to reach an optimal rate of order $N^{-\beta}$ for binary cartoons with $C^\beta$ regularity [16, 17]. Other notable dictionaries used for adaptive approximation include beamlets [39], platelets [113], and derivatives of wedgelets such as multiwedgelets [91] or smoothlets [90]. More recently, new adaptive schemes have emerged that use bases, e.g., bandelets [87], grouplets [98], and tetrolets [77]. For bandelets, the quasi-optimal approximation of general $C^\beta$-cartoons has been proved in [86], showing that the benchmark rate of order $N^{-\beta}$ is achievable, at least when resorting to adaptive approximation schemes.

As already mentioned above, for images that are smooth apart from point singularities, wavelets can reach a quasi-optimal approximation rate by a nonadaptive scheme, namely by a simple thresholding of the frame coefficients.

This raises the question if there also exist nonadaptive approximation methods performing quasi-optimally for certain cartoon classes, based on the thresholding of frame coefficients for example. Since, from an algorithmic perspective, such nonadaptive methods tend to be much simpler than adaptive schemes, they promise advantages for the implementation and lower computational cost. And indeed, the discovery of ridgelets and curvelets by Candès and Donoho showed that there exist frames with quasi-optimal approximation performance for certain cartoon classes.

Triggered by the invention of these first so-called directional representation systems, many novel constructions were introduced in the period that followed.

## 1.2 Directional Representation Systems

The key idea for the development of directional representation systems is to modify the original wavelet construction by incorporating some form of anisotropic scaling. Depending on the utilized type of scaling, this approach leads to many different systems. In the following we present some of the most prominent examples, but by no means this shall be a complete overview.

### 1.2.1 Ridgelets

Let us start with ridgelet systems which have been shown to yield quasi-optimal approximation [12, 64, 63] for cartoon-like functions if the edges of the cartoons are straight. Thereby the term 'ridgelet' is used for different types of constructions in the literature.

Originally, it was introduced by Candès [8] in 1998 to refer to systems consisting of translated, rotated, and dilated versions of some underlying ridge function whose profile is a univariate wavelet. Nowadays, these kind of ridgelets are called 'pure ridgelets'. They have been shown to provide quasi-optimal approximation for functions with straight line singularities in [12].

Since pure ridgelets are not square-integrable, the concept was slightly modified by Donoho to obtain frames or even bases for $L_2(\mathbb{R}^2)$. In [36] he constructed an orthonormal basis by allowing the ridgelets a slow decay along the ridge. These so-called 'orthonormal ridgelets' have similar properties as the original pure ridgelets. In particular, they share the same quasi-optimal approximation properties with respect to straight line singularities. The close relationship between orthonormal and pure ridgelets has been analyzed in [37]. A good introductory survey on the subject is given in [9].

Another ridgelet construction which coincides with the concept of '0-curvelets' is due to Grohs [57]. It is a special case of the $\alpha$-curvelet construction presented in Subsection 3.2.3 which for $\alpha = 0$ yields purely directionally scaled systems. In essence, those are obtained by performing rotations, translations, and directional scaling on a generator $g \in L_2(\mathbb{R}^2)$ with corresponding scaling matrix

$$A_{0,t} := \begin{pmatrix} t & 0 \\ 0 & 1 \end{pmatrix}, \qquad t > 0.$$

In [57, 60] tight ridgelet frames of this type were constructed that also provably provide quasi-optimal approximation of data with straight line singularities [63, 64].

For general cartoons with curved edges, however, neither of the above ridgelet systems provide a quasi-optimal performance.

### 1.2.2 Curvelets

An important milestone concerning the approximation of cartoon-like functions with curved edges was the introduction of curvelets by Candès and Donoho [14, 15]. They were introduced in 1999 representing the first frame to reach the optimal approximation order of $N^{-2}$ for general $C^2$-cartoons [14]. In 2002, a modification of the original system, the so-called second generation of curvelets [15], was introduced by the same authors. It is closely related to the frame of $\frac{1}{2}$-curvelets presented in Subsection 3.2.3 and features the same quasi-optimal approximation properties as the first generation.

The crucial ingredient in both curvelet constructions, first and second generation, is the use of parabolic scaling described by a matrix of the form

$$A_{\frac{1}{2},t} := \begin{pmatrix} t & 0 \\ 0 & \sqrt{t} \end{pmatrix}, \qquad t > 0. \tag{1.1}$$

This type of scaling can be considered as a compromise between directional scaling as used for ridgelets and isotropic scaling as used for wavelets. As the following heuristic shows, it is specifically adapted to the resolution of $C^2$-discontinuity curves.

Locally, at each point $p$ of the discontinuity, such a curve can be parametrized by $(E(x_2), x_2)$ with $E(0) = 0 = E'(0)$ using a Cartesian coordinate system $(x_1, x_2) \in \mathbb{R}^2$ which is centered at $p$ and whose $x_2$-axis coincides with the tangent. A Taylor expansion of $E$ then yields approximately $E(x_2) \approx \frac{1}{2}E''(0)x_2^2$ for small $x_2$ showing that parabolically scaled functions can optimally align with the discontinuity curve since the size of their essential support satisfies the relation *'width $\approx$ length$^2$'*.

It should be mentioned that in the actual construction of the second generation curvelets the translations and rotations are applied to a set of generators related to each other by a parabolic scaling law realised not by (1.1) but by dilations with respect to polar coordinates. This deviation from a strict affine construction allows for a simple realization of the Parseval frame property. Very similarly, as a special case of a more general $\alpha$-curvelet construction, the $\frac{1}{2}$-curvelets from Subsection 3.2.3 are obtained.

Meanwhile, many different variants of curvelet systems are available, among those even curvelet-like systems with compact support [99]. They cover a wide range of applications, for example in the field of image and seismic processing [93, 95, 34, 92], as PDE solvers [106], or in the study of turbulent flows [94]. A more thorough overview is provided in [96].

### 1.2.3 Shearlets

After the introduction of curvelets, many other systems based on parabolic scaling were developed. As examples, let us mention contourlets [33] by Do and Vetterli and shearlets going back to Guo, Kutyniok, Labate, Lim, and Weiss [81, 66]. One motivation behind those novel constructions was the desire to have systems with similar properties as curvelets but better suited for digital implementation.

The first shearlet construction was presented in 2005 by Kutyniok, Labate, Lim, and Weiss in [81]. It was an affine system obtained from a single band-limited generator using parabolic scaling, shearings, and translations. The novel ingredient and main difference to the construction of curvelets was that shearings, given by the matrices

$$S_\upsilon = \begin{pmatrix} 1 & 0 \\ \upsilon & 1 \end{pmatrix} \quad \text{and} \quad S_\upsilon^T = \begin{pmatrix} 1 & \upsilon \\ 0 & 1 \end{pmatrix}, \quad \upsilon \in \mathbb{R}, \tag{1.2}$$

and not rotations were used for the change of direction. This modification bears advantages in a discrete setting, since shearings leave the digital grid invariant, and allows for a unified treatment of the continuum and digital realm.

A drawback of the use of shearings is that those have an inherent bias towards one distinguished coordinate direction. To avoid large shear parameters and thus enable an unbiased treatment of all coordinate directions, the original shearlet construction was therefore later modified and so-called cone-adapted shearlet systems were introduced. Those have several generators with different orientations corresponding to different cones of the frequency domain. The first such construction was presented by Guo, Kutyniok, and Labate in [66]. For more details on this topic we refer to Section 3.3.

Following the initial constructions, also more sophisticated shearlet systems were developed, such as for example the cone-adapted Parseval frame of well-localized band-limited shearlets by Guo and Labate [70, 67] or systems of compactly supported shearlets by Kittipoom, Kutyniok, and Lim [76]. Like curvelets, shearlet systems provide quasi-optimal approximation for $C^2$-cartoons. For the cone-adapted band-limited shearlets this was established in [67], for those with compact support in [78].

It should be noted that, as for curvelets, many actual constructions of shearlet systems are not entirely faithful to the original idea of applying shears and parabolic scalings, using matrices (1.1) and (1.2), and translations to a finite set of generators. An example is the above mentioned cone-adapted shearlet system by Guo and Labate [70], where certain 'boundary' elements, corresponding to the boundary of the frequency cones, need to be modified to obtain good spatial localization.

Nowadays, shearlets are widely used directional representation systems with applications ranging from imaging science [40], simulations of inverse scattering problems [84] to solvers for transport equations [29]. More information can be found in the book [79].

### 1.2.4 $\alpha$-Scaling

Comparing the approximation properties of wavelets, curvelets, shearlets, and ridgelets reveals a distinct behavior with respect to their ability to resolve edges. Ridgelets are optimally suited to resolve straight edges, curvelets and shearlets are optimal for $C^2$ line singularities, and wavelets perform optimal with respect to point singularities. The origin of this characteristic behavior lies in the different scaling laws underlying the respective constructions: Isotropic scaling for wavelets, parabolic scaling for curvelets and shearlets, and directional scaling for ridgelets.

Introducing a parameter $\alpha \in \mathbb{R}$ and associated $\alpha$-scaling matrices

$$A_{\alpha,t} := \begin{pmatrix} t & 0 \\ 0 & t^{\alpha} \end{pmatrix}, \qquad t > 0, \tag{1.3}$$

it is possible to interpolate between these different types of scaling. In particular, one can construct corresponding $\alpha$-scaled representation systems, for instance $\alpha$-curvelets by incorporating $\alpha$-scaling in the classic curvelet construction. The scale of tight frames obtained in [60] for the range $\alpha \in [0, 1]$ constitutes a family of systems which encompass ridgelets (in the sense of [57]) for $\alpha = 0$, the classic curvelets for $\alpha = \frac{1}{2}$, and wavelets for $\alpha = 1$. Similarly, $\alpha$-shearlet systems can be defined by modifying the original parabolic shearlet constructions. They have been examined for example in [73, 83] (for the range $\alpha \in [\frac{1}{2}, 1)$).

A natural question concerning such $\alpha$-scaled representation systems is how their approximation properties are affected by the choice of the parameter $\alpha$. With regard to cartoon approximation, this question has been pursued in [60, 102] for $\alpha$-curvelet frames and in [73, 83] for $\alpha$-shearlet frames. In [60, 73, 83] it is shown that, if $\alpha \in [\frac{1}{2}, 1)$ and $\beta = \alpha^{-1}$, simple thresholding of the coefficients yields $N$-term approximations with an optimal convergence rate of order $N^{-\beta}$ for $C^\beta$-cartoons. The findings of [102] further extend this result. There it is shown that the best possible $N$-term approximation rate achievable for $C^\beta$-cartoons by $\alpha$-curvelets with $\alpha \in [0,1)$ is limited to at most $N^{-\frac{1}{1-\alpha}}$, independent of the smoothness $\beta > 0$. Moreover, if a simple thresholding scheme is used the achievable rate cannot even exceed $N^{-\frac{1}{\max\{\alpha,1-\alpha\}}}$.

These results show that for $C^\beta$-cartoons with $\beta \geq 2$ the classic parabolically scaled curvelets provide the best possible approximation performance among all $\alpha$-curvelet systems, at least when restricting to simple thresholding schemes, with an approximation rate of order $N^{-2}$. This confirms the special role of parabolic scaling for cartoon approximation. On the other hand, it becomes clear that the classic curvelets do not take advantage of cartoon regularity higher than $C^2$ since if $\beta > 2$ the obtainable approximation rate remains below the optimality benchmark of $N^{-\beta}$. For different choices of $\alpha$ the rate even deteriorates as $\alpha$ tends to 1 or 0. Consequently, $\alpha$-curvelets cannot provide optimal approximation for general $C^\beta$-cartoons if $\beta > 2$. In fact, up to now, no frame construction is known where a nonadaptive approximation scheme yields rates better than $N^{-2}$.

In [102] also the approximation of cartoons featuring only straight edges is considered. It is shown that by a simple thresholding scheme $\alpha$-curvelets can reach approximation rates of order $N^{-\min\{\alpha^{-1},\beta\}}$. Hence, here a smaller $\alpha$ is beneficial and even ensures quasi-optimal approximation if $\alpha \in [0, \beta^{-1}]$. This finding generalizes earlier results for ridgelets [63, 64].

## 1.3 A Common Framework

The directional systems described above are all constructed using the same idea: take a set of generators and then perform scalings with some degree of anisotropy, changes of direction using for example rotations or shearings, and finally translations. In addition, in order to obtain systems with desirable properties, usually some regularity conditions on the generators are posed. Having this in mind, it seems possible to regard all such systems as certain instances of a common more general concept.

First developments in this direction were the concepts of curvelet molecules [13] and shearlet molecules [68], conceived as a means to unify the analysis of curvelet-like and shearlet-like constructions, respectively. However, those concepts do not bridge the gap between rotation-based and shear-based constructions and are thus not able to unify those under one common roof. This was first achieved by the concept of parabolic molecules [62] using the idea of variable generators and parametrizations.

### 1.3.1 Parabolic Molecules

The concept of parabolic molecules was introduced in 2011 by Grohs and Kutyniok [62]. It has the ability to unify various parabolically scaled systems under one common roof. In particular, it allows to derive the classic curvelets and shearlets as special instances of the same general construction principle, although these specific constructions are rather different. Recall that for curvelets the scaling is done by a dilation with respect to polar coordinates and the orientation is enforced by rotations, whereas shearlets are based on affine scaling and the directionality is generated by the action of shear matrices.

The basic construction principle of a system of parabolic molecules thereby resembles that of an ordinary affine construction. Starting from a set of generating functions, the system elements are obtained by applying parabolic dilations, rotations, and translations. The essential novelty is that the generators can be chosen freely, apart from a certain time-frequency localization, and each molecule may thus have its own individual generator. This 'variability' of the generating set is the reason for the terminology 'molecules' (see also [47], for instance). Together with the utilization of so-called parametrizations to allow a generic indexing of the system elements, it provides the flexibility to cast rotation- and shear-based systems as products of the same underlying construction process. Moreover, as a nice side-effect, it becomes possible to relax the vanishing moment conditions usually imposed on the generators to achieve favorable approximation properties. Rather to demand a rigid condition as in most classic constructions, it suffices to require the moments to vanish asymptotically at high scales, without changing the asymptotic approximation behavior of the system.

In essence, the concept of parabolic molecules provides a high level description of parabolically scaled representation systems based solely on the time-frequency localization of the system elements. This has the advantage that the associated theory becomes independent of the specific constructions, allowing simultaneous investigations for many different systems. In particular, the theory is well-suited for applications in approximation theory since it is foremost the time-frequency localization of a system that is responsible for its approximation properties. As an example application, the theory of parabolic molecules was used in [62] to show that the classic curvelets and shearlets feature a similar approximation behavior.

Since nowadays higher dimensional data plays an ever increasing role a first step towards a theory for higher dimensions was pursued in [44], with an extension of the parabolic molecule framework from [62] to 3D. In the recent work [75, 61] another extension in a different direction was pursued. Here the theory of parabolic molecules was generalized to also include non-discrete systems. The resulting continuous theory is well-suited for microlocal analysis with applications for example in the theory of function spaces.

### 1.3.2 $\alpha$-Molecules

As the name already suggests, the scope of parabolic molecules is limited to parabolically scaled systems. In this thesis, we will put forward a more general framework which also includes differently scaled systems such as wavelets and ridgelets, for instance. This becomes possible by the utilization of $\alpha$-scaling (1.3), which allows to realize scalings with different degrees of anisotropy controlled by the parameter $\alpha \in [0, 1]$.

The fundamental notion are systems of $\alpha$-molecules which are obtained similarly as systems of parabolic molecules. Like those, they consist of dilated, rotated, and translated versions of a set of generators which are merely required to fulfill a common time-frequency localization. However, instead of parabolic scaling, more general $\alpha$-scaling is used. For the choice $\alpha = \frac{1}{2}$ the concept coincides with that of parabolic molecules, choosing $\alpha = 0$ or $\alpha = 1$, for example, ridgelet and wavelet systems can be obtained.

The concept of $\alpha$-molecules was first introduced in [59] as an extension of the discrete theory of parabolic molecules from [62]. In [45] the framework was then further generalized to arbitrary dimensions $d \in \mathbb{N}$ with $d \geq 2$. In this thesis, we further extend it to a continuous setting comprising then in particular the notion of continuous parabolic molecules from [75]. The theory presented in this thesis thus essentially builds upon the articles [59, 45, 75]. It is intended as an abstract tool enabling a unified treatment of a variety of directional multiscale systems, applicable for instance for the analysis of their approximation properties.

Some features of the theory are listed below.

- In Section 2.2 we prove in Theorem 2.2.2 that $\alpha$-molecule systems are almost orthogonal to each other with respect to a certain distance function on their respective indices. We further show in Theorem 2.2.12 that this so-called index distance induces a quasi-metric structure on their common underlying parameter space.
- A Daubechies-type frame criterion for a particular subclass of discrete $\alpha$-molecule systems is proved in Section 2.4. As direct corollaries we deduce two concrete frame criteria for $\alpha$-curvelet molecules and $\alpha$-shearlet molecules, in Theorem 3.2.5 and Theorem 3.3.7, respectively.
- A transfer of approximation results between different $\alpha$-molecule systems is enabled by the transfer principle, Theorem 2.3.6 proved in Section 2.3. In Chapters 5 and 6 we apply this result to determine bounds and guarantees for the approximation rates achievable by $\alpha$-molecule frames for cartoon-like functions. A multi-dimensional version of the transfer principle, Theorem 7.3.2, is proved in Chapter 7.
- The consistency of the $\alpha$-curvelet and $\alpha$-shearlet parametrizations, proved in Theorem 3.4.3 and Corollary 3.4.4, gives an explanation for the similar approximation properties of curvelet-like and shearlet-like constructions.
- The theory enables a unified structural treatment of coorbit spaces associated with $\alpha$-molecule systems. This is the topic of Chapter 4. More information on $\alpha$-molecule coorbit spaces and a short recollection of coorbit theory in general is provided in the next paragraph, Section 1.4.

Other applications, not handled in this thesis, include for example the microlocal analysis of signals on a generic $\alpha$-molecule level, as conducted with parabolic molecules in the article [75]. We further remark that, apart from the analysis aspects of the framework, the $\alpha$-molecule concept also promises new design approaches for novel multiscale constructions.

## 1.4 $\alpha$-Molecule Coorbit Spaces

The theory of coorbit spaces represents a unifying approach for the abstract description and investigation of function spaces. Starting in the 1980ies, the foundation of the theory was laid mainly by Feichtinger and Gröchenig [42, 54, 55]. The underlying idea is to use an abstract transform, called the voice transform, for the characterization of functions. Given some function class $Y$ on the associated transform domain, the term coorbit thereby refers to a retract of $Y$ in some suitable reservoir of signals.

In the original formulation, the voice transform stems from an integrable irreducible representation of a locally compact group on some Hilbert space $\mathcal{H}$. The classic example of such a transform is the continuous wavelet transform which is related to the $ax+b$-group.

Associated coorbit spaces are for example the homogeneous scales of the classic Besov and Triebel-Lizorkin spaces [107, 108, 109]. They correspond to certain mixed-norm Lebesgue spaces on the wavelet domain and were identified rigorously as coorbits in Ullrich [110]. Further extensions of these spaces were investigated by Lieang et al. [88, 89].

More general wavelet-type coorbit spaces, associated with a semidirect product $G = \mathbb{R}^d \rtimes H$, where the dilation group $H$ is a suitable subgroup of $GL(\mathbb{R}^d)$, have been studied in [48, 49] and could recently be identified with certain decomposition spaces on the Fourier domain [50]. Those in particular include shearlet coorbit spaces, first studied in [26], which are associated to the classic shearlet transform and the shearlet group.

Other group-based coorbit spaces, with a voice transform different from the wavelet transform, are for example modulation spaces [56, 41] related to the Weyl-Heisenberg group and the short-time Fourier transform or Bergman spaces [42]. Furthermore, the irreducibility and integrability conditions of the considered group representations have recently been relaxed [23], allowing for instance to treat Paley-Wiener spaces and spaces related to Shannon wavelets and Schrödingerlets as coorbits.

Whereas a group structure in the background is certainly a nice property, it also limits the reach of the theory. For example, it is not possible to treat the inhomogeneous scales of Besov-Triebel-Lizorkin spaces within the classic framework. Also shearlet spaces related to the cone-adapted version of the shearlet transform [79] do not fall into the group setting.

Therefore, in the meantime, many generalizations of the original setup have been pursued. With the aim to treat functions on manifolds, Dahlke, Steidl, and Teschke [27, 28, 24] replaced the group by a homogeneous space, for example, i.e., a quotient of a group with a subgroup. A frame-based approach, not relying on an underlying group structure at all, was developed by Fornasier and Rauhut [46]. Instead of a group representation, the starting point of this generalized theory is the notion of a continuous Hilbert frame, a notion which first appeared in [1]. The voice transform is then defined as the associated analysis operator. Intriguingly, many aspects of the original theory remain valid in this more general setup. In particular, analoga of the classic discretization results hold true.

To make the frame-based theory more accessible for applications, it was later revised and extended in [4]. Another expansion was conducted in [111], where the theory was used to characterize the inhomogeneous versions of the Besov-Triebel-Lizorkin spaces as coorbits with respect to an inhomogeneous continuous wavelet transform.

Other generalizations of the original theory due to Feichtinger and Gröchenig concern the requirements imposed on the function class $Y$ on the transform domain $\mathbb{X}$. In the classic setting, the class $Y$ is required to be a Banach function space. The group-based theory was then extended in [100] to a more general quasi-Banach setting, utilizing the idea of Wiener amalgams. In particular, this extension allows for coorbit characterizations of the homogeneous Besov-Triebel-Lizorkin spaces also in the quasi-Banach range.

Combining the approach in [111] with the idea from [100] leads to a group-less formulation of coorbit theory as presented in [74] which also comprises the quasi-Banach case. This generalized version of the theory is the foundation for our subsequent definition and analysis of $\alpha$-curvelet and $\alpha$-molecule coorbit spaces in Chapter 4. These spaces are associated with a continuous $\alpha$-curvelet transform, defined as a generalization of the parabolic curvelet transform from [11], and a more general continuous $\alpha$-molecule transform, respectively, whereby both of which are not naturally related to any group structure. A group-less formulation of coorbit theory is therefore a prerequisite for their definition.

Since the continuous $\alpha$-molecule transform in particular generalizes the cone-adapted version of the continuous $\alpha$-shearlet transform, $\alpha$-molecule coorbit spaces enable a unified description of curvelet and shearlet smoothness spaces. Thereby the latter, building on the concept of decomposition spaces [6], have been defined well before the coorbit descriptions given in this thesis, see e.g. [85]. Representing an alternative approach for the definition and investigation of function spaces, decomposition spaces have a close relationship to coorbit spaces. In fact, many function spaces can be described in both ways, see e.g. [50]. In particular, $\alpha$-curvelet and $\alpha$-shearlet decomposition spaces coincide with their respective $\alpha$-molecule coorbit counterparts.

## 1.5 Outline

The thesis is organized as follows.

After the introduction in Chapter 1, we begin with the development of the general theory in Chapter 2. Here we first restrict to a bivariate setting and introduce the notion of a system of $\alpha$-molecules in $L_2(\mathbb{R}^2)$. By definition, those are distinguished by their respective orders and parametrizations, i.e., associated mappings from their index sets into a common underlying parameter space. As the theory will show, for many investigations the knowledge of these characteristic parameters is sufficient information.

The parameter space is then equipped with a quasi-metric structure induced by an $\alpha$-scaled index distance, which is closely related to the cross-correlations of $\alpha$-molecules. One of the main results, Theorem 2.2.2, states that systems of $\alpha$-molecules are almost orthogonal to each other in the sense that cross-correlations are small whenever the index distance is large. We continue with some deeper investigation of certain subclasses of $\alpha$-molecule systems. In Theorem 2.3.6 we derive a sufficient condition for discrete $\alpha$-molecule systems to be sparsity equivalent. This condition, which is solely based on the order and the parametrization of the involved systems, gives rise to the so-called transfer principle since it enables the transfer of approximation properties within the framework. Finally, in Theorem 2.4.1 at the end of Chapter 2, we prove a Daubechies-type frame criterion for a specific subclass of discrete $\alpha$-molecule systems.

Some concrete examples of $\alpha$-molecule systems in $L_2(\mathbb{R}^2)$ are presented in Chapter 3. At first we construct a continuous frame of $\alpha$-curvelets and verify its frame property and that it is indeed a system of $\alpha$-molecules. Then we turn to discrete $\alpha$-molecule systems, whereby we distinguish two important subclasses, namely $\alpha$-curvelet and $\alpha$-shearlet molecules. Those are characterized by corresponding classes of parametrizations, called $\alpha$-curvelet and $\alpha$-shearlet parametrizations. As particular instances of these classes, discrete $\alpha$-curvelet frames and cone-adapted $\alpha$-shearlets are considered. A main result of this chapter is Theorem 3.4.3, a direct consequence of which is the fact that the $\alpha$-curvelet and $\alpha$-shearlet parametrizations are consistent with each other. We further show that wavelets and ridgelets, in the sense of 0-curvelets, fit into the framework.

Chapter 4 is devoted to an application of the concept of $\alpha$-molecules in the theory of function spaces. Based on the continuous $\alpha$-curvelet frame from Chapter 3, we introduce an associated continuous $\alpha$-molecule transform and Besov-type coorbit spaces corresponding to certain mixed-norm Lebesgue spaces on the transform domain. A main result, Theorem 4.3.8, which can be interpreted as another manifestation of the transfer principle, shows that these $\alpha$-molecule coorbit spaces coincide if the order of the $\alpha$-molecules is sufficiently high. The discrete characterization of Theorem 4.3.13 further allows to identify them with known scales of curvelet and shearlet smoothness spaces. Finally, the abstract machinery of coorbit theory yields two other discretization results, Theorem 4.4.19 and Theorem 4.4.21.

In Chapters 5 and 6 we turn to another application of the theory. Here we investigate the approximation performance of $\alpha$-molecule systems for certain classes of cartoon-like functions. Whereas Chapter 5 is concerned with bounds on the achievable approximation rates, the main results being Theorem 5.4.2, Theorem 5.4.4, and Theorem 5.4.6, in Chapter 6 actual guarantees for these rates are established, in Theorem 6.0.1 and Theorem 6.0.2.

An extension of the theory to multi-dimensions $d \in \mathbb{N}\backslash\{1\}$ is conducted in the last part of the thesis, Chapter 7. Both, the notion of $\alpha$-molecules and the notion of $\alpha$-shearlet molecules are transferred to $L_2(\mathbb{R}^d)$, which requires the parameter space as well as the index distance to be adapted to $d$ dimensions. As in the bivariate case, systems of such multivariate $\alpha$-molecules are almost orthogonal to each other, which is established in Theorem 7.2.2. Consequently, also a $d$-dimensional version of the transfer principle, Theorem 7.3.2, holds true. As an application, we finally investigate the approximation performance of parabolic molecules in 3D with respect to video data, leading to Theorem 7.5.8.

## 1.6 Preliminaries: *Notation and Conventions*

For clarity, let us shortly explain the general notation used throughout the thesis. The symbols $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{C}$ have the standard meaning, i.e., $\mathbb{N}$ stands for the natural numbers, $\mathbb{N}_0$ for the natural numbers including 0, $\mathbb{Z}$ for the integers, and $\mathbb{R}$ and $\mathbb{C}$ are the real numbers and complex numbers, respectively. The strictly positive real numbers are denoted by $\mathbb{R}_+$, i.e., $\mathbb{R}_+ := (0,\infty)$, whereas $\mathbb{R}_0^+ := [0,\infty)$ stands for the ray including 0.

The complex conjugate of a number $z \in \mathbb{C}$ is denoted by $\overline{z}$. For $x, y \in \mathbb{R}$ we put $(x,y)_+ := \max\{x,y\}$ and $(x)_+ := (x,0)_+ = \max\{x,0\}$. Further, the floor and ceiling functions are defined by $\lfloor x \rfloor := \max\{n \in \mathbb{Z} : n \le x\}$ and $\lceil x \rceil := \min\{n \in \mathbb{Z} : n \ge x\}$, respectively. A useful abbreviation is also the ubiquitous 'analyst's bracket' given by $\langle x \rangle := \sqrt{1+x^2}$.

For two entities $x, y \in \mathbb{R}$, dependent on a certain set of parameters, the notation $x \lesssim y$ shall mean that there exists a constant $C > 0$ such that $x \le Cy$, uniformly in the parameters. If the converse inequality holds true, we write $x \gtrsim y$ and if both inequalities hold we shall write $x \asymp y$.

The vector space $\mathbb{R}^d$ with $d \in \mathbb{N}$ is equipped with the usual Euclidean scalar product denoted by $\langle \cdot, \cdot \rangle$. The $p$-quasi-norm in the range $0 < p \le \infty$ of a vector $x \in \mathbb{R}^d$ is denoted by $|x|_p$. In case of the Euclidean norm $|x|_2 = \sqrt{\langle x, x \rangle}$, we will usually omit the subindex. For the unit sphere $\{x \in \mathbb{R}^d : |x| = 1\}$ in $\mathbb{R}^d$ the symbol $\mathbb{S}^{d-1}$ is used. The standard unit vectors are denoted by $e_1, \ldots, e_d$, and for a vector $x \in \mathbb{R}^d$ we use the notation $[x]_i := \langle x, e_i \rangle$, $i \in \{1, \ldots, d\}$, for the $i$:th component. In Chapter 7, also the short-hand notation $|x|_{[d-1]} := |([x]_1, \ldots, [x]_{d-1}, 0)|_2$ will be useful.

Besides Cartesian coordinates, we will often use polar coordinates for the representation of a vector $x \in \mathbb{R}^2$, i.e., a pair $(r,\phi) \in [0,\infty) \times [0, 2\pi)$, where $r = |x|$ is the length of the ray from the origin $(0,0)$ to $x$ and $\phi = \phi(x)$ measures the angle from the $x_1$-axis to this ray, in a counter-clockwise sense.

The usual Lebesgue spaces on a generic measure space $(\Omega, \mu)$ are denoted by $L_p(\Omega) := L_p(\Omega, \mu)$, where $0 < p \le \infty$, and the symbol $\| \cdot | L_p \|$ is used for the associated quasi-norms.

The inner product on the Hilbert space $L_2(\Omega)$ is given by

$$\langle f, g\rangle := \int_\Omega f(x)\overline{g(x)}\, d\mu(x), \qquad f, g \in L_2(\Omega),$$

whereby the same symbol $\langle\cdot,\cdot\rangle$ is used as for the scalar product on $\mathbb{R}^d$. In case $\Omega = \mathbb{R}^d$, we further introduce the space $L_p^{\mathrm{loc}}(\mathbb{R}^d)$ consisting of all Lebesgue-measurable functions $f$ on $\mathbb{R}^d$ which satisfy $f\mathcal{X}_K \in L_p(\mathbb{R}^d)$ for every compact subset $K \subset \mathbb{R}^d$. Thereby $\mathcal{X}_K$ is the characteristic function of $K$, i.e., $\mathcal{X}_K(x) = 1$ for $x \in K$ and $\mathcal{X}_K(x) = 0$ otherwise.

For the Lebesgue sequence spaces, corresponding to a countable index set $\Lambda$, we write $\ell^p(\Lambda)$. The weak versions of these spaces are denoted by $\omega\ell^p(\Lambda)$ with associated quasi-norms $\|\cdot\|_{\omega\ell^p}$. Their precise definition is recalled in Subsection 2.3.1.

Next, let us turn to the scale $C^\beta_{\mathrm{loc}}(\Omega)$, $\beta \in [0,\infty)$, of classic smoothness spaces on some domain $\Omega \subseteq \mathbb{R}^d$. For a multi-index $m = (m_1, \dots, m_d) \in \mathbb{N}_0^d$ we first introduce the notation $\partial^m := \partial_1^{m_1} \cdots \partial_d^{m_d}$, where $\partial_i$ is the partial derivative in the $i$-th coordinate direction, $i \in \{1, \dots, d\}$. Then we can define $C^\beta_{\mathrm{loc}}(\Omega)$ as the space comprising all functions on $\Omega$ which have continuous derivatives up to order $\lfloor\beta\rfloor$ such that $\mathit{H\ddot{o}l}_K(\partial^m f, \beta - \lfloor\beta\rfloor) < \infty$ for every compact subset $K \subset \mathbb{R}^d$ and every multi-index $m \in \mathbb{N}_0^d$ with $|m|_1 = \lfloor\beta\rfloor$. Hereby,

$$\mathit{H\ddot{o}l}_K(f,\gamma) := \sup_{x,y\in K\cap\Omega} \frac{|f(x) - f(y)|}{|x-y|^\gamma}$$

is the Hölder constant of a function $f : \Omega \to \mathbb{C}$ with respect to the exponent $\gamma \in [0,1]$ and the domain $K$. We further introduce the Banach space

$$C^\beta(\Omega) := \Big\{ f \in C^{\lfloor\beta\rfloor}_{\mathrm{loc}}(\Omega) : \|f\|_{C^\beta(\Omega)} := \|f\|_{C^{\lfloor\beta\rfloor}(\Omega)} + \sum_{|m|_1 = \lfloor\beta\rfloor} \mathit{H\ddot{o}l}(\partial^m f, \beta - \lfloor\beta\rfloor) < \infty \Big\},$$

where $\|f\|_{C^{\lfloor\beta\rfloor}(\Omega)} := \sum_{|m|_1 \le \lfloor\beta\rfloor} \sup_{x\in\Omega} |\partial^m f(x)|$ and $\mathit{H\ddot{o}l}(f,\gamma) := \mathit{H\ddot{o}l}_{\mathbb{R}^d}(f,\gamma)$. For convenience, the space of continuous functions $C^0(\Omega)$ is often denoted by $C(\Omega)$, a notation also used for continuous functions on general topological spaces $\Omega$. At last, we extend the definition of $C^\beta_{\mathrm{loc}}(\Omega)$ and $C^\beta(\Omega)$ to $\beta = \infty$ and let $C^\infty_{\mathrm{loc}}(\Omega) := \bigcap_{\beta\ge 0} C^\beta_{\mathrm{loc}}(\Omega)$ and $C^\infty(\Omega) := \bigcap_{\beta \ge 0} C^\beta(\Omega)$.

All functions $f \in C^\beta(\mathbb{R}^d)$, $\beta \in [0,\infty]$, whose support $\operatorname{supp} f$ is a compact subset of $\Omega$ are collected in the space $C^\beta_0(\Omega)$ which can be considered as a subspace of $C^\beta(\Omega)$ by identifying every $f \in C^\beta_0(\Omega)$ with its restriction $f|_\Omega$. Note that the functions in $C^\beta_0(\Omega)$ necessarily vanish on the boundary $\partial\Omega$. In contrast, the notation $C^\beta_c(\Omega)$ refers to the larger space of all compactly supported functions in $C^\beta(\Omega)$. Thereby the notation $C_c(\Omega)$ is again also used for general topological spaces $\Omega$.

The Schwartz space of rapidly decreasing functions on $\mathbb{R}^d$ is denoted by $\mathcal{S}(\mathbb{R}^d)$. Let us put $x^m := x_1^{m_1} \cdots x_d^{m_d}$ for $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ and a multi-index $m = (m_1, \dots, m_d) \in \mathbb{N}_0^d$. Then we have

$$\mathcal{S}(\mathbb{R}^d) := \Big\{ f \in C^\infty(\mathbb{R}^d, \mathbb{C}) \ : \ |f|_{\kappa,\nu} < \infty \text{ for all } (\kappa,\nu) \in \mathbb{N}_0^d \times \mathbb{N}_0^d \Big\}$$

with

$$|f|_{\kappa,\nu} := \sup_{x\in\mathbb{R}^d} \left|\xi^\kappa \partial^\nu f(\xi)\right|, \quad \kappa,\nu \in \mathbb{N}_0^d. \tag{1.4}$$

Furthermore, this space is topologized by the locally convex topology induced by the collection of semi-norms in (1.4).

The Fourier transform $\mathcal{F}f$ of a function $f \in \mathcal{S}(\mathbb{R}^d)$ shall be given by

$$\mathcal{F}f(\xi) := \int_{\mathbb{R}^d} f(x) \exp(-2\pi i \langle \xi, x \rangle)\, dx,$$

for which we will often use the short-hand notation $\widehat{f} = \mathcal{F}f$. We further remark that, as usual, the transform $\mathcal{F}$ is extended to the space of tempered distributions $\mathcal{S}'(\mathbb{R}^d)$, i.e., the topological dual of $\mathcal{S}(\mathbb{R}^d)$.

We finally mention another important transform which we will encounter in Chapter 6. It is the Radon transform $\mathcal{R}f$ defined via the line integral

$$\mathcal{R}f(t, \eta) := \int_{\mathfrak{L}_{t,\eta}} f\, ds, \tag{1.5}$$

whereby $(t, \eta) \in \mathbb{R} \times (-\pi/2, \pi/2]$ and $\mathfrak{L}_{t,\eta} := \{(x_1, x_2) \in \mathbb{R}^2 : \sin(\eta)x_1 + \cos(\eta)x_2 = t\}$.

After these preliminaries, we are now ready to turn to the development of the basic theory of $\alpha$-molecules in Chapter 2.

# Chapter 2

# Bivariate $\alpha$-Molecules

In this chapter we lay the foundation for the theory of $\alpha$-molecules in $L_2(\mathbb{R}^2)$. As already explained in the introduction, $\alpha$-molecules are envisioned as a common framework for different directional multi-scale systems, encompassing in particular the classic constructions of wavelets, ridgelets, curvelets, and shearlets. They are intended as an abstract tool for the applied harmonic analyst, enabling a simultaneous treatment of such systems and thus simplifying many considerations.

The subsequent exposition is mainly based on the article [59], some additional results are presented in Section 2.2 and Section 2.4. Since we also want to investigate continuous $\alpha$-molecule systems, especially in Chapter 4, the discrete theory in [59] is further transferred to a continuous setting. The presentation is then in line with the theory of continuous parabolic molecules put forward in [75]. Technically, this transfer mainly just requires an adaption of the formulation, whereas the underlying proofs of the results essentially remain the same.

The structure of the exposition is as follows. The first section deals with the basic notions of the theory. Here the general definition of an $\alpha$-molecule system in $L_2(\mathbb{R}^2)$ of a certain order is given and the corresponding parameter space, also called the phase space, together with the concept of parametrizations is introduced.

In the next section, the parameter space is equipped with a natural quasi-metric giving rise to a notion of distance between $\alpha$-molecules in phase space. According to Theorem 2.2.2, whose proof is given at the end of the chapter, this so-called *index distance* is in correspondence with the size of the cross-correlations of the respective $\alpha$-molecules, i.e., their scalar products. This is a central result and will play a pivotal role throughout the whole theory. The remainder of the section is devoted to a thorough analysis of the induced quasi-metric structure of the phase space, a notable result being Theorem 2.2.12.

We turn to approximation theoretic considerations in the third section. Based on the index distance, a notion of consistency of parametrizations is introduced and we prove in Theorem 2.3.6 that discrete $\alpha$-molecule frames with consistent parametrizations are sparsity equivalent if their orders are sufficiently high. In terms of approximation, this means that the approximation rates can be transferred between such frames, wherefore this result is also called the *transfer principle*.

In the fourth section we proceed with a short investigation of frame properties of discrete $\alpha$-molecule systems. The main result is Theorem 2.4.1, a sufficient frame criterion of Daubechies-type applicable to a certain class of $\alpha$-molecule systems.

## 2.1 The Concept of $\alpha$-Molecules in $L_2(\mathbb{R}^2)$

Modern directional multi-scale systems such as ridgelets, curvelets, and shearlets, have evolved from classical wavelet systems whose multi-scale structure is solely based on translations and dilations. With scale and position being the only degrees of freedom of a wavelet, a suitable wavelet parameter space – for the bivariate case – is given by $\mathbb{R}^2 \times \mathbb{R}_+$.

In contrast, directional multi-scale systems possess orientation as an additional parameter. Every element not only corresponds to a certain scale and position, but also to a certain orientation. This necessitates an appropriate extension of the parameter space. Adding a new variable corresponding to orientation leads to the following definition.

Note that in contrast to [59, Def. 2.7] we use the full circle of orientations as in [45]. At this stage, this seems to be the most natural choice.

**Definition 2.1.1** (compare [59])**.** The *parameter space* $\mathbb{P}$ is defined by

$$\mathbb{P} := \mathbb{R}^2 \times \mathbb{T} \times \mathbb{R}_+, \tag{2.1}$$

where here and throughout the thesis $\mathbb{R}_+ := (0, \infty)$ and $\mathbb{T} := [0, 2\pi)$.

This parameter space will also be referred to as *phase space*. Its points $\mathbf{x} = (x, \eta, t) \in \mathbb{P}$ carry information on the *scale* $t \in \mathbb{R}_+$, the *orientation* $\eta \in \mathbb{T}$, and the *location* $x \in \mathbb{R}^2$ of the yet to be defined $\alpha$-molecules. By convention, the orientation represented by a value $\eta \in \mathbb{T}$ is expressed explicitly by the vector

$$e_\eta := (\cos(\eta), -\sin(\eta)) = R_\eta^{-1} e_1, \tag{2.2}$$

where $e_1 := (1, 0) \in \mathbb{R}^2$ is the first unit vector in $\mathbb{R}^2$ and $R_\eta$ denotes the *rotation matrix*

$$R_\eta := \begin{pmatrix} \cos(\eta) & -\sin(\eta) \\ \sin(\eta) & \cos(\eta) \end{pmatrix}, \qquad \eta \in \mathbb{R}. \tag{2.3}$$

In the sequel, the interval $\mathbb{T} = [0, 2\pi)$ will often be identified with the unit sphere $\mathbb{S}^1 \subset \mathbb{R}^2$ via the correspondence $\eta \mapsto e_\eta$.

One problem that occurs, when aiming for a common framework able to unify different directional multi-scale systems, is the fact that the index sets of the various systems usually differ from each other. However, using $\mathbb{P}$ as a common parameter space and the concept of *parametrizations*, it is possible to include systems independent of their specific indexing.

**Definition 2.1.2** ([59])**.** A *parametrization* is a pair $(\Lambda, \Phi_\Lambda)$ consisting of an index set $\Lambda$ and a mapping

$$\Phi_\Lambda : \Lambda \to \mathbb{P}, \quad \lambda \mapsto \mathbf{x}_\lambda = (x_\lambda, \eta_\lambda, t_\lambda),$$

which associates to each index $\lambda \in \Lambda$ a point $\mathbf{x}_\lambda = (x_\lambda, \eta_\lambda, t_\lambda) \in \mathbb{P}$, specifying a *scale* $t_\lambda \in \mathbb{R}_+$, an *orientation* $\eta_\lambda \in \mathbb{T}$, and a *location* $x_\lambda \in \mathbb{R}^2$.

The general construction of a system of $\alpha$-molecules shall follow the same principles used for the construction of a typical directional multi-scale system. Such a system is usually obtained from a set of generating functions by applying a scaling operation in connection with certain transformations to adjust the orientation and location of its elements.

The first question that arises is which type of scaling should be used for $\alpha$-molecules. Whereas wavelets scale isotropically, curvelets and shearlets are based on parabolic scaling, ridgelets only scale in one coordinate direction. Since the framework of $\alpha$-molecules is supposed to be general enough to comprise all these classic systems, different scaling anisotropies need to be accounted for.

A convenient way to do this is to introduce a parameter $\alpha \in [0,1]$ and associated $\alpha$-*scaling matrices*

$$A_{\alpha,t} := \begin{pmatrix} t & 0 \\ 0 & t^\alpha \end{pmatrix}, \qquad t \in \mathbb{R}_+. \tag{2.4}$$

With these matrices different degrees of anisotropy of the scaling can be realized, ranging from isotropic scaling for $\alpha = 1$ to pure directional scaling for $\alpha = 0$. The parameter $\alpha = \frac{1}{2}$ corresponds to parabolic scaling.

The next question concerns the transformations which should be used for the adjustment of the orientation and location of the $\alpha$-molecules. Since the envisioned framework is mainly a theoretical framework, rotations and translations seem to be the most natural choice. Note however, that in practice – due to numerical and computational advantages – often other means for the orientation change are used. A prominent example are shearlet systems where shearings take the place of rotations. Intruigingly, the choice of rotations in the definition of $\alpha$-molecules does not confine this concept to rotation-based constructions. In Chapter 3 we will prove for example that shearlets are still included in the framework.

Finally, we come to the main conceptual ingredient for the construction of an $\alpha$-molecule system $\{m_\lambda\}_{\lambda\in\Lambda}$. Since we want to ensure maximal flexibility, we allow the generators to change with *each* index $\lambda \in \Lambda$, i.e., we employ an associated family $\{g_\lambda\}_{\lambda\in\Lambda}$ of variable generators which are merely subject to a common time-frequency localization. This localization condition is specified by a set of *control parameters* $(L, M, N_1, N_2)$, where $L$ describes the spatial localization of the generators, $M$ their number of directional almost vanishing moments, and $N_1, N_2$ their smoothness.

It is this construction principle which explains the use of the term 'molecule'. In the theory of atomic decompositions (see e.g. [42]), 'atoms' usually refer to bounded functions with compact support and many vanishing moments. Replacing the compact support condition by some weaker decay requirement leads to the notion of a 'molecule'. Furthermore, atomic decompositions are typically obtained by transforming a set of fixed generators, resembling the construction of a system of $\alpha$-molecules. There are also notable differences however. Whereas the utilized transformations are typically obtained from an underlying group action, there is no natural group structure related to the parameter space $\mathbb{P}$.

After these explanations, we are ready for the formal definition of a system of $\alpha$-molecules $\{m_\lambda\}_{\lambda\in\Lambda}$. The definition given below corresponds to [59, Def. 2.9], with the difference that the scale variable $t \in \mathbb{R}_+$ is inverted, i.e., a small $t \in \mathbb{R}_+$ now corresponds to a high scale. By this modification, our exposition is more in line with the continuous setting which has not been considered before for $\alpha$-molecules but was already the subject of investigation for parabolic molecules [75].

As for the notation, we use the so-called *analyst's bracket* $\langle x\rangle := (1+x^2)^{\frac{1}{2}}$ defined for $x \in \mathbb{R}$. Further, the notation $a \lesssim b$ indicates that the entities $a, b$ satisfy $a \leq Cb$ for an *implicit constant* $C > 0$, independent of the intrinsic parameters.

**Definition 2.1.3** (compare [59, Def. 2.9])**.** Let $\alpha \in [0,1]$, and let $L, M, N_1, N_2 \in \mathbb{N}_0 \cup \{\infty\}$. Further, let $(\Lambda, \Phi_\Lambda)$ be a parametrization where

$$\Phi_\Lambda \;:\; \Lambda \to \mathbb{P} = \mathbb{R}^2 \times \mathbb{T} \times \mathbb{R}_+ \,, \quad \lambda \mapsto (x_\lambda, \eta_\lambda, t_\lambda).$$

A family $\{m_\lambda\}_{\lambda\in\Lambda}$ of functions contained in $L_2(\mathbb{R}^2)$ is called a *system of $\alpha$-molecules of order* $(L, M, N_1, N_2)$ *with respect to the parametrization* $(\Lambda, \Phi_\Lambda)$, if its elements can be written as

$$m_\lambda(\cdot) = t_\lambda^{-(1+\alpha)/2} g_\lambda(A_{\alpha,t_\lambda}^{-1} R_{\eta_\lambda}(\cdot - x_\lambda))$$

with *generators* $g_\lambda \in L_2(\mathbb{R}^2)$ satisfying for all $\rho \in \mathbb{N}_0^2$ with $|\rho|_1 \le L$

$$|\partial^\rho \hat{g}_\lambda(\xi)| \lesssim \min\left\{1, t_\lambda + |\xi_1| + t_\lambda^{1-\alpha}|\xi_2|\right\}^M \langle|\xi|\rangle^{-N_1} \langle\xi_2\rangle^{-N_2}. \tag{2.5}$$

The implicit constant is required to be uniform over all $\lambda \in \Lambda$ and $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$. If a control parameter equals infinity, this means that the respective quantity can be chosen arbitrarily large in (2.5).

Note that, as desired, the general building principles of a typical directional multi-scale system are reflected by Definition 2.1.3. Each molecule $m_\lambda$ is obtained from a corresponding generator $g_\lambda$ by a scaling operation and a subsequent adjustment of orientation and location. A molecule $m_\lambda$ with *phase space coordinates* $\mathbf{x}_\lambda = (x_\lambda, \eta_\lambda, t_\lambda) \in \mathbb{P}$ corresponds to the scale $t_\lambda \in \mathbb{R}_+$ and is located at the point $x_\lambda \in \mathbb{R}^2$. Its orientation, represented by $\eta_\lambda \in \mathbb{T}$, is given by the vector $e_{\eta_\lambda} = (\cos(\eta_\lambda), -\sin(\eta_\lambda)) \in \mathbb{R}^2$ as in (2.2).

The uniform time-frequency localization of the generators $g_\lambda$ is specified in (2.5). As a consequence of this condition, the frequency support of each $\alpha$-molecule $m_\lambda$ is essentially contained in a pair of opposite wedges in the frequency domain, whereby the location of these wedges is determined solely by the scale $t_\lambda$ and the orientation $\eta_\lambda$ of the respective $\alpha$-molecule.

In order to see this, we use a representation of $\hat{m}_\lambda$ in polar coordinates. Let $\xi(r,\phi) := (r\cos(\phi), r\sin(\phi))$ for $r \ge 0$ and $\phi \in \mathbb{T}$. Then the function $\hat{m}_\lambda$ can be easily computed to satisfy

$$|\hat{m}_\lambda(\xi(r,\phi))| \lesssim t_\lambda^{(1+\alpha)/2} \cdot \min\{1, t_\lambda(1+r)\}^M \cdot \langle \min\{t_\lambda^\alpha, t_\lambda\} r\rangle^{-N_1} \cdot \langle t_\lambda^\alpha r \sin(\phi + \eta_\lambda)\rangle^{-N_2}. \tag{2.6}$$

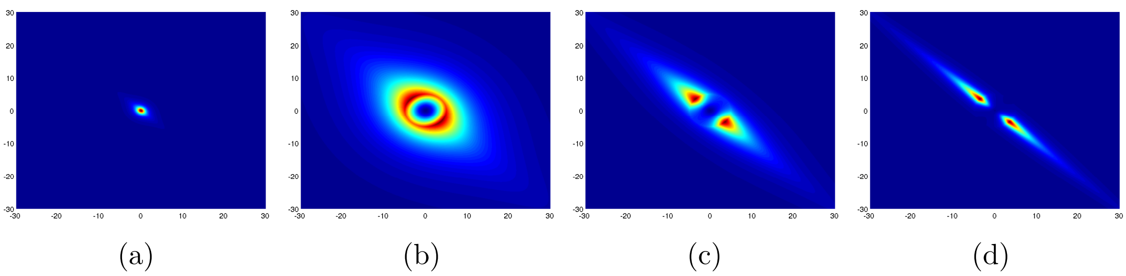


Figure 2.1: Frequency support of $\alpha$-molecules ($N_1 = 2$, $N_2 = 1$, $M = 3$, $\eta = \frac{\pi}{4}$) with (a): $t = 1$ and $\alpha$ arbitrary, (b): $t = \frac{1}{6}$ and $\alpha = 1$, (c): $t = \frac{1}{6}$ and $\alpha = \frac{1}{2}$, (d): $t = \frac{1}{6}$ and $\alpha = 0$.

As an illustration of (2.6), the essential frequency support of several $\alpha$-molecules with different phase space coordinates is depicted in Figure 2.1.

On the spatial side, the essential support of $m_\lambda$ can be thought of as being contained in the rectangle $x_\lambda + R_{\eta_\lambda}^{-1} A_{\alpha,t_\lambda} Q$ of dimensions $t_\lambda \times t_\lambda^\alpha$, where $Q := [-1,1]^2$.

Finally, we remark that the normalization factor occurring in the $\alpha$-molecule definition ensures the equality $\|m_\lambda|L_2\| = \|g_\lambda|L_2\|$ for all $\lambda \in \Lambda$. In combination with condition (2.5), we can deduce that an $\alpha$-molecule system of order $(L, M, N_1, N_2)$ is $L_2$-bounded under the mild assumption $N_1 > 1$.

**Lemma 2.1.4.** *Let $\mathfrak{M}_\alpha = \{m_\lambda\}_{\lambda\in\Lambda}$ be a system of $\alpha$-molecules of order $(L, M, N_1, N_2)$ with respect to some arbitrary parametrization $(\Lambda, \Phi_\Lambda)$. If $N_1 > 1$ its elements satisfy*

$$\sup_{\lambda\in\Lambda} \|m_\lambda|L_2\| < \infty.$$

*Proof.* Let $N_1 > 1$. Then we have uniformly for all $\lambda \in \Lambda$

$$\|m_\lambda|L_2\|^2 = \|g_\lambda|L_2\|^2 = \|\hat{g}_\lambda|L_2\|^2 \lesssim \int_{\mathbb{R}^2} (1 + |\xi|^2)^{-N_1} \, d\xi < \infty.$$

□

Finally, with a viable notion of $\alpha$-molecules at hand, we could verify the unifying qualities of this notion by providing some concrete examples. However, we postpone this investigation to Chapter 3 and instead proceed with the development of the general theory.

## 2.2 Metrization of the Parameter Space

Our next goal is to develop appropriate tools enabling the analysis of $\alpha$-molecule systems on a generic level. For instance, one might be interested in frame or approximation properties of such systems, which we will in fact investigate later in Sections 2.3 and 2.4. Also the properties of associated transforms might be of interest, which will be relevant for us in Chapter 4.

A fundamental tool for the analysis of function systems in general is given by the so-called *cross-Gramian matrices* $\mathcal{G}[\mathfrak{M}, \widetilde{\mathfrak{M}}]$ associated to any two systems $\mathfrak{M} := \{m_\lambda\}_{\lambda\in\Lambda}$ and $\widetilde{\mathfrak{M}} := \{\tilde{m}_\mu\}_{\mu\in\Delta}$ in $L_2(\mathbb{R}^2)$. Their entries are the scalar products $\langle m_\lambda, \tilde{m}_\mu\rangle$, also called the *cross-correlations*, of the individual functions from $\mathfrak{M}$ and $\widetilde{\mathfrak{M}}$. Depending on the index sets $\Lambda$ and $\Delta$, the cross-Gramian $\mathcal{G}[\mathfrak{M}, \widetilde{\mathfrak{M}}]$ is thus the possibly infinite-dimensional matrix

$$\mathcal{G}[\mathfrak{M}, \widetilde{\mathfrak{M}}] := \left\{\langle m_\lambda, \tilde{m}_\mu\rangle\right\}_{\lambda\in\Lambda, \mu\in\Delta}. \tag{2.7}$$

In case of a single system, i.e., when $\mathfrak{M} = \widetilde{\mathfrak{M}}$, the notation is simplified $\mathcal{G}[\mathfrak{M}] := \mathcal{G}[\mathfrak{M}, \mathfrak{M}]$ and we call the matrix $\mathcal{G}[\mathfrak{M}]$ just the *Gramian matrix* of the system $\mathfrak{M}$.

The cross-Gramian matrix (2.7) of two function systems contains essential information about the mutual relationship of the system elements. A careful analysis enables a comparison of the different systems and can also reveal many inherent properties of the systems themselves. In particular, cross-Gramians play a pivotal role in frame theory and coorbit theory for example.

Turning our attention to systems $\mathfrak{M}_\alpha$ and $\widetilde{\mathfrak{M}}_\alpha$ of $\alpha$-molecules, a fundamental result will be the fact that the associated cross-Gramian $\mathcal{G}[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha]$ can be bounded based solely on the order and the parametrization of the respective systems. This is shown in Theorem 2.2.2, the main result of the next subsection. Moreover, it is possible to define a distance function $\omega_\alpha : \mathbb{P} \times \mathbb{P} \to [1, \infty)$ such that for $\alpha$-molecules $m_\lambda \in \mathfrak{M}$ and $\tilde{m}_\mu \in \widetilde{\mathfrak{M}}$ a higher distance $\omega_\alpha(\mathbf{x}_\lambda, \mathbf{x}_\mu)$ of their respective phase-space coordinates corresponds to a lower cross-correlation.

### 2.2.1 The Index Distance $\omega_\alpha$

The distance $\omega_\alpha(\mathbf{x}, \mathbf{y})$ between two points $\mathbf{x} = (x, \eta, t)$, $\mathbf{y} = (y, \theta, u) \in \mathbb{P}$ must certainly take into account their spatial, scale, and orientational relations. Hence, we first introduce a notion of distance on each single component of $\mathbb{P} = \mathbb{R}^2 \times \mathbb{T} \times \mathbb{R}_+$. Later, we will use those to assemble the desired distance $\omega_\alpha$ on $\mathbb{P}$.

A canonical choice for the distance between two points $x, y \in \mathbb{R}^2$ ist the Euclidean distance $|x - y|$ induced by the Euclidean norm $|\cdot|$.

On $\mathbb{T} = [0, 2\pi)$, we first define the distance function

$$d_{\mathbb{S}}(\eta, \theta) := \arccos(\langle e_\eta, e_\theta \rangle) \in [0, \pi], \quad \eta, \theta \in \mathbb{T}, \tag{2.8}$$

which essentially measures the distance of the associated orientation vectors $e_\eta$ and $e_\theta$ given as in (2.2) on the sphere $\mathbb{S}^1$. Due to the symmetries of (2.5), the distance $d_{\mathbb{S}}(\eta, \theta)$ is then further projected onto the interval $\mathbf{T} := [-\pi/2, \pi/2)$. For this we introduce the so-called *projective bracket*. With respect to a given half-open interval $\mathbf{I} \subset \mathbb{R}$ of finite length $|\mathbf{I}| < \infty$, this is the function

$$\{\cdot\}_{\mathbf{I}} : \mathbb{R} \to \mathbf{I}, \quad \eta \mapsto \{\eta\}_{\mathbf{I}}, \tag{2.9}$$

which maps a number $\eta \in \mathbb{R}$ to the unique element $\{\eta\}_{\mathbf{I}}$ in the set $\{\eta + m|\mathbf{I}| : m \in \mathbb{Z}\} \cap \mathbf{I}$. Since we will use this bracket mainly for the interval $\mathbf{T} = [-\pi/2, \pi/2)$, we further introduce the abbreviation $\{\cdot\} := \{\cdot\}_{\mathbf{T}}$. A suitable measure for the orientational distance is then given by $|\{\eta - \theta\}| = |\{d_{\mathbb{S}}(\eta, \theta)\}|$.

Finally, due to the multiplicative structure of the ray $\mathbb{R}_+$, the ratio $\max\{t/u, u/t\}$ is a natural way to measure the distance between different scales $t, u \in \mathbb{R}_+$.

Now we are ready to define the $\alpha$-scaled index distance $\omega_\alpha$ on $\mathbb{P}$ analogous to [59]. It can be viewed as a natural extension of Hart Smith's pseudo-distance [104, Def. 2.1].

**Definition 2.2.1** (compare [59, Def. 4.1])**.** Let $\alpha \in [0, 1]$. The *$\alpha$-scaled index distance* $\omega_\alpha : \mathbb{P} \times \mathbb{P} \to [1, \infty)$ is defined for two points $\mathbf{x} = (x, \eta, t) \in \mathbb{P}$ and $\mathbf{y} = (y, \theta, u) \in \mathbb{P}$ as

$$\omega_\alpha(\mathbf{x}, \mathbf{y}) := \max\left\{\frac{t}{u}, \frac{u}{t}\right\}(1 + d_\alpha(\mathbf{x}, \mathbf{y})), \tag{2.10}$$

with $d_\alpha(\mathbf{x}, \mathbf{y})$ being defined by

$$d_\alpha(\mathbf{x}, \mathbf{y}) := t_0^{-2\alpha}|x - y|^2 + t_0^{-2(1-\alpha)}|\{\eta - \theta\}|^2 + \frac{t_0^{-2}|\langle e_\eta, x - y\rangle|^2}{1 + t_0^{-2(1-\alpha)}|\{\eta - \theta\}|^2}, \tag{2.11}$$

where $t_0 := \max\{t, u\}$ and $e_\eta = R_\eta^{-1} e_1$ is the orientation vector from (2.2).

We now come to a core result of the theory of $\alpha$-molecules, Theorem 2.2.2, which provides a bound for the cross-Gramian $\mathcal{G}[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha]$ of two systems of $\alpha$-molecules $\mathfrak{M}_\alpha = \{m_\lambda\}_{\lambda \in \Lambda}$ and $\widetilde{\mathfrak{M}}_\alpha = \{\tilde{m}_\mu\}_{\mu \in \Delta}$. It draws a connection between the size of the cross-correlations $\langle m_\lambda, \tilde{m}_\mu \rangle$, the order of the respective $\alpha$-molecules $m_\lambda$ and $\tilde{m}_\mu$, and their positions in phase space.

**Theorem 2.2.2** (compare [59, Thm. 4.2])**.** *Let $\alpha \in [0, 1]$, and let $\{m_\lambda\}_{\lambda \in \Lambda}$ and $\{\tilde{m}_\mu\}_{\mu \in \Delta}$ be two systems of $\alpha$-molecules of order $(L, M, N_1, N_2)$ with respective parametrizations $(\Lambda, \Phi_\Lambda)$ and $(\Delta, \Phi_\Delta)$. Further assume that there exists some constant $C > 0$ such that*

$$t_\lambda, t_\mu \le C \quad \textit{for all } \lambda \in \Lambda, \mu \in \Delta, \ \textit{where } (x_\lambda, \eta_\lambda, t_\lambda) := \Phi_\Lambda(\lambda), (x_\mu, \eta_\mu, t_\mu) := \Phi_\Delta(\mu).$$

*Then, for every positive integer $N \in \mathbb{N}$ satisfying*

$$L \geq 2N, \quad M > 3N - \frac{3-\alpha}{2}, \quad N_1 \geq N + \frac{1+\alpha}{2}, \quad N_2 \geq 2N,$$

*there exists a corresponding constant $C_N > 0$ such that*

$$|\langle m_\lambda, \tilde{m}_\mu \rangle| \leq C_N \omega_\alpha \big(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu)\big)^{-N} \quad \textit{for all } \lambda \in \Lambda, \mu \in \Delta.$$

*Proof.* The proof requires some preparation and is therefore outsourced to the appendix, Section 2.5. □

As a consequence of this theorem, under appropriate assumptions on the parametrizations, the cross-Gramian $\mathcal{G}[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha]$ of two systems of $\alpha$-molecules $\mathfrak{M}_\alpha = \{m_\lambda\}_{\lambda \in \Lambda}$ and $\widetilde{\mathfrak{M}}_\alpha = \{\tilde{m}_\mu\}_{\mu \in \Delta}$ is well-localized in the sense of a fast off-diagonal decay with respect to the index distance $\omega_\alpha$. Put differently, $\mathcal{G}[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha]$ is then close to a diagonal matrix and the corresponding systems $\mathfrak{M}_\alpha$ and $\widetilde{\mathfrak{M}}_\alpha$ are almost orthogonal to each other, a fact which is sometimes referred to as the '*almost orthogonality of systems of $\alpha$-molecules*'.

This property has many implications, see for instance [62, 58, 65]. We will use it to derive Theorem 2.3.6, and Theorems 4.5.5 and 4.5.7.

### A simplified version of $\omega_\alpha$

The index distance $\omega_\alpha$ given in Definition 2.2.1 is not the only possible way to introduce a meaningful distance on $\mathbb{P}$. Another simpler version was put forward in [65] for example. To distinguish it from the distance $\omega_\alpha$, we will subsequently call it the *simplified index distance*. Its definition is as follows.

**Definition 2.2.3** ([65])**.** Let $\alpha \in [0,1]$. The *simplified $\alpha$-scaled index distance* $\omega_\alpha^{\mathrm{sim}} : \mathbb{P} \times \mathbb{P} \to [1,\infty)$ is defined for two points $\mathbf{x} = (x,\eta,t) \in \mathbb{P}$ and $\mathbf{y} = (y,\theta,u) \in \mathbb{P}$ as

$$\omega_\alpha^{\mathrm{sim}}(\mathbf{x},\mathbf{y}) := \max\left\{\frac{t}{u}, \frac{u}{t}\right\}\big(1 + d_\alpha^{\mathrm{sim}}(\mathbf{x},\mathbf{y})\big),$$

with $d_\alpha^{\mathrm{sim}}(\mathbf{x},\mathbf{y})$ given by

$$d_\alpha^{\mathrm{sim}}(\mathbf{x},\mathbf{y}) := t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2 + t_0^{-2\alpha}|x-y|^2 + t_0^{-1}|\langle e_\eta, x-y\rangle|,$$

where $t_0 = \max\{t,u\}$ and $e_\eta = (\cos(\eta), -\sin(\eta)) = R_\eta^{-1} e_1$.

The simplified index distance $\omega_\alpha^{\mathrm{sim}}$ shares many properties with the distance $\omega_\alpha$. In particular, Theorem 2.2.2 still holds true, which is a consequence of the following lemma.

**Lemma 2.2.4.** *We have uniformly for all $\mathbf{x},\mathbf{y} \in \mathbb{P}$*

$$d_\alpha^{\mathrm{sim}}(\mathbf{x},\mathbf{y}) \lesssim d_\alpha(\mathbf{x},\mathbf{y}) \quad \textit{and} \quad \omega_\alpha^{\mathrm{sim}}(\mathbf{x},\mathbf{y}) \lesssim \omega_\alpha(\mathbf{x},\mathbf{y}).$$

*Proof.* Using the inequality of the arithmetic and geometric means, we obtain

$$\begin{aligned} t_0^{-1}|\langle e_\eta, x-y\rangle| &= \left(\Big(1 + t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2\Big)\frac{t_0^{-2}|\langle e_\eta, x-y\rangle|^2}{1 + t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2}\right)^{1/2} \\ &\leq \frac{1}{2}\left(1 + t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2 + \frac{t_0^{-2}|\langle e_\eta, x-y\rangle|^2}{1 + t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2}\right). \end{aligned}$$

This establishes

$$\mathrm{d}_\alpha^{\mathrm{sim}}(\mathbf{x},\mathbf{y}) \le \frac{3}{2} d_\alpha(\mathbf{x},\mathbf{y}) \quad \text{and in turn also} \quad \omega_\alpha^{\mathrm{sim}}(\mathbf{x},\mathbf{y}) \le \frac{3}{2}\omega_\alpha(\mathbf{x},\mathbf{y}).$$

□

Due to the previous lemma, we have for arbitrary $N \in \mathbb{N}$ and every $\mathbf{x},\mathbf{y} \in \mathbb{P}$

$$\omega_\alpha(\mathbf{x},\mathbf{y})^{-N} \lesssim \omega_\alpha^{\mathrm{sim}}(\mathbf{x},\mathbf{y})^{-N}.$$

Hence, Theorem 2.2.2 still holds true for $\omega_\alpha^{\mathrm{sim}}$. Let us record this observation.

**Remark 2.2.5.** Theorem 2.2.2 still holds true if $\omega_\alpha$ is replaced by $\omega_\alpha^{\mathrm{sim}}$.

Clearly, a suitable index distance shall mirror the decay of the cross-Gramian as closely as possible. Unfortunately, as a trade-off for its simplicity, the distance $\omega_\alpha^{\mathrm{sim}}$ is weaker than $\omega_\alpha$ since the opposite estimates $d_\alpha(\mathbf{x},\mathbf{y}) \lesssim d_\alpha^{\mathrm{sim}}(\mathbf{x},\mathbf{y})$ and $\omega_\alpha(\mathbf{x},\mathbf{y}) \lesssim \omega_\alpha^{\mathrm{sim}}(\mathbf{x},\mathbf{y})$ do not hold true. Therefore, we prefer $\omega_\alpha$. However, let us mention that in other publications $\omega_\alpha^{\mathrm{sim}}$ has been used, e.g. in [65].

### 2.2.2 Metric Properties of $\omega_\alpha$

Recalling Definition 2.2.1, we next observe that due to $\{\eta-\theta\} = \{\{\eta\}-\{\theta\}\}$ the distance $\omega_\alpha(\mathbf{x},\mathbf{y})$ between two points $\mathbf{x}=(x,\eta,t)$ and $\mathbf{y}=(y,\theta,u) \in \mathbb{P}$ only depends on the values $\{\eta\}$ and $\{\theta\}$. Hence it makes sense to define the *reduced parameter space* below.

**Definition 2.2.6.** The *reduced parameter space* $\mathbf{P}$ is defined by

$$\mathbf{P} := \mathbb{R}^2 \times \mathbf{T} \times \mathbb{R}_+,$$

where here and throughout the thesis $\mathbf{T} := [-\pi/2, \pi/2)$.

The parameter space $\mathbb{P}$ is mapped onto the reduced space $\mathbf{P}$ via the *canonical projection*

$$\mathfrak{p} : \mathbb{P} \to \mathbf{P}, \quad (x,\eta,t) \mapsto (x.\{\eta\},t), \tag{2.12}$$

where $\{\cdot\} = \{\cdot\}_{\mathbf{T}} : \mathbb{T} \to \mathbf{T}$ denotes the projective bracket introduced in (2.9). This projection also induces an equivalence relation $\mathbf{x} \sim_{\mathfrak{p}} \mathbf{y}$ on $\mathbb{P}$. Each point $\mathbf{x}=(x,\eta,t)\in\mathbb{P}$ belongs to an equivalence class $[\mathbf{x}]_{\mathfrak{p}} := \mathfrak{p}^{-1}(\mathfrak{p}(\mathbf{x}))$ consisting precisely of two points, namely

$$[\mathbf{x}]_{\mathfrak{p}} = \Big\{(x,\eta,t), (x,(\eta+\pi)_{2\pi},t)\Big\}, \tag{2.13}$$

where here the short-hand notation $(\eta+\pi)_{2\pi} := (\eta+\pi) \bmod 2\pi$ is used.

Since $\omega_\alpha(\mathbf{x},\mathbf{y}) = \omega_\alpha(\tilde{\mathbf{x}},\tilde{\mathbf{y}})$ for points $\mathbf{x},\mathbf{y},\tilde{\mathbf{x}},\tilde{\mathbf{y}} \in \mathbb{P}$ whenever $\mathbf{x} \sim_{\mathfrak{p}} \tilde{\mathbf{x}}$ and $\mathbf{y} \sim_{\mathfrak{p}} \tilde{\mathbf{y}}$, the index distance $\omega_\alpha$ from Definition 2.2.1 gives rise to a distance on $\mathbf{P}$, for which we will use the same notation. It will always be clear from the context, which distance we refer to. The main result of this subsection will be that the induced distance $\omega_\alpha$ is a *multiplicative quasi-metric* on $\mathbf{P}$, a notion made precise by the following definition.

**Definition 2.2.7.** A *multiplicative quasi-metric* on $\mathbf{P}$ is a function $\omega : \mathbf{P}\times\mathbf{P} \to [1,\infty)$ which satisfies the following three axioms, where $\mathbf{x},\mathbf{y},\mathbf{z} \in \mathbf{P}$ are arbitrary:

($\tilde{Q}1$) $\omega(\mathbf{x},\mathbf{y}) = 1 \quad \Leftrightarrow \quad \mathbf{x} = \mathbf{y}$,

($\tilde{Q}2$) $\omega(\mathbf{x},\mathbf{y}) \le \big(\omega(\mathbf{y},\mathbf{x})\big)^{C_S}$ for some constant $C_S \ge 1$,

($\tilde{Q}3$) $\omega(\mathbf{x},\mathbf{y}) \le \big(\omega(\mathbf{x},\mathbf{z})\omega(\mathbf{z},\mathbf{y})\big)^{C_T}$ for some constant $C_T \ge 1$.

The axioms ($\tilde{Q}1$)-($\tilde{Q}3$) basically state that $\omega$ is – in a multiplicative sense – positive definite, quasi-symmetric, and satisfies a quasi-triangle inequality. An associated additive quasi-metric is obtained by taking the logarithm of $\omega$, see Definition 2.2.13 in the next subsection.

We call two multiplicative quasi-metrics $\omega : \mathbf{P} \times \mathbf{P} \to [1,\infty)$ and $\tilde{\omega} : \mathbf{P} \times \mathbf{P} \to [1,\infty)$ *Lipschitz equivalent in multiplicative sense* if there exists a constant $C \ge 1$ such that

$$\big(\omega(\mathbf{x},\mathbf{y})\big)^{1/C} \le \tilde{\omega}(\mathbf{x},\mathbf{y}) \le \big(\omega(\mathbf{x},\mathbf{y})\big)^{C} \quad \text{for all } \mathbf{x},\mathbf{y} \in \mathbf{P}.$$

They are called *Lipschitz equivalent in additive sense* if there exists a constant $C \ge 1$ with

$$\frac{1}{C}\omega(\mathbf{x},\mathbf{y}) \le \tilde{\omega}(\mathbf{x},\mathbf{y}) \le C\omega(\mathbf{x},\mathbf{y}) \quad \text{for all } \mathbf{x},\mathbf{y} \in \mathbf{P}.$$

In the sequel, we will prove that the reduced distance $\omega_\alpha$ on $\mathbf{P}$ satisfies the axioms ($\tilde{Q}1$)-($\tilde{Q}3$) of a multiplicative quasi-metric with constants $C_S = 3$ and $C_T = 6$. This will be stated in Theorem 2.2.12 whose proof requires some preparation.

Let us first look at the axiom ($\tilde{Q}1$). As follows directly from Definition 2.2.1, the index distance $\omega_\alpha$ satisfies the relation

$$\omega_\alpha(\mathbf{x},\mathbf{y}) = 1 \quad \Leftrightarrow \quad \mathbf{x} \sim_{\mathfrak{p}} \mathbf{y}. \tag{2.14}$$

Property ($\tilde{Q}1$) of the induced distance $\omega_\alpha$ on $\mathbf{T}$ is thus evident.

Concerning the axioms ($\tilde{Q}2$) and ($\tilde{Q}3$), it is more convenient to investigate those properties of $\omega_\alpha$ directly on $\mathbb{P}$. Our investigation will show that $\omega_\alpha : \mathbb{P} \times \mathbb{P} \to [1,\infty)$ satisfies

$$\omega_\alpha(\mathbf{x},\mathbf{y}) \le \big(\omega_\alpha(\mathbf{y},\mathbf{x})\big)^{C_S} \quad \text{and} \quad \omega_\alpha(\mathbf{x},\mathbf{y}) \le \big(\omega_\alpha(\mathbf{x},\mathbf{z})\omega_\alpha(\mathbf{z},\mathbf{y})\big)^{C_T} \tag{2.15}$$

with constants $C_S, C_T \ge 1$ independent of $\mathbf{x},\mathbf{y} \in \mathbb{P}$. As a consequence, ($\tilde{Q}2$) and ($\tilde{Q}3$) then clearly also hold true for the induced distance $\omega_\alpha : \mathbf{T} \times \mathbf{T} \to [1,\infty)$.

Our investigation will further establish that for all $\mathbf{x},\mathbf{y} \in \mathbb{P}$

$$\omega_\alpha(\mathbf{x},\mathbf{y}) \le C_S\omega_\alpha(\mathbf{y},\mathbf{x}) \quad \text{and} \quad \omega_\alpha(\mathbf{x},\mathbf{y}) \le C_T\omega_\alpha(\mathbf{x},\mathbf{z})\omega_\alpha(\mathbf{z},\mathbf{y}), \tag{2.16}$$

a property which has also been proved for the simplified distance $\omega_\alpha^{\mathrm{sim}}$ in [65].

In our investigation of (2.15) and (2.16), let us first focus on the symmetry properties. We note that the last term on the right-hand side of (2.11) prevents the index distance $\omega_\alpha$ from being symmetric. This is somewhat unsatisfactory since, in view of Theorem 2.2.2 and the symmetry of the Gramian, a symmetric version of $\omega_\alpha$ seems more appropriate.

It is possible however to symmetrize $\omega_\alpha$ by adding a fourth term in (2.11). We define for $\mathbf{x} = (x,\eta,t), \mathbf{y} = (y,\theta,u) \in \mathbb{P}$

$$d_\alpha^{\mathrm{sym}}(\mathbf{x},\mathbf{y}) := d_\alpha(\mathbf{x},\mathbf{y}) + \frac{t_0^{-2}|\langle e_\theta, x-y\rangle|^2}{1 + t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2},$$

where $t_0 := \max\{t, u\}$ and $e_\theta = R_\theta^{-1} e_1$. The resulting *symmetric distance*

$$\omega_\alpha^{\mathrm{sym}}(\mathbf{x}, \mathbf{y}) := \max\Big\{\frac{t}{u}, \frac{u}{t}\Big\}(1 + d_\alpha^{\mathrm{sym}}(\mathbf{x}, \mathbf{y}))$$

then still satisfies Theorem 2.2.2, which is due to the symmetry of the Gramian. Even more, since $\omega_\alpha \le \omega_\alpha^{\mathrm{sym}}$, the symmetric distance $\omega_\alpha^{\mathrm{sym}}$ is at least as strong as $\omega_\alpha$ and thus even seems to strengthen the statement of this theorem. However, as we will see below, $\omega_\alpha^{\mathrm{sym}}$ is in fact Lipschitz equivalent to $\omega_\alpha$ and can thus be considered as just another version of $\omega_\alpha$. Due to its more complicated structure, we prefer the distance $\omega_\alpha$.

Also other modifications of the index distance $\omega_\alpha$ are possible. For example, the definition of $\omega_\alpha$ is rather robust with respect to perturbations of $e_\eta$. To see this, let us define the subset $V_\eta(\theta)$ of the sphere $\mathbb{S}^1 \subset \mathbb{R}^2$, depending on $\eta, \theta \in \mathbb{T}$, by

$$V_\eta(\theta) := \Big\{ e \in \mathbb{S}^1 \ : \ |\langle e_\eta, e\rangle| \ge |\langle e_\eta, e_\theta\rangle| \Big\},$$

where $e_\eta$ and $e_\theta$ denote the orientation vectors from (2.2). Further, let us assign unit vectors $e(\mathbf{x}, \mathbf{y}) \in V_\eta(\theta)$ to all pairs $(\mathbf{x}, \mathbf{y}) \in \mathbb{P} \times \mathbb{P}$ and then modify the definition of the distance $\omega_\alpha$ by replacing the term $d_\alpha(\mathbf{x}, \mathbf{y})$ in (2.10) with the expression

$$\tilde{d}_\alpha(\mathbf{x}, \mathbf{y}) := t_0^{-2(1-\alpha)}|\{\eta - \theta\}|^2 + t_0^{-2\alpha}|x - y|^2 + \frac{t_0^{-2}|\langle e(\mathbf{x}, \mathbf{y}), x - y\rangle|^2}{1 + t_0^{-2(1-\alpha)}|\{\eta - \theta\}|^2}. \tag{2.17}$$

This gives rise to a new index distance $\tilde{\omega}_\alpha$ on $\mathbb{P}$ of the form

$$\tilde{\omega}_\alpha(\mathbf{x}, \mathbf{y}) := \max\Big\{\frac{t}{u}, \frac{u}{t}\Big\}(1 + \tilde{d}_\alpha(\mathbf{x}, \mathbf{y})). \tag{2.18}$$

As a consequence of the following lemma, like $\omega_\alpha^{\mathrm{sym}}$, the new distance $\tilde{\omega}_\alpha$ is Lipschitz equivalent to $\omega_\alpha$, both in additive and multiplicative sense. Hence, yet again, in essence the distance $\tilde{\omega}_\alpha$ is just another version of the original distance $\omega_\alpha$.

**Lemma 2.2.8.** *With $d_\alpha$ and $\omega_\alpha$ given as in Definition 2.2.1 and $\tilde{d}_\alpha$ and $\tilde{\omega}_\alpha$ given as in* (2.17) *and* (2.18)*, it holds uniformly for all points $\mathbf{x}, \mathbf{y} \in \mathbb{P}$*

$$\frac{1}{3} d_\alpha(\mathbf{x}, \mathbf{y}) \le \tilde{d}_\alpha(\mathbf{x}, \mathbf{y}) \le 3 d_\alpha(\mathbf{x}, \mathbf{y}) \quad \text{and} \quad \frac{1}{3}\omega_\alpha(\mathbf{x}, \mathbf{y}) \le \tilde{\omega}_\alpha(\mathbf{x}, \mathbf{y}) \le 3\omega_\alpha(\mathbf{x}, \mathbf{y}).$$

*Furthermore, for all $\mathbf{x}, \mathbf{y} \in \mathbb{P}$*

$$\big(\omega_\alpha(\mathbf{x}, \mathbf{y})\big)^{1/3} \le \tilde{\omega}_\alpha(\mathbf{x}, \mathbf{y}) \le \big(\omega_\alpha(\mathbf{x}, \mathbf{y})\big)^3.$$

*Proof.* Assume first that $e \in V_\eta(\theta)$ is such that $\langle e, e_\eta\rangle \ge |\langle e_\eta, e_\theta\rangle|$. Then

$$|e - e_\eta|^2 = 2 - 2\langle e, e_\eta\rangle \le 2 - 2|\langle e_\eta, e_\theta\rangle| = \min\{|e_\eta - e_\theta|^2, |e_\eta + e_\theta|^2\},$$

and hence

$$|e - e_\eta| \le \min\{|e_\eta - e_\theta|, |e_\eta + e_\theta|\} \le |\{\eta - \theta\}|.$$

It follows

$$\frac{t_0^{-2}|\langle e - e_\eta, x - y\rangle|^2}{1 + t_0^{-2(1-\alpha)}|\{\eta - \theta\}|^2} \le \frac{t_0^{-2(1-\alpha)}|\{\eta - \theta\}|^2}{1 + t_0^{-2(1-\alpha)}|\{\eta - \theta\}|^2} t_0^{-2\alpha}|x - y|^2 \le t_0^{-2\alpha}|x - y|^2.$$

We conclude, now for arbitrary $e \in V_\eta(\theta)$,

$$\frac{t_0^{-2}|\langle e, x-y\rangle|^2}{1+t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2} \le 2\Big(\frac{t_0^{-2}|\langle e_\eta, x-y\rangle|^2}{1+t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2} + t_0^{-2\alpha}|x-y|^2\Big), \tag{2.19}$$

and analogously

$$\frac{t_0^{-2}|\langle e_\eta, x-y\rangle|^2}{1+t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2} \le 2\Big(\frac{t_0^{-2}|\langle e, x-y\rangle|^2}{1+t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2} + t_0^{-2\alpha}|x-y|^2\Big).$$

This proves

$$\tilde{d}_\alpha(\mathbf{x},\mathbf{y}) \le 3d_\alpha(\mathbf{x},\mathbf{y}) \quad \text{and} \quad d_\alpha(\mathbf{x},\mathbf{y}) \le 3\tilde{d}_\alpha(\mathbf{x},\mathbf{y}),$$

and in turn also

$$\tilde{\omega}_\alpha(\mathbf{x},\mathbf{y}) \le 3\omega_\alpha(\mathbf{x},\mathbf{y}) \quad \text{and} \quad \omega_\alpha(\mathbf{x},\mathbf{y}) \le 3\tilde{\omega}_\alpha(\mathbf{x},\mathbf{y}).$$

Further, we deduce

$$\begin{aligned}\omega_\alpha(\mathbf{x},\mathbf{y}) = \max\Big\{\frac{t}{u},\frac{u}{t}\Big\}(1+d_\alpha(\mathbf{x},\mathbf{y})) &\le \max\Big\{\frac{t}{u},\frac{u}{t}\Big\}(1+3\tilde{d}_\alpha(\mathbf{x},\mathbf{y})) \\ &\le \max\Big\{\frac{t}{u},\frac{u}{t}\Big\}^3(1+\tilde{d}_\alpha(\mathbf{x},\mathbf{y}))^3 = \big(\tilde{\omega}_\alpha(\mathbf{x},\mathbf{y})\big)^3.\end{aligned}$$

The other direction $\tilde{\omega}_\alpha(\mathbf{x},\mathbf{y}) \le \big(\omega_\alpha(\mathbf{x},\mathbf{y})\big)^3$ follows analogously. □

As a direct corollary of Lemma 2.2.8, we can now deduce the additive and multiplicative quasi-symmetry of $\omega_\alpha$.

**Corollary 2.2.9.** *The index distance $\omega_\alpha : \mathbb{P}\times\mathbb{P} \to [1,\infty)$ is quasi-symmetric, both in an additive and multiplicative sense, with associated quasi-symmetry constant $C_S = 3$.*

*Proof.* Let $\tilde{\omega}_\alpha(\mathbf{x},\mathbf{y})$ be defined as in (2.18) and choose $e(\mathbf{x},\mathbf{y}) := e_\theta$ in (2.17) for all $\mathbf{x} = (x,\eta,t) \in \mathbb{P}$ and $\mathbf{y} = (y,\theta,u) \in \mathbb{P}$. Since $e_\theta \in V_\eta(\theta)$ holds true for all $\eta,\theta \in \mathbb{T}$, Lemma 2.2.8 can be applied, showing that $\tilde{\omega}_\alpha$ is Lipschitz equivalent to $\omega_\alpha$, both in additive and multiplicative sense. The observation $\tilde{\omega}_\alpha(\mathbf{x},\mathbf{y}) = \omega_\alpha(\mathbf{y},\mathbf{x})$ finishes the proof. □

With Lemma 2.2.8 as a tool, we can also show that the symmetric distance $\omega_\alpha^{\mathrm{sym}}$ is additively and multiplicatively Lipschitz equivalent to $\omega_\alpha$. Clearly, for all $\mathbf{x},\mathbf{y} \in \mathbb{P}$

$$d_\alpha(\mathbf{x},\mathbf{y}) \le d_\alpha^{\mathrm{sym}}(\mathbf{x},\mathbf{y}) \le d_\alpha(\mathbf{x},\mathbf{y}) + d_\alpha(\mathbf{y},\mathbf{x}) \le 4d_\alpha(\mathbf{x},\mathbf{y}).$$

From this we get

$$\omega_\alpha(\mathbf{x},\mathbf{y}) \le \omega_\alpha^{\mathrm{sym}}(\mathbf{x},\mathbf{y}) \le \omega_\alpha(\mathbf{x},\mathbf{y}) + \omega_\alpha(\mathbf{y},\mathbf{x}) \le 4\omega_\alpha(\mathbf{x},\mathbf{y})$$

and

$$\omega_\alpha(\mathbf{x},\mathbf{y}) \le \omega_\alpha^{\mathrm{sym}}(\mathbf{x},\mathbf{y}) \le \big(\omega_\alpha(\mathbf{x},\mathbf{y})\big)^4.$$

Finally, let us remark that Lemma 2.2.8 can also be used to obtain another symmetric variant of $\omega_\alpha$. To this end, we define the unit vectors $e_{\eta,\theta} \in \mathbb{S}^1$ 'in the middle' of $e_\eta$ and $e_\theta$ as follows

$$e_{\eta,\theta} := \begin{cases} |e_\eta + e_\theta|^{-1}(e_\eta + e_\theta) & , \langle e_\eta, e_\theta \rangle \geq 0\,, \\ |e_\eta - e_\theta|^{-1}(e_\eta - e_\theta) & , \langle e_\eta, e_\theta \rangle < 0\,. \end{cases}$$

In case $e_\eta \neq \pm e_\theta$, the vector $e_{\eta,\theta}$ is the unique unit vector $e \in \mathbb{S}^1$ characterized by $\langle e, e_\eta \rangle > 0$, $\langle e, e_\theta \rangle > 0$, and $|\langle e, e_\eta \rangle| = |\langle e, e_\theta \rangle|$. Since $e_{\eta,\theta} = \pm e_{\theta,\eta}$, choosing $e(\mathbf{x},\mathbf{y}) := e_{\eta,\theta}$ in (2.17) for $\mathbf{x} = (x,\eta,t)$, $\mathbf{y} = (y,\theta,u) \in \mathbb{P}$ yields a symmetric index distance $\tilde{\omega}_\alpha$ in (2.18) which is equivalent to the original distance $\omega_\alpha$. Let us record this fact.

**Remark 2.2.10.** A symmetric version of $\tilde{\omega}_\alpha$ is obtained by choosing $e(\mathbf{x},\mathbf{y}) = e_{\eta,\theta}$ in (2.17).

Up to now, we have verified ($\tilde{Q}1$) and studied symmetry properties of $\omega_\alpha$, thereby proving the symmetry relations in (2.15) and (2.16), and as a consequence also axiom ($\tilde{Q}2$). It remains to prove the triangle inequalities in (2.15) and (2.16), which then also give ($\tilde{Q}3$). For this we need the following elementary observation.

**Lemma 2.2.11.** *For $\eta, \theta, \kappa \in \mathbb{R}$ the following triangle inequality holds true*

$$|\{\eta - \theta\}| \leq |\{\eta - \kappa\}| + |\{\kappa - \theta\}|.$$

*Proof.* Let us first assume $\theta = 0$, $\eta, \kappa \in \mathbf{T} = [-\pi/2, \pi/2)$. Then indeed $|\eta| \leq |\{\eta - \kappa\}| + |\kappa|$. Next, for general $\eta, \kappa$ we plug $\tilde{\eta} = \eta - k_1\pi \in \mathbf{T}$ and $\tilde{\kappa} = \kappa - k_2\pi \in \mathbf{T}$ into the former inequality, where $k_1, k_2 \in \mathbb{Z}$ are chosen appropriately. One obtains $|\{\eta\}| \leq |\{\eta - \kappa\}| + |\{\kappa\}|$. Finally, we substitute $\eta - \theta$ for $\eta$ and $\kappa - \theta$ for $\kappa$, the proof is finished. $\square$

Now we are in the position to prove the main result of this subsection.

**Theorem 2.2.12.** *Let $\alpha \in [0,1]$. The index distance $\omega_\alpha : \mathbb{P} \times \mathbb{P} \to [1,\infty)$ introduced in Definition 2.2.1 satisfies* (2.14)*,* (2.15)*, and* (2.16)*. The quasi-symmetry constant can be chosen as $C_S = 3$, the quasi-triangle constant as $C_T = 6$. In particular, the induced distance $\omega_\alpha : \mathbf{P} \times \mathbf{P} \to [1,\infty)$ is a multiplicative quasi-metric on $\mathbf{P}$ with the same constants.*

*Proof.* Axiom ($\tilde{Q}1$) is clear, the quasi-symmetry with constant $C_S = 3$ was established in Corollary 2.2.9. To prove the triangle inequalities in (2.15) and (2.16), we take arbitrary $\mathbf{x} = (x,\eta,t), \mathbf{y} = (y,\theta,u), \mathbf{z} = (z,\kappa,v) \in \mathbb{P}$ and abbreviate $t_0 = \max\{t,u\}$, $u_0 = \max\{u,v\}$, $v_0 = \max\{v,t\}$. Our first observation is that for $t,u,v \in \mathbb{R}_+$

$$\max\left\{\frac{t}{u}, \frac{u}{t}\right\} \leq \max\left\{\frac{t}{v}, \frac{v}{t}\right\} \max\left\{\frac{v}{u}, \frac{u}{v}\right\}. \tag{2.20}$$

Further, it holds

$$M(t,u,v) := \max\left\{1, \frac{v^2}{t_0^2}\right\} \leq \max\left\{\frac{t}{u}, \frac{u}{t}\right\}^{-1} \max\left\{\frac{u}{v}, \frac{v}{u}\right\} \max\left\{\frac{v}{t}, \frac{t}{v}\right\}.$$

The quasi-triangle inequalities with $C_T = 6$ then follow if we can prove the validity of the inequality

$$d_\alpha(\mathbf{x},\mathbf{y}) \leq 6M(t,u,v)\Big(d_\alpha(\mathbf{x},\mathbf{z}) + d_\alpha(\mathbf{z},\mathbf{y}) + d_\alpha(\mathbf{x},\mathbf{z})d_\alpha(\mathbf{z},\mathbf{y})\Big). \tag{2.21}$$

Indeed, altogether, (2.20) and (2.21) yield

$$\begin{aligned}\omega_\alpha(\mathbf{x},\mathbf{y}) &= \max\Big\{\frac{t}{u},\frac{u}{t}\Big\}(1+d_\alpha(\mathbf{x},\mathbf{y}))\\ &\le \max\Big\{\frac{t}{v},\frac{v}{t}\Big\}\max\Big\{\frac{v}{u},\frac{u}{v}\Big\}\Big(1+6\big(d_\alpha(\mathbf{x},\mathbf{z})+d_\alpha(\mathbf{z},\mathbf{y})+d_\alpha(\mathbf{x},\mathbf{z})d_\alpha(\mathbf{z},\mathbf{y})\big)\Big).\end{aligned}$$

From here, one directly obtains

$$\omega_\alpha(\mathbf{x},\mathbf{y}) \le \big(\omega_\alpha(\mathbf{x},\mathbf{z})\omega_\alpha(\mathbf{z},\mathbf{y})\big)^6 \quad \text{and} \quad \omega_\alpha(\mathbf{x},\mathbf{y}) \le 6\big(\omega_\alpha(\mathbf{x},\mathbf{z})\omega_\alpha(\mathbf{z},\mathbf{y})\big).$$

In order to verify (2.21) we treat the different components of $d_\alpha(\mathbf{x},\mathbf{y})$ separately. Let us first record that in the range $\alpha \in [0,1]$ always

$$t_0^{-2\alpha} \le M(t,u,v)v_0^{-2\alpha} \quad \text{and} \quad t_0^{-2(1-\alpha)} \le M(t,u,v)v_0^{-2(1-\alpha)}. \tag{2.22}$$

Applying the triangle-inequality, $M(t,u,v) \ge 1$, and (2.22), yields

$$t_0^{-2\alpha}|x-y|^2 \le 2t_0^{-2\alpha}\Big(|x-z|^2+|z-y|^2\Big) \le 2M(t,u,v)\Big(t_0^{-2\alpha}|x-z|^2+v_0^{-2\alpha}|z-y|^2\Big). \tag{2.23}$$

Invoking Lemma 2.2.11, we analogously get

$$\begin{aligned}t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2 &\le 2t_0^{-2(1-\alpha)}\Big(|\{\eta-\kappa\}|^2+|\{\kappa-\theta\}|^2\Big)\\ &\le 2M(t,u,v)\Big(t_0^{-2(1-\alpha)}|\{\eta-\kappa\}|^2+v_0^{-2(1-\alpha)}|\{\kappa-\theta\}|^2\Big).\end{aligned} \tag{2.24}$$

To bound the last term of $d_\alpha(\mathbf{x},\mathbf{y})$, we choose the sign of $\tilde{e}_\kappa := \pm e_\kappa$ in such a way that $\langle e_\eta, \tilde{e}_\kappa\rangle \ge 0$ and then expand

$$\langle e_\eta, x-y\rangle = \langle e_\eta, x-z\rangle + \langle \tilde{e}_\kappa, z-y\rangle + \langle e_\eta - \tilde{e}_\kappa, z-y\rangle.$$

This leads to the estimate

$$|\langle e_\eta, x-y\rangle|^2 \le 3\Big(|\langle e_\eta, x-z\rangle|^2 + |\langle e_\kappa, z-y\rangle|^2 + |\langle e_\eta - \tilde{e}_\kappa, z-y\rangle|^2\Big).$$

With the triangle inequality from Lemma 2.2.11 we deduce

$$\frac{(1+t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2)(1+t_0^{-2(1-\alpha)}|\{\theta-\kappa\}|^2)}{1+t_0^{-2(1-\alpha)}|\{\eta-\kappa\}|^2} \ge \frac{1}{2}.$$

We conclude, with the help of (2.22) and $M(t,u,v) \ge 1$ in the second step,

$$\begin{aligned}\frac{t_0^{-2}|\langle e_\eta, x-z\rangle|^2}{1+t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2} &\le 2\frac{t_0^{-2}|\langle e_\eta, x-z\rangle|^2(1+t_0^{-2(1-\alpha)}|\{\theta-\kappa\}|^2)}{1+t_0^{-2(1-\alpha)}|\{\eta-\kappa\}|^2}\\ &\le 2M(t,u,v)\frac{t_0^{-2}|\langle e_\eta, x-z\rangle|^2(1+v_0^{-2(1-\alpha)}|\{\kappa-\theta\}|^2)}{1+t_0^{-2(1-\alpha)}|\{\eta-\kappa\}|^2}.\end{aligned}$$

Similarly, we first verify

$$\frac{t_0^{-2}|\langle e_\kappa, z-y\rangle|^2}{1+t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2} \le 2\frac{t_0^{-2}|\langle e_\kappa, z-y\rangle|^2(1+t_0^{-2(1-\alpha)}|\{\kappa-\eta\}|^2)}{1+t_0^{-2(1-\alpha)}|\{\kappa-\theta\}|^2}.$$

Assuming $v_0 \le t_0$ and using $M(t,u,v) \ge 1$ in the second step, we then obtain

$$\begin{aligned}\frac{t_0^{-2}|\langle e_\kappa, z-y\rangle|^2}{1+t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2} &\le 2\frac{v_0^{-2}|\langle e_\kappa, z-y\rangle|^2(1+t_0^{-2(1-\alpha)}|\{\eta-\kappa\}|^2)}{1+v_0^{-2(1-\alpha)}|\{\kappa-\theta\}|^2} \\ &\le 2M(t,u,v)\frac{v_0^{-2}|\langle e_\kappa, z-y\rangle|^2(1+t_0^{-2(1-\alpha)}|\{\eta-\kappa\}|^2)}{1+v_0^{-2(1-\alpha)}|\{\kappa-\theta\}|^2}.\end{aligned}$$

If $v_0 > t_0$ we argue differently, using $t_0^{-2} \le M(t,u,v)v_0^{-2}$ in the first step,

$$\begin{aligned}\frac{t_0^{-2}|\langle e_\kappa, z-y\rangle|^2}{1+t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2} &\le 2M(t,u,v)\frac{v_0^{-2}|\langle e_\kappa, z-y\rangle|^2(1+t_0^{-2(1-\alpha)}|\{\eta-\kappa\}|^2)}{1+t_0^{-2(1-\alpha)}|\{\kappa-\theta\}|^2} \\ &\le 2M(t,u,v)\frac{v_0^{-2}|\langle e_\kappa, z-y\rangle|^2(1+t_0^{-2(1-\alpha)}|\{\eta-\kappa\}|^2)}{1+v_0^{-2(1-\alpha)}|\{\kappa-\theta\}|^2}.\end{aligned}$$

Finally, we have

$$\frac{t_0^{-2}|\langle e_\eta-\tilde{e}_\kappa, z-y\rangle|^2}{1+t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2} \le M(t,u,v)t_0^{-2(1-\alpha)}v_0^{-2\alpha}|\{\eta-\kappa\}|^2|z-y|^2,$$

because of

$$|\langle e_\eta-\tilde{e}_\kappa, z-y\rangle|^2 \le |e_\eta-\tilde{e}_\kappa|^2|z-y|^2 \le |\{\eta-\kappa\}|^2|z-y|^2,$$

where $\langle e_\eta, \tilde{e}_\kappa\rangle \ge 0$ was used for the last estimate, and the inequality

$$\frac{t_0^{-2}}{1+t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2} \le t_0^{-2} \le M(t,u,v)t_0^{-2(1-\alpha)}v_0^{-2\alpha},$$

where again (2.22) was used in the last step.

All in all, we arrive at

$$\begin{aligned}\frac{t_0^{-2}|\langle e_\eta, x-y\rangle|^2}{1+t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2} \le 6M(t,u,v)\Big(&t_0^{-2(1-\alpha)}v_0^{-2\alpha}|\{\eta-\kappa\}|^2|z-y|^2+ \\ &+\frac{v_0^{-2}|\langle e_\kappa, z-y\rangle|^2 t_0^{-2(1-\alpha)}|\{\eta-\kappa\}|^2}{1+v_0^{-2(1-\alpha)}|\{\kappa-\theta\}|^2}+\frac{t_0^{-2}|\langle e_\eta, x-z\rangle|^2 v_0^{-2(1-\alpha)}|\{\kappa-\theta\}|^2}{1+t_0^{-2(1-\alpha)}|\{\eta-\kappa\}|^2}\Big).\end{aligned}$$

This bound together with (2.23) and (2.24) establishes (2.21), and the proof is finished. □

In this subsection, we have proved that $\omega_\alpha$ constitutes a multiplicative quasi-metric on the reduced parameter space $\mathbf{P}$. Further, we have seen that many different equivalent versions of $\omega_\alpha$ can be defined, even symmetric ones. For the subsequent theory it does not matter which version we use, since only the constants would be affected. The reason, why we stick to the original definition of $\omega_\alpha$ from Definition 2.2.1 is its simpler structure compared to the other versions.

### 2.2.3 $\alpha$-Balls

We will now discuss in more detail the quasi-metric structure of $\mathbb{P}$ induced by $\omega_\alpha$. In particular, we will investigate the 'balls' obtained from $\omega_\alpha$. To shift this investigation into more familiar territory, we first associate to $\omega_\alpha$ a corresponding *additive quasi-metric*. The axioms of this notion are recalled below.

**Definition 2.2.13.** An *additive quasi-metric* on $\mathbf{P}$ is a function $\omega : \mathbf{P} \times \mathbf{P} \to [0, \infty)$ which satisfies the following three axioms, where $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{P}$ are arbitrary:

(Q1) $\omega(\mathbf{x}, \mathbf{y}) = 0 \quad \Leftrightarrow \quad \mathbf{x} = \mathbf{y}$,

(Q2) $\omega(\mathbf{x}, \mathbf{y}) \le C_S \omega(\mathbf{y}, \mathbf{x})$ for some constant $C_S \ge 1$,

(Q3) $\omega(\mathbf{x}, \mathbf{y}) \le C_T\big(\omega(\mathbf{x}, \mathbf{z}) + \omega(\mathbf{z}, \mathbf{y})\big)$ for some constant $C_T \ge 1$.

Since, according to Theorem 2.2.12, $\omega_\alpha$ constitutes a multiplicative quasi-metric on $\mathbf{P}$ satisfying the axioms $(\tilde{Q}1)$-$(\tilde{Q}3)$ of Definition 2.2.7, clearly the function

$$\omega_\alpha^{\log}(\mathbf{x}, \mathbf{y}) := \log_2(\omega_\alpha(\mathbf{x}, \mathbf{y})) = |\log_2(t/u)| + \log_2\left(1 + d_\alpha(\mathbf{x}, \mathbf{y})\right)$$

defines an additive quasi-metric on $\mathbf{P}$. We further note that due to (2.15) $\omega_\alpha^{\log}$ also fulfills the axioms $(Q2)$ and $(Q3)$ as a function on $\mathbb{P} \times \mathbb{P}$.

The quasi-metric $\omega_\alpha^{\log}$ gives rise to an associated family of balls in $\mathbb{P}$. The *$\alpha$-balls of radius* $\tau \ge 0$ are defined by

$$B_\tau^\alpha(\mathbf{x}) := \Big\{\mathbf{y} \in \mathbb{P} \ : \ \omega_\alpha^{\log}(\mathbf{x}, \mathbf{y}) \le \tau\Big\} = \Big\{\mathbf{y} \in \mathbb{P} \ : \ \omega_\alpha(\mathbf{x}, \mathbf{y}) \le 2^\tau\Big\}. \tag{2.25}$$

Due to the non-symmetry of $\omega_\alpha^{\log}$, there also exist the *dual $\alpha$-balls of radius* $\tau \ge 0$ given by

$$B_\tau^{\prime;\alpha}(\mathbf{x}) := \Big\{\mathbf{y} \in \mathbb{P} \ : \ \omega_\alpha^{\log}(\mathbf{y}, \mathbf{x}) \le \tau\Big\} = \Big\{\mathbf{y} \in \mathbb{P} \ : \ \omega_\alpha(\mathbf{y}, \mathbf{x}) \le 2^\tau\Big\}.$$

In general, the dual $\alpha$-balls $B_\tau^{\prime;\alpha}(\mathbf{x})$ do not coincide with the primal $\alpha$-balls $B_\tau^\alpha(\mathbf{x})$. The relation between the two types of balls is expressed by the equivalence

$$\mathbf{y} \in B_\tau^\alpha(\mathbf{x}) \Leftrightarrow \mathbf{x} \in B_\tau^{\prime;\alpha}(\mathbf{y}).$$

Subsequently, we will mostly be interested in the primal balls $B_\tau^\alpha(\mathbf{x})$, in particular in a more explicit representation. Clearly, $B_0^\alpha(\mathbf{x}) = [\mathbf{x}]_{\mathfrak{p}}$ for every $\mathbf{x} \in \mathbb{P}$, where $[\mathbf{x}]_{\mathfrak{p}} = \mathfrak{p}^{-1}(\mathfrak{p}(\mathbf{x}))$ is the equivalence class (2.13) induced by the canonical projection $\mathfrak{p} : \mathbb{P} \to \mathbf{P}$ from (2.12). In case of radii $\tau > 0$, let us consider $\mathbf{y} = (y, \theta, u) \in B_\tau^\alpha(\mathbf{x})$ at some fixed position $\mathbf{x} = (x, \eta, t) \in \mathbb{P}$. We then derive the following necessary conditions from the definition of $\omega_\alpha$.

First we see that, due to $\omega_\alpha(\mathbf{x}, \mathbf{y}) \le 2^\tau$ and $1 + d_\alpha(\mathbf{x}, \mathbf{y}) \ge 1$, we necessarily have

$$\max\left\{\frac{t}{u}, \frac{u}{t}\right\} \le 2^\tau \quad \text{or equivalently} \quad |\log_2(t/u)| \le \tau.$$

Using $\alpha \in [0, 1]$ and $1 \le \max\{t/u, u/t\} \le 2^\tau$, we also get

$$t^{-2(1-\alpha)}|\{\eta - \theta\}|^2 \le 2^{2\tau} t_0^{-2(1-\alpha)}|\{\eta - \theta\}|^2 \le 2^{2\tau}(2^\tau - 1) \le 2^{4\tau} - 1.$$

Further, we deduce

$$t^{-2}|\langle e_1, R_\eta(x-y)\rangle|^2 \le 2^{2\tau} t_0^{-2}|\langle e_1, R_\eta(x-y)\rangle|^2 \le 2^{3\tau}(2^\tau - 1) \le 2^{4\tau} - 1,$$

where we applied

$$t_0^{-2}|\langle e_\eta, x-y\rangle|^2 \le (2^\tau - 1)(1 + t_0^{-2(1-\alpha)}|\{\eta - \theta\}|^2) \le 2^\tau(2^\tau - 1).$$

At last, we bound $t^{-2\alpha}|\langle e_2, R_\eta(x-y)\rangle|^2$ by

$$2^{2\tau} t_0^{-2\alpha}|\langle e_2, R_\eta(x-y)\rangle|^2 \le 2^{2\tau} t_0^{-2\alpha}|x-y|^2 \le 2^{2\tau}(2^\tau - 1) \le 2^{4\tau} - 1.$$

These estimates motivate the definition of another distance function on $\mathbb{P}$, namely

$$\mathbf{w}_\alpha(\mathbf{x}, \mathbf{y}) := \max\Big\{ t^{-\alpha}|\langle e_2, R_\eta(x-y)\rangle|, t^{-1}|\langle e_1, R_\eta(x-y)\rangle|, t^{-(1-\alpha)}|\{\eta - \theta\}|, |\log_2(t/u)| \Big\}.$$

Further, we introduce corresponding subsets of $\mathbb{P}$, for $\mathbf{x} = (x, \eta, t) \in \mathbb{P}$ and $\tau \ge 0$ the sets

$$V_\tau^\alpha(\mathbf{x}) := \Big(x + R_\eta^{-1} A_{\alpha,t} Q^\tau\Big) \times \Big(\eta + t^{1-\alpha} I^\tau\Big)_{2\pi} \times \Big(t J^\tau\Big), \tag{2.26}$$

where $Q^\tau := [-\tau, \tau]^2$, $I^\tau := [-\tau, \tau]$, $J^\tau := [2^{-\tau}, 2^\tau]$, and $(\cdot)_{2\pi} := (\cdot) \bmod 2\pi$, and collect these in

$$\mathcal{V}_\tau^\alpha := \mathcal{V}_\tau^\alpha[\mathbb{P}] := \Big\{ V_\tau^\alpha(\mathbf{x}) \ : \ \mathbf{x} \in \mathbb{P} \Big\}.$$

Then, for every $\mathbf{x} = (x, \eta, t) \in \mathbb{P}$ and $\tau \ge 0$ we can write

$$V_\tau^\alpha([\mathbf{x}]_\mathfrak{p}) := V_\tau^\alpha((x, \eta, t)) \cup V_\tau^\alpha((x, (\eta + \pi)_{2\pi}, t)) = \Big\{ \mathbf{y} \in \mathbb{P} \ : \ \mathbf{w}_\alpha(\mathbf{x}, \mathbf{y}) \le \tau \Big\},$$

where as above $[\mathbf{x}]_\mathfrak{p}$ denotes the equivalence class (2.13).

We will see below that the sets $V_\tau^\alpha([\mathbf{x}]_\mathfrak{p})$ are good approximations of the $\alpha$-balls $B_\tau^\alpha(\mathbf{x})$ as long as the scale parameter remains small. In Chapter 4, where we will only consider the subspace $\mathbb{X} = \mathbb{R}^2 \times \mathbb{T} \times (0, 1] \subset \mathbb{P}$, we can thus use them as convenient substitutes for $B_\tau^\alpha(\mathbf{x})$, which are easier to handle since they can be expressed explicitly by (2.26).

Again, there also exist dual sets which have a slightly more complicated structure. They are given by

$$\begin{aligned} V_\tau^{\prime,\alpha}(\mathbf{x}) &:= \Big\{ \mathbf{y} \in \mathbb{P} \ : \ \mathbf{x} \in V_\tau^\alpha(\mathbf{y}) \Big\} \\ &= \Big\{ \mathbf{y} = (y, \theta, u) \in \mathbb{P} \ : \ y \in x + R_\theta^{-1} A_{\alpha,u} Q^\tau, \theta \in \Big(\eta + u^{1-\alpha} I^\tau\Big)_{2\pi}, u \in t J^\tau \Big\}. \end{aligned}$$

In the following, we will investigate the relation of the sets $V_\tau^\alpha([\mathbf{x}]_\mathfrak{p})$ to the $\alpha$-balls $B_\tau^\alpha(\mathbf{x})$. As an immediate consequence of the definition of $\mathbf{w}_\alpha$, we obtain the lemma below.

**Lemma 2.2.14.** *Let $\alpha \in [0, 1]$. For all $\mathbf{x} \in \mathbb{P}$ and $\tau \ge 0$*

$$0 \le \sup_{\mathbf{y} \in B_\tau^\alpha(\mathbf{x})} \mathbf{w}_\alpha(\mathbf{x}, \mathbf{y}) \le \sqrt{2^{4\tau} - 1}.$$

*Proof.* Let $\mathbf{y} \in B_\tau^\alpha(\mathbf{x})$ for some fixed $\mathbf{x} \in \mathbb{P}$ and $\tau \geq 0$. Then, according to the estimates above,

$$\max\left\{t^{-\alpha}|\langle e_2, R_\eta(x-y)\rangle|, t^{-1}|\langle e_1, R_\eta(x-y)\rangle|, t^{-(1-\alpha)}|\{\eta-\theta\}|\right\} \leq \sqrt{2^{4\tau}-1}.$$

Further, it also holds

$$|\log_2(t/u)| \leq \tau \leq \sqrt{2^{4\tau}-1}.$$

□

Put differently, Lemma 2.2.14 yields the following inclusion, valid for all $\mathbf{x} \in \mathbb{P}$ and $\tau \geq 0$,

$$B_\tau^\alpha(\mathbf{x}) \subseteq V^\alpha_{\sqrt{2^{4\tau}-1}}([\mathbf{x}]_\mathfrak{p}).$$

To obtain a reverse inclusion, we need the following lemma.

**Lemma 2.2.15.** *Let $\alpha \in [0,1]$. For all $\mathbf{x} = (x,\eta,t) \in \mathbb{P}$ and $\tau \geq 0$*

$$1 \leq \sup_{\mathbf{y} \in V_\tau^\alpha(\mathbf{x})} \omega_\alpha(\mathbf{x},\mathbf{y}) \leq 2^\tau(1+4\tau^2 k(\mathbf{x})^2)$$

*with a factor $k(\mathbf{x}) := \max\{1,t\}^{1-\alpha}$ which is relevant only if $t > 1$ and $\alpha \neq 1$.*

*Proof.* Let $\mathbf{y} \in V_\tau^\alpha(\mathbf{x})$ for some fixed $\mathbf{x} \in \mathbb{P}$ and $\tau \geq 0$. Then we can deduce from (2.26) that $\max\{t/u, u/t\} \leq 2^\tau$ and, with $d_\mathbb{S}(\eta,\theta)$ given as in (2.8),

$$d_\mathbb{S}(\eta,\theta) \leq \tau t^{1-\alpha}, \quad |\langle e_1, R_\eta(x-y)\rangle| \leq \tau t, \quad |\langle e_2, R_\eta(x-y)\rangle| \leq \tau t^\alpha.$$

Since $|\{\eta-\theta\}| \leq d_\mathbb{S}(\eta,\theta)$ and $t_0 = \max\{t,u\} \geq t$, this implies

$$t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2 \leq \tau^2.$$

Further, we get

$$\frac{t_0^{-2}|\langle e_\eta, x-y\rangle|^2}{1+t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2} \leq t_0^{-2}|\langle e_1, R_\eta(x-y)\rangle|^2 \leq \tau^2.$$

Finally, we deduce

$$|x-y|^2 = |\langle e_1, R_\eta(x-y)\rangle|^2 + |\langle e_2, R_\eta(x-y)\rangle|^2 \leq \tau^2(t^2+t^{2\alpha}).$$

If $t \leq 1$, whence $t \leq t^\alpha$, this leads to

$$t_0^{-2\alpha}|x-y|^2 \leq 2t_0^{-2\alpha}t^{2\alpha}\tau^2 \leq 2\tau^2.$$

In case $t > 1$, we have $t \geq t^\alpha$ and obtain

$$t_0^{-2\alpha}|x-y|^2 \leq 2t_0^{-2\alpha}t^2\tau^2 \leq 2\tau^2 t^{2(1-\alpha)}.$$

Altogether, our estimates prove

$$\omega_\alpha(\mathbf{x},\mathbf{y}) \leq 2^\tau(1+2\tau^2+2\tau^2 k(\mathbf{x})^2) \leq 2^\tau(1+4\tau^2 k(\mathbf{x})^2)$$

with $k(\mathbf{x}) = \max\{1,t\}^{1-\alpha}$. □

Note that in the isotropic case, when $\alpha = 1$, the estimate in Lemma 2.2.15 is uniform in $\mathbf{x} \in \mathbb{P}$ since $k(\mathbf{x}) = 1$ for all $\mathbf{x} \in \mathbb{P}$. Unfortunately, in the anisotropic case, the factor $k(\mathbf{x})$ is relevant and it is not possible to get rid of the dependence on $\mathbf{x}$. However, in Chapter 4, where we will merely work on the subdomain $\mathbb{X} = \mathbb{R}^2 \times \mathbb{T} \times (0, 1]$ of $\mathbb{P}$, the condition $t \le 1$ will always be satisfied and the factor becomes void.

Together, Lemma 2.2.14 and Lemma 2.2.15 yield the following proposition.

**Proposition 2.2.16.** *Let $\alpha \in [0, 1]$. For all $\mathbf{x} = (x, \eta, t) \in \mathbb{P}$ and for all $\tau \ge 0$ it holds*

$$B^{\alpha}_{\log_2(1+\tau^2)/4}(\mathbf{x}) \subseteq V^{\alpha}_{\tau}([\mathbf{x}]_{\mathfrak{p}}) \subseteq B^{\alpha}_{\tau+\log_2(1+4\tau^2 k(\mathbf{x})^2)}(\mathbf{x}).$$

*with $k(\mathbf{x}) = \max\{1, t\}^{1-\alpha}$.*

*Proof.* This follows from Lemma 2.2.14 and Lemma 2.2.15. $\square$

As a consequence of Proposition 2.2.16, the sets $V^{\alpha}_{\tau}([\mathbf{x}]_{\mathfrak{p}})$ constitute suitable substitutes for the corresponding $\alpha$-balls $B^{\alpha}_{\tau}(\mathbf{x})$ if $t$ is small. They have the advantage that due to their explicit form they are more easily accessible. A disadvantage of the sets $V^{\alpha}_{\tau}([\mathbf{x}]_{\mathfrak{p}})$ compared to $B^{\alpha}_{\tau}(\mathbf{x})$ is, however, that, unlike $\omega^{\log}_{\alpha}$, the generating distance function $\mathbf{w}_{\alpha}$ is not a quasi-metric in the sense of Definition 2.2.13. At least, we have the properties listed in Lemma 2.2.17 below, where (i)-(iii) can be interpreted as relaxations of the axioms $(Q1)$-$(Q3)$.

**Lemma 2.2.17.** *Let $\alpha \in [0, 1]$ and $\mathbf{x} = (x, \eta, t) \in \mathbb{P}$ be fixed. For $\tau \ge 0$ define the function $m_t(\tau) := 2^{\tau}(1 + \tau \max\{1, t\}^{2(1-\alpha)})$. For $\tau, \sigma \ge 0$ the following holds:*

*i)* $V^{\alpha}_0(\mathbf{x}) = \bigcap_{\tau>0} V^{\alpha}_{\tau}(\mathbf{x}) = \{\mathbf{x}\}$ *and* $V^{\alpha}_{\tau}(\mathbf{x}) \subset V^{\alpha}_{\sigma}(\mathbf{x})$ *if* $\tau < \sigma$.

*ii)* $\mathbf{y} \in V^{\alpha}_{\tau}(\mathbf{x}) \Rightarrow \mathbf{x} \in V^{\alpha}_{\tau m_t(\tau)}(\mathbf{y})$.

*iii)* $\mathbf{y} \in V^{\alpha}_{\tau}(\mathbf{x})$ *and* $\mathbf{z} \in V^{\alpha}_{\sigma}(\mathbf{y}) \Rightarrow \mathbf{z} \in V^{\alpha}_{f_t(\tau,\sigma)}(\mathbf{x})$ *with* $f_t(\tau, \sigma) := \tau + \sigma m_t(\tau)$.

*iv)* $\mathbf{y} \in V^{\alpha}_{\tau}(\mathbf{x})$ *and* $\mathbf{z} \in V^{\alpha}_{\sigma}(\mathbf{x}) \Rightarrow \mathbf{z} \in V^{\alpha}_{g_t(\tau,\sigma)}(\mathbf{y})$ *with* $g_t(\tau, \sigma) := (\tau + \sigma)m_t(\tau)$.

*v)* $\mathbf{x} \in V^{\alpha}_{\tau}(\mathbf{y}) \cap V^{\alpha}_{\sigma}(\mathbf{z}) \Rightarrow \mathbf{z} \in V^{\alpha}_{h_{u,v}(\tau,\sigma)}(\mathbf{y})$ *with* $h_{u,v}(\tau, \sigma) = \tau + \sigma m_u(\tau)m_v(\sigma)$.

*Proof.* ad (i): Clear.
ad (ii): If $\mathbf{y} \in V^{\alpha}_{\tau}(\mathbf{x})$ we have

$$u/t \in J^{\tau}\,, \quad d_{\mathbb{S}}(\theta, \eta) \le t^{1-\alpha}\tau\,, \quad y - x \in R^{-1}_{\eta} A_{\alpha,t} Q^{\tau}. \tag{2.27}$$

This implies

$$t/u \in J^{\tau}\,, \quad d_{\mathbb{S}}(\eta, \theta) \le u^{1-\alpha}\tilde{\tau}\,, \quad x - y \in R^{-1}_{\theta} A_{\alpha,u} T Q^{\tau},$$

where $\tilde{\tau} := (t/u)^{1-\alpha}\tau$ and $T := T(\theta, u, \eta, t)$ is the 'transfer matrix' given by

$$T(\theta, u, \eta, t) := A^{-1}_{\alpha,u} R_{\theta} R^{-1}_{\eta} A_{\alpha,t}.$$

Since $\tilde{\tau} \le 2^{\tau}\tau \le \tau m_t(\tau)$ and $\|T(\theta, u, \eta, t)\|_{\infty\to\infty} = \|T(\eta, t, \theta, u)^{-1}\|_{\infty\to\infty} \le m_t(\tau)$ according to Lemma 2.2.18 we obtain $\mathbf{x} \in V^{\alpha}_{\tau m_t(\tau)}(\mathbf{y})$.

ad (iii): In addition to the assumptions (2.27), we now also have

$$v/u \in J^{\sigma}\,, \quad d_{\mathbb{S}}(\kappa,\theta) \le u^{1-\alpha}\sigma\,, \quad z-y \in R_{\theta}^{-1}A_{\alpha,u}Q^{\sigma}.$$

We deduce

$$v/t \in J^{\tau+\sigma}\,, \quad d_{\mathbb{S}}(\kappa,\eta) \le t^{1-\alpha}\tilde{\tau}\,, \quad z-x \in R_{\eta}^{-1}A_{\alpha,t}(Q^{\tau}+TQ^{\sigma})$$

with $\tilde{\tau} = \tau + (u/t)^{1-\alpha}\sigma$ and the transfer matrix $T := T(\eta,t,\theta,u) = A_{\alpha,t}^{-1}R_{\eta}R_{\theta}^{-1}A_{\alpha,u}$. Due to Lemma 2.2.18 we have $\|T(\eta,t,\theta,u)\|_{\infty\to\infty} \le m_t(\tau)$. Further $\tilde{\tau} \le \tau + 2^{\tau}\sigma \le \tau + \sigma m_t(\tau)$.
ad (iv): In addition to the assumptions (2.27), we now also have

$$v/t \in J^{\sigma}\,, \quad d_{\mathbb{S}}(\kappa,\eta) \le t^{1-\alpha}\sigma\,, \quad z-x \in R_{\eta}^{-1}A_{\alpha,t}Q^{\sigma}.$$

Hence, we get with $\tilde{\tau} := \tau + \sigma$

$$v/u \in J^{\tilde{\tau}}\,, \quad d_{\mathbb{S}}(\kappa,\theta) \le t^{1-\alpha}\tilde{\tau}\,, \quad z-y \in R_{\theta}^{-1}A_{\alpha,u}TQ^{\tilde{\tau}},$$

where $T := T(\theta,u,\eta,t) = A_{\alpha,u}^{-1}R_{\theta}R_{\eta}^{-1}A_{\alpha,t}$ is the same transfer matrix as in (ii). We already know that $\|T(\theta,u,\eta,t)\|_{\infty\to\infty} \le m_t(\tau)$ due to Lemma 2.2.18, further $\tilde{\tau} \le (\tau+\sigma)m_t(\tau)$.
ad (v): Here, we now have the assumptions $t/u \in J^{\tau}$, $t/v \in J^{\sigma}$, $d_{\mathbb{S}}(\eta,\theta) \le u^{1-\alpha}\tau$, $d_{\mathbb{S}}(\eta,\kappa) \le v^{1-\alpha}\sigma$, and $x-y \in R_{\theta}^{-1}A_{\alpha,u}Q^{\tau}$, $x-z \in R_{\kappa}^{-1}A_{\alpha,v}Q^{\sigma}$. We deduce

$$v/u \in J^{\tau+\sigma}\,, \quad d_{\mathbb{S}}(\kappa,\theta) \le u^{1-\alpha}\tilde{\tau}\,, \quad z-y \in R_{\theta}^{-1}A_{\alpha,u}\Big(Q^{\tau}+TQ^{\sigma}\Big),$$

where $\tilde{\tau} := \tau + (v/u)^{1-\alpha}\sigma$ and

$$T := A_{\alpha,u}^{-1}R_{\theta}R_{\kappa}^{-1}A_{\alpha,v} = \big(A_{\alpha,u}^{-1}R_{\theta}R_{\eta}^{-1}A_{\alpha,t}\big)\big(A_{\alpha,t}^{-1}R_{\eta}R_{\kappa}^{-1}A_{\alpha,v}\big) = T(\theta,u,\eta,t)T(\eta,t,\kappa,v).$$

Due to Lemma 2.2.18, we have $\|T(\theta,u,\eta,t)\|_{\infty\to\infty} \le m_u(\tau)$ and $\|T(\eta,t,\kappa,v)\|_{\infty\to\infty} = \|T(\kappa,v,\eta,t)^{-1}\|_{\infty\to\infty} \le m_v(\sigma)$. □

In the proof of Lemma 2.2.17 matrices of the form $T = T(\eta,t,\theta,u)$ play an essential role. Such matrices will also be important later in Subsection 4.3.4. The entries of such matrices are investigated in the following lemma.

**Lemma 2.2.18.** *Consider the matrix* $T := T(\eta,t,\theta,u) := A_{\alpha,t}^{-1}R_{\eta}R_{\theta}^{-1}A_{\alpha,u}$ *with* $\eta,\theta \in \mathbb{T}$, $t,u \in \mathbb{R}_+$, *and let*

$$T =: \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix} \quad \text{and} \quad T^{-1} =: \begin{pmatrix} \tilde{t}_{11} & \tilde{t}_{12} \\ \tilde{t}_{21} & \tilde{t}_{22} \end{pmatrix}.$$

*Let us further assume that for some* $\tau \ge 0$ *we have*

$$\max\{t/u, u/t\} \le 2^{\tau} \quad , \quad d_{\mathbb{S}}(\eta,\theta) \le t^{1-\alpha}\tau.$$

*Then the entries of* $T$ *and* $T^{-1}$ *are uniformly bounded as follows*

$$\begin{array}{cccc} |t_{11}| \le 2^{\tau} \quad , & |t_{12}| \le 2^{\tau}\tau \quad , & |t_{21}| \le 2^{\tau}\tau t^{2(1-\alpha)} \quad , & |t_{22}| \le 2^{\tau}, \\ |\tilde{t}_{11}| \le 2^{\tau} \quad , & |\tilde{t}_{12}| \le 2^{\tau}\tau \quad , & |\tilde{t}_{21}| \le 2^{\tau}\tau t^{2(1-\alpha)} \quad , & |\tilde{t}_{22}| \le 2^{\tau}. \end{array}$$

*Further, writing* $m_t(\tau) := 2^{\tau}(1+\tau\max\{1,t\}^{2(1-\alpha)})$, *this leads to the estimates*

$$\|T\|_{\infty\to\infty}\,, \quad \|T^{-1}\|_{\infty\to\infty}\,, \quad \|T^{T}\|_{\infty\to\infty}\,, \quad \|T^{-T}\|_{\infty\to\infty} \le m_t(\tau).$$

*Proof.* The matrix $T$ has the form

$$T = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix} = \begin{pmatrix} ut^{-1}\cos(\eta-\theta) & -u^{\alpha}t^{-1}\sin(\eta-\theta) \\ ut^{-\alpha}\sin(\eta-\theta) & u^{\alpha}t^{-\alpha}\cos(\eta-\theta) \end{pmatrix}.$$

To estimate the entries, we use $|\sin(\eta-\theta)| \le \min\{1, d_{\mathbb{S}}(\eta,\theta)\}$ and $|\cos(\eta-\theta)| \le 1$. We obtain

$$\begin{aligned} |t_{11}| &= |(u/t)\cos(\eta-\theta)| \le u/t \le \max\{t/u, u/t\}, \\ |t_{22}| &= |(u/t)^{\alpha}\cos(\eta-\theta)| \le (u/t)^{\alpha} \le \max\{t/u, u/t\}^{\alpha}. \end{aligned}$$

Further,

$$\begin{aligned} |t_{12}| = |(u^{\alpha}/t)\sin(\eta-\theta)| &\le (u^{\alpha}/t)d_{\mathbb{S}}(\eta,\theta) \le (u^{\alpha}/t)\max\{t,u\}^{1-\alpha}\tau \\ &= (u^{\alpha}/t)t^{1-\alpha}\max\{u/t,1\}^{1-\alpha}\tau = (u/t)^{\alpha}\max\{u/t,1\}^{1-\alpha}\tau \\ &\le \max\{t/u, u/t\}\tau. \end{aligned}$$

We also get

$$\begin{aligned} |t_{21}| = |(u/t^{\alpha})\sin(\eta-\theta)| &\le (u/t^{\alpha})d_{\mathbb{S}}(\eta,\theta) \le (u/t^{\alpha})t^{1-\alpha}\tau \\ &= (u/t)^{\alpha}u^{1-\alpha}t^{1-\alpha}\tau = t^{2(1-\alpha)}(u/t)^{1-\alpha}(u/t)^{\alpha}\tau \\ &\le \max\{t/u, u/t\}t^{2(1-\alpha)}\tau. \end{aligned}$$

To obtain the results for $T^{-1}$, we use that $T(\eta,t,\theta,u)^{-1} = T(\theta,u,\eta,t)$. The estimates of the entries $|\tilde{t}_{11}|$, $|\tilde{t}_{22}|$, and $|\tilde{t}_{12}|$ are then analogous to above. For $|\tilde{t}_{21}|$ we get

$$\begin{aligned} |\tilde{t}_{21}| = |(t/u^{\alpha})\sin(\theta-\eta)| &\le (t/u^{\alpha})d_{\mathbb{S}}(\theta,\eta) \le (t/u^{\alpha})t^{1-\alpha}\tau \\ &= (t/u)^{\alpha}t^{2(1-\alpha)}\tau \le \max\{t/u, u/t\}^{\alpha}t^{2(1-\alpha)}\tau. \end{aligned}$$

We obtain estimates for the row-sum and column-sum norms of $T$ and $T^{-1}$ directly from these estimates of the entries. This establishes the result. ☐

At last, we arrive at the following corollary which will be useful in the proof of Lemma 4.3.9.

**Corollary 2.2.19.** *Consider the matrix $T := T(\eta,t,\theta,u) := A_{\alpha,t}^{-1}R_{\eta}R_{\theta}^{-1}A_{\alpha,u}$ with $\eta,\theta \in \mathbb{T}$, $t,u \in \mathbb{R}_+$. Assume that for some $\tau \ge 0$ we have*

$$\max\{t/u, u/t\} \le 2^{\tau} \quad , \quad d_{\mathbb{S}}(\eta,\theta) \le \min\{t, t^{-1}\}^{1-\alpha}\tau.$$

*Then we have the estimates*

$$\|T\|_{\infty\to\infty}\,, \quad \|T^{-1}\|_{\infty\to\infty}\,, \quad \|T^{T}\|_{\infty\to\infty}\,, \quad \|T^{-T}\|_{\infty\to\infty} \le 2^{\tau}(1+\tau).$$

*Proof.* If $t \le 1$ the statement is a direct consequence of Lemma 2.2.18. In case $t > 1$, we define $\tilde{t} := t^{-1}$ and $\tilde{u} := u^{-1}$ and apply Lemma 2.2.18 to the matrix $T(\eta,\tilde{t},\theta,\tilde{u})$. Since $T(\eta,t,\theta,u) = T(\eta,\tilde{t},\theta,\tilde{u})^{-T}$ the assertion follows. ☐

### 2.2.4 Stability of $\omega_\alpha$

An important property of the distance function $\omega_\alpha$ is its stability with respect to small perturbations of the indices $\mathbf{x}, \mathbf{y} \in \mathbb{P}$. This is a direct consequence of the triangle-inequality in (2.16). Since we want to formulate the stability with respect to the sets $V_\tau^\alpha(\mathbf{x})$ from (2.26), we first need another auxiliary result which complements Lemma 2.2.15.

**Lemma 2.2.20.** *Let $\alpha \in [0,1]$. For all $\mathbf{x} \in \mathbb{P}$ and $\tau \geq 0$*

$$1 \leq \sup_{\mathbf{y} \in V_\tau^\alpha(\mathbf{x})} \omega_\alpha(\mathbf{y}, \mathbf{x}) \leq 2^\tau (1 + 9\tau^2 k(\mathbf{x})^2),$$

*where the factor $k(\mathbf{x}) = \max\{1, t\}^{1-\alpha}$ is relevant only if $t > 1$ and $\alpha \neq 1$.*

*Proof.* For $\mathbf{y} \in V_\tau^\alpha(\mathbf{x})$, as in the proof of Lemma 2.2.15, we have $\max\{t/u, u/t\} \leq 2^\tau$ and

$$|\{\eta - \theta\}| \leq \tau t^{1-\alpha}, \quad |\langle e_1, R_\eta(x-y)\rangle| \leq \tau t, \quad |\langle e_2, R_\eta(x-y)\rangle| \leq \tau t^\alpha.$$

Further, we observe that the terms we need to estimate coincide with those in the proof of Lemma 2.2.15, with the exception of the following term, which is estimated as in (2.19),

$$\frac{t_0^{-2}|\langle e_\theta, x-y\rangle|^2}{1 + t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2} \leq 2\left(\frac{t_0^{-2}|\langle e_\eta, x-y\rangle|^2}{1 + t_0^{-2(1-\alpha)}|\{\eta-\theta\}|^2} + t_0^{-2\alpha}|x-y|^2\right) \leq 2\tau^2 + 4\tau^2 k(\mathbf{x})^2.$$

We obtain

$$\omega_\alpha(\mathbf{x}, \mathbf{y}) \leq 2^\tau(1 + 3\tau^2 + 6\tau^2 k(\mathbf{x})^2) \leq 2^\tau(1 + 9\tau^2 k(\mathbf{x})^2).$$

□

Now we can prove that $\omega_\alpha(\mathbf{x}, \mathbf{y})$ is stable with respect to perturbations in both arguments.

**Proposition 2.2.21.** *Let $\alpha \in [0,1]$ and let $C_T \geq 1$ denote the constant from (2.16). For $\tau \geq 0$ and $\mathbf{x}, \mathbf{y} \in \mathbb{P}$ we have*

$$\begin{aligned}
C_T^{-1} 2^{-\tau}(1 + 9\tau^2 k(\mathbf{y})^2)^{-1}\omega_\alpha(\mathbf{x}, \mathbf{y}) &\leq \inf_{\mathbf{z} \in V_\tau^\alpha(\mathbf{y})} \omega_\alpha(\mathbf{x}, \mathbf{z}) \\
&\leq \sup_{\mathbf{z} \in V_\tau^\alpha(\mathbf{y})} \omega_\alpha(\mathbf{x}, \mathbf{z}) \leq C_T 2^\tau (1 + 4\tau^2 k(\mathbf{y})^2)\omega_\alpha(\mathbf{x}, \mathbf{y}), \\
C_T^{-1} 2^{-\tau}(1 + 4\tau^2 k(\mathbf{x})^2)^{-1}\omega_\alpha(\mathbf{x}, \mathbf{y}) &\leq \inf_{\mathbf{z} \in V_\tau^\alpha(\mathbf{x})} \omega_\alpha(\mathbf{z}, \mathbf{y}) \\
&\leq \sup_{\mathbf{z} \in V_\tau^\alpha(\mathbf{x})} \omega_\alpha(\mathbf{z}, \mathbf{y}) \leq C_T 2^\tau (1 + 9\tau^2 k(\mathbf{x})^2)\omega_\alpha(\mathbf{x}, \mathbf{y}),
\end{aligned}$$

*where $k(\mathbf{x}) = \max\{1, t\}^{1-\alpha}$ and $k(\mathbf{y}) = \max\{1, u\}^{1-\alpha}$.*

*Proof.* Let $\mathbf{x} = (x, \eta, t) \in \mathbb{X}$ and $\mathbf{z} \in (z, \kappa, v) \in \mathbb{X}$. As a consequence of the triangle inequality in (2.16), for every $\mathbf{y} = (y, \theta, u) \in \mathbb{X}$

$$\omega_\alpha(\mathbf{x}, \mathbf{z}) \leq C_T \omega_\alpha(\mathbf{x}, \mathbf{y})\omega_\alpha(\mathbf{y}, \mathbf{z}).$$

Together with Lemma 2.2.15, this yields

$$\sup_{\mathbf{z}\in V_\tau^\alpha(\mathbf{y})} \omega_\alpha(\mathbf{x},\mathbf{z}) \le C_T\omega_\alpha(\mathbf{x},\mathbf{y}) \sup_{\mathbf{z}\in V_\tau^\alpha(\mathbf{y})} \omega_\alpha(\mathbf{y},\mathbf{z}) \le C_T 2^\tau (1+4\tau^2 k(\mathbf{y})^2)\omega_\alpha(\mathbf{x},\mathbf{y}).$$

The triangle inequality also yields $\omega_\alpha(\mathbf{x},\mathbf{y}) \le C_T\omega_\alpha(\mathbf{x},\mathbf{z})\omega_\alpha(\mathbf{z},\mathbf{y})$ which implies

$$C_T^{-1}\omega_\alpha(\mathbf{z},\mathbf{y})^{-1}\omega_\alpha(\mathbf{x},\mathbf{y}) \le \omega_\alpha(\mathbf{x},\mathbf{z}).$$

We deduce with Lemma 2.2.20

$$\begin{aligned}\inf_{\mathbf{z}\in V_\tau^\alpha(\mathbf{y})} \omega_\alpha(\mathbf{x},\mathbf{z}) &\ge C_T^{-1} \inf_{\mathbf{z}\in V_\tau^\alpha(\mathbf{y})} (\omega_\alpha(\mathbf{z},\mathbf{y})^{-1})\omega_\alpha(\mathbf{x},\mathbf{y}) \\ &= C_T^{-1}\Big(\sup_{\mathbf{z}\in V_\tau^\alpha(\mathbf{y})} \omega_\alpha(\mathbf{z},\mathbf{y})\Big)^{-1}\omega_\alpha(\mathbf{x},\mathbf{y}) \ge C_T^{-1}2^{-\tau}(1+9\tau^2 k(\mathbf{y})^2)^{-1}\omega_\alpha(\mathbf{x},\mathbf{y}).\end{aligned}$$

Similarly, the second inequality is shown. □

We can draw the following conclusion, which will be relevant in Chapter 4.

**Corollary 2.2.22.** *Let $C>0$ and $\tau \ge 0$ be fixed. Then we have*

$$\sup_{\mathbf{a}\in V_\tau^\alpha(\mathbf{x})} \sup_{\mathbf{b}\in V_\tau^\alpha(\mathbf{y})} \omega_\alpha(\mathbf{a},\mathbf{b}) \lesssim \omega_\alpha(\mathbf{x},\mathbf{y}) \lesssim \inf_{\mathbf{a}\in V_\tau^\alpha(\mathbf{x})} \inf_{\mathbf{b}\in V_\tau^\alpha(\mathbf{y})} \omega_\alpha(\mathbf{a},\mathbf{b})$$

*uniformly for all $\mathbf{x}=(x,\eta,t)\in\mathbb{P}$ and $\mathbf{y}=(y,\theta,u)\in\mathbb{P}$ with $t,u\le C$.*

## 2.3 Transfer Principle for Discrete $\alpha$-Molecule Frames

In this section, we take a first step towards the analysis of approximation properties of $\alpha$-molecule systems. Our main result, Theorem 2.3.6 ([59, Thm. 5.6]), will lay the foundation for a systematic comparison of their approximation performance. It is referred to as the *transfer principle* and will later be used in Chapters 5 and 6 to derive approximation rates of $\alpha$-molecules for cartoon-like data.

Our subsequent considerations are restricted to discrete systems of $\alpha$-molecules $\{m_\lambda\}_{\lambda\in\Lambda}$, i.e., systems with a countable index set $\Lambda$. Further, we require $\{m_\lambda\}_{\lambda\in\Lambda}$ to be a frame, a notion recalled in the first subsection below. In the context of frame approximation, the concept of *sparsity equivalence* from [62, Def. 4.2] is a useful tool. It allows to put the approximation properties of different frame systems into relation to each other.

As we will see, for frames of $\alpha$-molecules an analysis of their sparsity equivalence is possible based solely on the order and the parametrizations of the respective systems. According to the transfer principle, a sufficiently high order already implies sparsity equivalence if the parametrizations are consistent in a suitable sense.

### 2.3.1 Frame Approximation and Sparsity Equivalence

Before turning to $\alpha$-molecules, let us briefly recall some aspects of approximation theory in an abstract Hilbert space $\mathcal{H}$ with associated scalar product $\langle\cdot,\cdot\rangle$. Thereby we assume $\mathcal{H}$ to be separable, which in view of our later application is no restriction. In this setting, we now first discuss the question of suitable representations of signals $f\in\mathcal{H}$.

A standard way is to use the so-called *analysis coefficients* $\{\langle f, m_\lambda \rangle\}_{\lambda\in\Lambda}$ with respect to some fixed *dictionary* $\mathfrak{M} := \{m_\lambda\}_{\lambda\in\Lambda}$, where the index set $\Lambda$ can be assumed to be countable due to the separability of $\mathcal{H}$. In order to obtain a faithful representation for all signals $f \in \mathcal{H}$, the *analysis operator* of the dictionary, i.e., the mapping $f \mapsto \{\langle f, m_\lambda \rangle\}_{\lambda\in\Lambda}$, needs to be injective. This is the case precisely if $\overline{\operatorname{span}}\{m_\lambda\}_{\lambda\in\Lambda} = \mathcal{H}$, which can therefore be considered as a minimal assumption for a suitable dictionary $\mathfrak{M}$.

A particular useful class of dictionaries are so-called *frame systems* (see e.g. [20]). These are systems, where the analysis operator is an injective linear map from $\mathcal{H}$ to $\ell^2(\Lambda)$ and where in addition it is bounded from above and below. Frames thus not only ensure faithful representation of the signals but also a stable measurement of the coefficients and a stable reconstruction. They are characterized by the property that there exist constants $0 < A \leq B < \infty$, called the *frame bounds*, such that

$$A\|f\|^2 \leq \sum_{\lambda\in\Lambda} |\langle f, m_\lambda \rangle|^2 \leq B\|f\|^2 \text{ for all } f \in \mathcal{H}.$$

If $A$ and $B$ can be chosen equal, the frame is called *tight*. In case $A = B = 1$, one speaks of a *Parseval frame*.

Frame systems also naturally lend themselves for the synthesis of signals, since for every sequence $\{c_\lambda\}_\lambda \in \ell^2(\Lambda)$ the sum

$$f = \sum_{\lambda\in\Lambda} c_\lambda m_\lambda, \tag{2.28}$$

converges unconditionally in $\mathcal{H}$. The associated operator from $\ell^2(\Lambda)$ to $\mathcal{H}$ is surjective and called the *synthesis operator* of the frame. It allows to alternatively use the so-called *synthesis coefficients* $\{c_\lambda\}_\lambda$ in the expansion (2.28) for the representation of $f$. This sequence however is usually not unique since in general the synthesis operator is not injective. Unlike a basis, a frame allows for a certain redundancy of its elements.

The composition of the synthesis operator and the analysis operator is called the *frame operator*. It is an isomorphism $S : \mathcal{H} \to \mathcal{H}$ and given explicitly by $Sf = \sum_{\lambda\in\Lambda} \langle f, m_\lambda \rangle m_\lambda$. It can be used to compute the so-called *canonical dual frame* $\{\tilde{m}_\lambda\}_{\lambda\in\Lambda}$ of $\{m_\lambda\}_{\lambda\in\Lambda}$ defined by $\tilde{m}_\lambda := S^{-1} m_\lambda$. We then have the following decomposition and reconstruction formulas

$$f = \sum_{\lambda\in\Lambda} \langle f, \tilde{m}_\lambda \rangle m_\lambda = \sum_{\lambda\in\Lambda} \langle f, m_\lambda \rangle \tilde{m}_\lambda.$$

In general, any frame $\{\tilde{m}_\lambda\}_{\lambda\in\Lambda}$ satisfying these formulas is called an associated *dual frame* of $\{m_\lambda\}_{\lambda\in\Lambda}$. The synthesis coefficients $\{c_\lambda\}_\lambda$ given by

$$c_\lambda = \langle f, S^{-1} m_\lambda \rangle, \quad \lambda \in \Lambda.$$

are called the *canonical frame coefficients*. They have the distinct property that they minimize the $\ell^2$-norm among all possible synthesis coefficient sequences.

For practical applications, as important as the question of faithful representation of a signal $f \in \mathcal{H}$ is the question of good approximation. Thereby one usually restricts to finite expansions in (2.28). This motivates the following definition. Given some arbitrary dictionary $\mathfrak{M} := \{m_\lambda\}_\lambda$, the associated, possibly non-linear, space of $N$-term expansions is denoted by $\Sigma_N := \Sigma_N[\mathfrak{M}]$ and consists of all linear combinations

$$\sum_{\lambda\in\Lambda_N} c_\lambda m_\lambda \quad \text{with } \Lambda_N \subseteq \Lambda, \quad \#\Lambda_N \leq N. \tag{2.29}$$

The *error of $N$-term approximation* of $f$ with respect to $\mathfrak{M}$ is defined by

$$\sigma_N(f) := \inf_{g\in\Sigma_N} \|f - g\|. \tag{2.30}$$

For an efficient approximation of $f$, it is desirable to find dictionaries which provide good sparse approximations, in the sense that the error $\sigma_N(f)$ decays quickly for $N \to \infty$.

Sometimes there exist vectors $f_N \in \Sigma_N$ which minimize the $N$-term approximation error, i.e., for which $\|f - f_N\| = \sigma_N(f)$ holds true. Such vectors $f_N$ are called *best $N$-term approximations* of $f$ with respect to $\mathfrak{M}$. They are given by

$$f_N = \operatorname*{arg\,min}_{g=\sum_{\lambda\in\Lambda_N} c_\lambda m_\lambda} \|f - g\| \quad \text{s.t.} \quad \Lambda_N \subseteq \Lambda, \quad \#\Lambda_N \leq N. \tag{2.31}$$

For general dictionaries, however, their existence is not guaranteed and usually hinges on additional assumptions, such as for example a polynomial depth search constraint as discussed in Section 5.1.

Even in the frame setting, best $N$-term approximations need not exist and, if they exist, their computation is not yet well-understood. The delicacy of this problem can for instance be seen in [52]. A typical approach to circumvent this problem is to consider not the best $N$-term approximation of a frame but the $N$-term approximation $f_N$ obtained by keeping the $N$ largest coefficients. This type of approximation is better understood and, due to $\sigma_N(f) \leq \|f - f_N\|$, also provides a bound for the best $N$-term approximation error.

The achievable $N$-term approximation rate can thus be estimated by the decay of $\|f - f_N\|$ as $N \to \infty$, which in turn depends on the decay of the corresponding frame coefficients. We subsequently quantify the decay of a sequence $\{c_\lambda\}_\lambda$ by its weak $\ell^p$-quasi-norm. For $p > 0$ this is the quantity

$$\|\{c_\lambda\}_\lambda\|_{\omega\ell^p} := \Big(\sup_{\varepsilon>0} \varepsilon^p \cdot \#\{\lambda : |c_\lambda| > \varepsilon\}\Big)^{1/p}.$$

The associated sequence space is denoted by $\omega\ell^p(\Lambda)$ and consists of all sequences $\{c_\lambda\}_\lambda$ with $\|\{c_\lambda\}_\lambda\|_{\omega\ell^p} < \infty$. Note that, together with $\|\cdot\|_{\omega\ell^p}$, this space is a quasi-normed space.

This space can also be characterized as the space of null sequences $\{c_\lambda\}_\lambda$ which possess a non-increasing rearrangement $(c_n^*)_{n\in\mathbb{N}}$ such that $\sup_{n>0} n^{1/p}|c_n^*| < \infty$. For the sequences in $\omega\ell^p(\Lambda)$, we even have the equality

$$\sup_{n>0} n^{1/p}|c_n^*| = \|\{c_\lambda\}_\lambda\|_{\omega\ell^p}.$$

The result below, whose proof can be found e.g. in [83, Sec. 3.2] or [59, Lem. 5.1], relates the decay of the synthesis coefficients of a frame to the $N$-term approximation rate.

**Lemma 2.3.1.** *Let $\{m_\lambda\}_{\lambda\in\Lambda}$ be a frame in $\mathcal{H}$ and $f = \sum c_\lambda m_\lambda$ an expansion of $f \in \mathcal{H}$ with respect to this frame. If $\{c_\lambda\}_\lambda \in \omega\ell^{2/(p+1)}(\Lambda)$ for some $p > 0$, then the $N$-term approximation rate for $f$ achieved by keeping the $N$ largest coefficients is at least of order $N^{-p/2}$, i.e.*

$$\|f - f_N\|_2^2 \lesssim N^{-p}.$$

According to Lemma 2.3.1, a fast decay of the frame coefficients implies good $N$-term approximation rates. Concretely, if the synthesis coefficients of $\{m_\lambda\}_{\lambda\in\Lambda}$ satisfy $\{c_\lambda\}_\lambda \in \omega\ell^p(\Lambda)$ for $p < 2$ an $N$-term approximation rate at least of order $N^{-(1/p-1/2)}$ is achieved. Conversely, if the sequence $\{\langle f, m_\lambda\rangle\}_{\lambda\in\Lambda}$ of analysis coefficients lies in $\omega\ell^p(\Lambda)$ , the best approximation rate of any dual frame $\{\tilde{m}_\lambda\}_{\lambda\in\Lambda}$ is at least of this order. In terms of signal compression this is exactly what one hopes for: From simply keeping the $N$ largest frame coefficients, which can be encoded by order $N$ bits, we can reconstruct the original signal $f$ up to a precision of order $N^{-(1/p-1/2)}$.

### Sparsity Equivalence

Let us now compare the approximation performance of different frame systems in $\mathcal{H}$. In view of Lemma 2.3.1, it makes sense to analyze the decay of the corresponding coefficient sequences. For a signal $f \in \mathcal{H}$ and two frames $\{m_\lambda\}_{\lambda\in\Lambda}$ and $\{p_\mu\}_{\mu\in\Delta}$, let $\{c_\lambda\}_\lambda$ be a sequence of synthesis coefficients for $f$ with respect to $\{m_\lambda\}_{\lambda\in\Lambda}$. Then, formally, the analysis coefficients with respect to $\{p_\mu\}_{\mu\in\Delta}$ can be calculated by

$$\langle f, p_\mu\rangle = \Big\langle \sum_{\lambda\in\Lambda} c_\lambda m_\lambda, p_\mu \Big\rangle = \sum_{\lambda\in\Lambda} c_\lambda \langle m_\lambda, p_\mu\rangle . \tag{2.32}$$

Hence, they are obtained by a multiplication of $\{c_\lambda\}_\lambda$ with the cross-Gramian of the two systems. This observation leads to the following result from [59].

**Proposition 2.3.2** ([59, Prop. 5.2])**.** *Let $0 < p < 2$, and let $\{m_\lambda\}_{\lambda\in\Lambda}$ and $\{p_\mu\}_{\mu\in\Delta}$ be two discrete frames in a Hilbert space $\mathcal{H}$ such that*

$$\Big\| \{\langle m_\lambda, p_\mu\rangle\}_{\lambda\in\Lambda,\mu\in\Delta} \Big\|_{\ell^p\to\ell^p} < \infty.$$

*Then for every signal $f \in \mathcal{H}$ the membership $\{\langle f, \tilde{m}_\lambda\rangle\}_\lambda \in \ell^p(\Lambda)$, where $\{\tilde{m}_\lambda\}_{\lambda\in\Lambda}$ denotes a dual frame of $\{m_\lambda\}_{\lambda\in\Lambda}$, implies $\{\langle f, p_\mu\rangle\}_\mu \in \ell^p(\Delta)$. In particular, $f$ can be encoded by the $N$ largest frame coefficients from $\{\langle f, p_\mu\rangle\}_\mu$ up to accuracy $\lesssim N^{-(1/p-1/2)}$.*

*Proof.* Define $c_\lambda := \langle f, \tilde{m}_\lambda\rangle$ for $\lambda \in \Lambda$. Then $\{c_\lambda\}_\lambda$ is a sequence of synthesis coefficients for $f$ and by assumption $\{c_\lambda\}_\lambda \in \ell^p(\Lambda)$ with $0 < p < 2$. Due to (2.32) and $\| \{\langle m_\lambda, p_\mu\rangle\}_{\lambda,\mu} \|_{\ell^p\to\ell^p} < \infty$, this implies $\{\langle f, p_\mu\rangle\}_\mu \in \ell^p(\Delta)$. □

Proposition 2.3.2 motivates the following notion of *sparsity equivalence* ([59, Def. 5.3]), initially introduced in [62, Def. 4.2] for parabolic molecules. The intuition behind this concept is that sparsity equivalent frames should provide frame coefficients with a similar decay. We remark however that, contrary to what the name suggests, this notion does not provide an equivalence relation.

**Definition 2.3.3** ([59, Def. 5.3])**.** Two discrete frames $\{m_\lambda\}_{\lambda\in\Lambda}$ and $\{p_\mu\}_{\mu\in\Delta}$ in a Hilbert space $\mathcal{H}$ are called *sparsity equivalent in $\ell^p$*, $0 < p < 2$, if

$$\Big\| \{\langle m_\lambda, p_\mu\rangle\}_{\lambda\in\Lambda,\mu\in\Delta} \Big\|_{\ell^p\to\ell^p} < \infty.$$

A useful tool for the verification of sparsity equivalence is Schur's test, which yields a simple estimate for the operator norm of matrices acting on discrete $\ell^p$ spaces. The version below can be found in [62].

**Lemma 2.3.4** ([62, Lem. 4.4])**.** *Let $\Lambda$ and $\Delta$ be countable index sets, and assume that $\mathbf{A} := \{A_{\lambda,\mu}\}_{\lambda\in\Lambda,\,\mu\in\Delta}$ is a (possibly infinite-dimensional) matrix with entries $A_{\lambda,\mu} \in \mathbb{C}$ such that*

$$\sup_{\lambda\in\Lambda} \sum_{\mu\in\Delta} |A_{\lambda,\mu}|^q < \infty \quad \textit{and} \quad \sup_{\mu\in\Delta} \sum_{\lambda\in\Lambda} |A_{\lambda,\mu}|^q < \infty,$$

*where $p > 0$ and $q := \min\{1, p\}$. Then $\mathbf{A} : \ell^p(\Lambda) \to \ell^p(\Delta)$ is a bounded linear operator with the bound*

$$\|\mathbf{A}\|_{\ell^p(\Lambda)\to\ell^p(\Delta)} \leq \max\Big\{ \sup_{\lambda\in\Lambda} \sum_{\mu\in\Delta} |A_{\lambda,\mu}|^q, \sup_{\mu\in\Delta} \sum_{\lambda\in\Lambda} |A_{\lambda,\mu}|^q \Big\}^{1/q}.$$

*Proof.* The proof for $p < 1$ follows easily from the fact that $|x+y|^p \leq |x|^p + |y|^p$ for $x, y \in \mathbb{R}$. To show the case $p \geq 1$ one proves the assertion for $p = 1$ and $p = \infty$. The claim then follows by interpolation. □

With this result, our excursion into abstract Hilbert space theory ends and we turn back to the topic of $\alpha$-molecules.

### 2.3.2 Transfer Principle and Consistency of Parametrizations

Let us now investigate the concept of sparsity equivalence in the realm of discrete $\alpha$-molecule frames in $L_2(\mathbb{R}^2)$. Our main result, Theorem 2.3.6, will provide a sufficient condition ensuring sparsity equivalence for such frames. The condition depends on the one hand on the respective orders of the systems, on the other hand, the respective parametrizations play a role.

In view of Schur's test, i.e., Lemma 2.3.4 above, and the estimate of the cross-Gramian in Theorem 2.2.2 the following notion of $(\alpha, k)$-consistency is reasonable.

**Definition 2.3.5** ([59, Def. 5.5])**.** Let $\alpha \in [0, 1]$ and $k > 0$. Two parametrizations $(\Lambda, \Phi_\Lambda)$ and $(\Delta, \Phi_\Delta)$ with countable index sets $\Lambda$ and $\Delta$ are called *$(\alpha, k)$-consistent*, if

$$\sup_{\lambda\in\Lambda} \sum_{\mu\in\Delta} \omega_\alpha\big(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu)\big)^{-k} < \infty \quad \text{and} \quad \sup_{\mu\in\Delta} \sum_{\lambda\in\Lambda} \omega_\alpha\big(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu)\big)^{-k} < \infty.$$

Note that $(\alpha, k)$-consistency implies $(\alpha, k')$-consistency for $k' \geq k$, due to $\omega_\alpha \geq 1$. Using this notion, we can now formulate a convenient sufficient condition for the sparsity equivalence of discrete $\alpha$-molecule frames.

**Theorem 2.3.6** ([59, Thm. 5.6])**.** *Let $\alpha \in [0, 1]$, $0 < p \leq 1$, and $k > 0$. Let $\{m_\lambda\}_{\lambda\in\Lambda}$ and $\{p_\mu\}_{\mu\in\Delta}$ be two discrete frames of $\alpha$-molecules of order $(L, M, N_1, N_2)$ with $(\alpha, k)$-consistent parametrizations $(\Lambda, \Phi_\Lambda)$ and $(\Delta, \Phi_\Delta)$ satisfying*

$$L \geq 2\frac{k}{p}, \quad M > 3\frac{k}{p} - \frac{3-\alpha}{2}, \quad N_1 \geq \frac{k}{p} + \frac{1+\alpha}{2}, \quad \textit{and} \quad N_2 \geq 2\frac{k}{p}.$$

*Then $\{m_\lambda\}_{\lambda\in\Lambda}$ and $\{p_\mu\}_{\mu\in\Delta}$ are sparsity equivalent in $\ell^q$ for all $p \leq q < 2$.*

*Proof.* By Lemma 2.3.4, it suffices to prove that

$$\max\Big\{ \sup_{\lambda\in\Lambda} \sum_{\mu\in\Delta} |\langle m_\lambda, p_\mu\rangle|^p, \sup_{\mu\in\Delta} \sum_{\lambda\in\Lambda} |\langle m_\lambda, p_\mu\rangle|^p \Big\}^{1/p} < \infty.$$

Since, by Theorem 2.2.2, we have

$$|\langle m_\lambda, p_\mu \rangle| \lesssim \omega_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu))^{-\frac{k}{p}},$$

we can estimate

$$\max\Big\{ \sup_{\lambda\in\Lambda} \sum_{\mu\in\Delta} |\langle m_\lambda, p_\mu\rangle|^p, \sup_{\mu\in\Delta} \sum_{\lambda\in\Lambda} |\langle m_\lambda, p_\mu\rangle|^p \Big\}$$
$$\lesssim \max\Big\{ \sup_{\lambda\in\Lambda} \sum_{\mu\in\Delta} \omega_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu))^{-k}, \sup_{\mu\in\Delta} \sum_{\lambda\in\Lambda} \omega_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu))^{-k} \Big\}.$$

Due to the $(\alpha, k)$-consistency of the parametrizations $(\Lambda, \Phi_\Lambda)$ and $(\Delta, \Phi_\Delta)$, the right-hand side is finite and the proof is finished. □

We see that, as long as the parametrizations are consistent, the sparsity equivalence of two frames of $\alpha$-molecules can be controlled by the order of the systems. Hereby, the smaller $p$ is, i.e., the more sparsity is promoted, and the less consistent the two frames are, the higher the order of the molecules needs to be for sparsity equivalence, i.e., better time-frequency localization and higher moments of the molecules are required.

Theorem 2.3.6 is called the *transfer principle* for discrete $\alpha$-molecule frames, since in conjunction with Proposition 2.3.2 it allows to transfer approximation properties from one anchor frame to other frames, if the coefficient decay of the anchor frame is known. It will be used in Chapters 5 and 6.

## 2.4 A Sufficient Condition for Discrete $\alpha$-Molecule Frames

As we have already stated in the previous section, for many reasons representation systems which constitute frames play an outstanding role in signal analysis. In practice, the frame property is often verified directly for the concrete systems of interest at hand. Within the framework of $\alpha$-molecules, a more generic approach is possible, however.

In this subsection we want to find a sufficient condition for $\alpha$-molecules to constitute a frame for $L_2(\mathbb{R}^2)$. Clearly, the frame condition is not fulfilled for $\alpha$-molecule systems per se as for instance the zero function is a trivial example of an $\alpha$-molecule system. Hence, to achieve this goal, additional assumptions to ensure the frame property are necessary.

As in the previous section, we only focus on discrete systems of $\alpha$-molecules $\{m_\lambda\}_{\lambda\in\Lambda}$ which are indexed by some countable index set $\Lambda$. For such systems, a sufficient frame condition similar to Daubechies' criterion for wavelets [30] was derived in the Bachelor's thesis [53], under certain additional assumptions on the associated parametrization $(\Lambda, \Phi_\Lambda)$. Our result below is mainly based on this result, but generalizes slightly on the utilized parametrizations.

As in [53], we require the index set $\Lambda$ to be of the form

$$\Lambda = \Delta \times \mathbb{Z}^2 \tag{2.33}$$

for some countable index set $\Delta$ of generic indices. Further, the parametrization map $\Phi_\Lambda$ shall have the special structure

$$\Phi_\Lambda : \Lambda = \Delta \times \mathbb{Z}^2 \to \mathbb{P}, \quad (\mu, k) \mapsto (x_{\mu,k}, \eta_\mu, t_\mu) \text{ with } x_{\mu,k} := R_{\eta_\mu}^{-1} A_{\alpha,t_\mu} T_\mu M_c k, \tag{2.34}$$

where $T_\mu \in \mathbb{R}^{2\times 2}$ is a matrix with $|\det(T_\mu)| = 1$ for each $\mu \in \Delta$ and

$$M_c := \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix} \quad \text{for some fixed parameter } c = (c_1, c_2) \in \mathbb{R}_+^2. \tag{2.35}$$

This condition is slightly less restrictive than the condition imposed on the parametrization map in [53], where $T_\mu$ needs to be the identity matrix for every $\mu \in \Delta$. Finally, as in [53], we assume that the corresponding generators $g_\lambda = g_{\mu,k}$ do not vary with $k \in \mathbb{Z}^2$. For simplicity we write this as $g_\lambda = g_\mu$.

The corresponding $\alpha$-molecules $m_\lambda = m_{\mu,k}$ then have the form

$$m_{\mu,k}(\cdot) = t_\mu^{-(1+\alpha)/2} g_\mu(A_{\alpha,t_\mu}^{-1} R_{\eta_\mu} \cdot - T_\mu M_c k)\,. \tag{2.36}$$

Note that the indices $\mu \in \Delta$ determine the scale and the orientation of the $\alpha$-molecules, whereas the indices $k \in \mathbb{Z}^2$ correspond to translations along the grid $\{T_\mu M_c k : k \in \mathbb{Z}^2\}$.

Parametrizations $(\Lambda, \Phi_\Lambda)$ of the form (2.33) and (2.34) might at first glance seem quite restrictive, but we will see in Sections 3.2 and 3.3 of Chapter 3 that they in particular comprise discrete $\alpha$-curvelet and $\alpha$-shearlet parametrizations. Hence, in view of Theorems 3.2.5 and 3.3.7, the criterion developed below in particular applies to $\alpha$-curvelet and $\alpha$-shearlet molecules, which constitute important subclasses of $\alpha$-molecules. In particular, it generalizes the frame criterion developed in [76] for $\alpha$-shearlets. The criterion from [32, Thm. 3.3] for systems of Gaussian wavepackets is also included, since those can be interpreted as systems of $\alpha$-curvelet molecules.

Before we formulate the statement of the theorem, let us observe that for a function system to form a frame in $L_2(\mathbb{R}^2)$, the spatial as well as the frequency support of its elements needs to cover the whole plane. Furthermore, the energy of the functions may not accumulate too much at any given point. Hence, intuitively, a certain 'spreading' of the functions in phase space is necessary for the frame condition to hold.

In general, Daubechies-type frame criteria are based on the investigation of associated (auto-)correlation functions $\Phi : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ and $\Gamma : \mathbb{R}^2 \to \mathbb{R}$, which measure the extent of the overlaps of the frequency support. For $\alpha$-molecules of the form (2.36) with associated generators $\{g_\mu\}_{\mu\in\Delta}$, these functions take the form

$$\begin{aligned} \Phi(\xi, \omega) &:= \sum_{\mu\in\Delta} |\hat{g}_\mu(A_{\alpha,t_\mu} R_{\eta_\mu} \xi)| |\hat{g}_\mu(A_{\alpha,t_\mu} R_{\eta_\mu} \xi + T_\mu^{-T} \omega)|, \\ \Gamma(\omega) &:= \operatorname*{ess\,sup}_{\xi\in\mathbb{R}^2} \Phi(\xi, \omega). \end{aligned}$$

The statement of the theorem now depends on the size of the following two quantities

$$L_{\inf} := \operatorname*{ess\,inf}_{\xi\in\mathbb{R}^2} \Phi(\xi, 0) \quad \text{and} \quad L_{\sup} := \Gamma(0) = \operatorname*{ess\,sup}_{\xi\in\mathbb{R}^2} \Phi(\xi, 0),$$

as well as the function $R : \mathbb{R}_+^2 \to [0, \infty)$ defined for $c \in \mathbb{R}_+^2$ by

$$R(c) := \sum_{m\in\mathbb{Z}^2\setminus\{0\}} \left[\Gamma(M_c^{-1} m) \Gamma(-M_c^{-1} m)\right]^{1/2}.$$

**Theorem 2.4.1.** *Let $\{m_\lambda\}_{\lambda\in\Lambda}$ be a system of $\alpha$-molecules with respect to a parametrization $(\Lambda, \Phi_\Lambda)$, where $\Lambda = \Delta \times \mathbb{Z}^2$ is an index set of the form (2.33) and $\Phi_\Lambda$ is a map as in (2.34) with $c = (c_1, c_2) \in \mathbb{R}^2_+$ fixed. Further, assume that the generators $g_\lambda = g_{\mu,k}$ do not vary with $k \in \mathbb{Z}^2$. Under these assumptions, the condition*

$$R(c) < L_{\inf} \le L_{\sup} < \infty \tag{2.37}$$

*ensures that $\{m_\lambda\}_{\lambda\in\Lambda}$ constitutes a frame for $L_2(\mathbb{R}^2)$ with frame bounds $A, B > 0$ satisfying*

$$\frac{L_{\inf} - R(c)}{|\det M_c|} \le A \le B \le \frac{L_{\sup} + R(c)}{|\det M_c|}.$$

The technique used for the proof of this theorem goes back to [30], and has also been used in [76], [32], and [53]. Our exposition follows the proof in the latter reference, with marginal modifications.

*Proof.* For fixed $\mu \in \Delta$ we first consider the sum $S_\mu := \sum_{k\in\mathbb{Z}^2} |\langle f, m_{\mu,k}\rangle|^2 = \sum_{k\in\mathbb{Z}^2} |\langle \hat{f}, \hat{m}_{\mu,k}\rangle|^2$. Plugging in (2.36), we have

$$S_\mu = \sum_{k\in\mathbb{Z}^2} \Big| \int_{\mathbb{R}^2} t_\mu^{(1+\alpha)/2} \hat{f}(\xi) \overline{\hat{g}_\mu(A_{\alpha,t_\mu} R_{\eta_\mu} \xi)} \exp(2\pi i \langle (R_{\eta_\mu}^{-1} A_{\alpha,t_\mu} T_\mu M_c)^T \xi, k\rangle)\, d\xi \Big|^2,$$

and the substitution $(R_{\eta_\mu}^{-1} A_{\alpha,t_\mu} T_\mu M_c)^T \xi \mapsto \xi$ yields

$$S_\mu = \sum_{k\in\mathbb{Z}^2} \Big| \int_{\mathbb{R}^2} t_\mu^{-(1+\alpha)/2} |\det(M_c)|^{-1} \hat{f}(R_{\eta_\mu}^{-1} A_{\alpha,t_\mu}^{-1} T_\mu^{-T} M_c^{-1}\xi) \overline{\hat{g}_\mu(T_\mu^{-T} M_c^{-1} \xi)} \exp(2\pi i \langle \xi, k\rangle)\, d\xi \Big|^2.$$

With $\Omega := [-\frac{1}{2}, \frac{1}{2})^2$, we can write

$$\begin{aligned} S_\mu &= \sum_{k\in\mathbb{Z}^2} \Big| \sum_{\ell\in\mathbb{Z}^2} \int_{\ell+\Omega} t_\mu^{-(1+\alpha)/2} |\det(M_c)|^{-1} \hat{f}(R_{\eta_\mu}^{-1} A_{\alpha,t_\mu}^{-1} T_\mu^{-T} M_c^{-1}\xi) \overline{\hat{g}_\mu(T_\mu^{-T} M_c^{-1} \xi)} \\ &\qquad \cdot \exp(2\pi i \langle \xi, k\rangle)\, d\xi \Big|^2 \\ &= \sum_{k\in\mathbb{Z}^2} \Big| \int_{\Omega} \sum_{\ell\in\mathbb{Z}^2} t_\mu^{-(1+\alpha)/2} |\det(M_c)|^{-1} \hat{f}(R_{\eta_\mu}^{-1} A_{\alpha,t_\mu}^{-1} T_\mu^{-T} M_c^{-1}(\xi+\ell)) \overline{\hat{g}_\mu(T_\mu^{-T} M_c^{-1} (\xi+\ell))} \\ &\qquad \cdot \exp(2\pi i \langle \xi, k\rangle) \exp(2\pi i \langle \ell, k\rangle)\, d\xi \Big|^2. \end{aligned}$$

Using the Parseval identity in $L_2(\Omega)$, we deduce

$$S_\mu = \int_{\Omega} \Big| \sum_{\ell\in\mathbb{Z}^2} t_\mu^{-(1+\alpha)/2} |\det(M_c)|^{-1} \hat{f}(R_{\eta_\mu}^{-1} A_{\alpha,t_\mu}^{-1} T_\mu^{-T} M_c^{-1}(\xi+\ell)) \overline{\hat{g}_\mu(T_\mu^{-T} M_c^{-1} (\xi+\ell))} \Big|^2 \, d\xi.$$

We continue with

$$\begin{aligned}
S_\mu &= \int_\Omega t_\mu^{-(1+\alpha)} |\det(M_c)|^{-2} \sum_{k,\ell\in\mathbb{Z}^2} \hat{f}(R_{\eta_\mu}^{-1} A_{\alpha,t_\mu}^{-1} T_\mu^{-T} M_c^{-1}(\xi+k)) \overline{\hat{g}_\mu(T_\mu^{-T} M_c^{-1}(\xi+k))} \\
&\quad \cdot \overline{\hat{f}(R_{\eta_\mu}^{-1} A_{\alpha,t_\mu}^{-1} T_\mu^{-T} M_c^{-1}(\xi+\ell))} \hat{g}_\mu(T_\mu^{-T} M_c^{-1}(\xi+\ell))\, d\xi \\
&= \sum_{\ell\in\mathbb{Z}^2} \int_{\ell+\Omega} t_\mu^{-(1+\alpha)} |\det(M_c)|^{-2} \sum_{k\in\mathbb{Z}^2} \hat{f}(R_{\eta_\mu}^{-1} A_{\alpha,t_\mu}^{-1} T_\mu^{-T} M_c^{-1}(\xi+k-\ell)) \\
&\quad \cdot \overline{\hat{g}_\mu(T_\mu^{-T} M_c^{-1}(\xi+k-\ell))}\, \overline{\hat{f}(R_{\eta_\mu}^{-1} A_{\alpha,t_\mu}^{-1} T_\mu^{-T} M_c^{-1}\xi)} \hat{g}_\mu(T_\mu^{-T} M_c^{-1}\xi)\, d\xi \\
&= \int_{\mathbb{R}^2} \sum_{m\in\mathbb{Z}^2} t_\mu^{-(1+\alpha)} |\det M_c|^{-2} \hat{f}(R_{\eta_\mu}^{-1} A_{\alpha,t_\mu}^{-1} T_\mu^{-T} M_c^{-1}\xi) \overline{\hat{g}_\mu(T_\mu^{-T} M_c^{-1}\xi)} \\
&\quad \cdot \overline{\hat{f}(R_{\eta_\mu}^{-1} A_{\alpha,t_\mu}^{-1} T_\mu^{-T} M_c^{-1}(\xi+m))} \hat{g}_\mu(T_\mu^{-T} M_c^{-1}(\xi+m)).
\end{aligned}$$

Finally, we substitute $R_{\eta_\mu}^{-1} A_{\alpha,t_\mu}^{-1} T_\mu^{-T} M_c^{-1}\xi \mapsto \xi$ and arrive at

$$\begin{aligned}
S_\mu = |\det M_c|^{-1} \int_{\mathbb{R}^2} \sum_{m\in\mathbb{Z}^2} &\hat{f}(\xi) \overline{\hat{g}_\mu(A_{\alpha,t_\mu} R_{\eta_\mu}\xi)} \\
&\cdot \overline{\hat{f}(\xi + R_{\eta_\mu}^{-1} A_{\alpha,t_\mu}^{-1} T_\mu^{-T} M_c^{-1} m)} \hat{g}_\mu(A_{\alpha,t_\mu} R_{\eta_\mu}\xi + T_\mu^{-T} M_c^{-1} m).
\end{aligned}$$

Now we take the sum over $\mu \in \Delta$ and split

$$\sum_{\mu\in\Delta} S_\mu = T_1 + T_2$$

into a term $T_1$, corresponding to $m = 0$,

$$T_1 := |\det M_c|^{-1} \int_{\mathbb{R}^2} \sum_{\mu\in\Delta} |\hat{f}(\xi)|^2 |\hat{g}_\mu(A_{\alpha,t_\mu} R_{\eta_\mu}\xi)|^2\, d\xi,$$

and a term $T_2$, corresponding to $m \neq 0$,

$$\begin{aligned}
T_2 := |\det M_c|^{-1} \int_{\mathbb{R}^2} \sum_{\mu\in\Delta} \sum_{m\in\mathbb{Z}^2\setminus\{0\}} &\hat{f}(\xi) \overline{\hat{f}(\xi + R_{\eta_\mu}^{-1} A_{\alpha,t_\mu}^{-1} T_\mu^{-T} M_c^{-1} m)} \\
&\cdot \overline{\hat{g}_\mu(A_{\alpha,t_\mu} R_{\eta_\mu}\xi)} \hat{g}_\mu(A_{\alpha,t_\mu} R_{\eta_\mu}\xi + T_\mu^{-T} M_c^{-1} m)\, d\xi.
\end{aligned}$$

Using the definition of $L_{\inf}$ and $L_{\sup}$, we directly obtain for $T_1$

$$\frac{L_{\inf}}{|\det(M_c)|} \|f\|^2 \leq T_1 \leq \frac{L_{\sup}}{|\det(M_c)|} \|f\|^2. \tag{2.38}$$

Further, for $T_2$ we will prove

$$|T_2| \leq \frac{R(c)}{|\det(M_c)|} \|f\|^2. \tag{2.39}$$

For this, we first estimate

$$\begin{aligned}
|T_2| \leq |\det M_c|^{-1} \sum_{m\in\mathbb{Z}^2\setminus\{0\}} \int_{\mathbb{R}^2} \sum_{\mu\in\Delta} &|\hat{f}(\xi)| \left| \hat{g}_\mu(A_{\alpha,t_\mu} R_{\eta_\mu}\xi) \hat{g}_\mu(A_{\alpha,t_\mu} R_{\eta_\mu}\xi + T_\mu^{-T} M_c^{-1} m) \right|^{1/2} \\
\cdot |\hat{f}(\xi + R_{\eta_\mu}^{-1} A_{\alpha,t_\mu}^{-1} T_\mu^{-T} M_c^{-1} m)| &\left| \hat{g}_\mu(A_{\alpha,t_\mu} R_{\eta_\mu}\xi) \hat{g}_\mu(A_{\alpha,t_\mu} R_{\eta_\mu}\xi + T_\mu^{-T} M_c^{-1} m) \right|^{1/2} d\xi.
\end{aligned}$$

Then we proceed with a double application of the Cauchy-Schwarz inequality, first to the sum and then to the integral. We obtain

$$|T_2| \leq |\det M_c|^{-1} \sum_{m\in\mathbb{Z}^2\setminus\{0\}} \Big( \int_{\mathbb{R}^2} \sum_{\mu\in\Delta} |\hat{f}(\xi)|^2 |\hat{g}_\mu(A_{\alpha,t_\mu}R_{\eta_\mu}\xi)||\hat{g}_\mu(A_{\alpha,t_\mu}R_{\eta_\mu}\xi + T_\mu^{-T}M_c^{-1}m)|\,d\xi\Big)^{1/2}$$
$$\cdot \Big( \int_{\mathbb{R}^2} \sum_{\mu\in\Delta} |\hat{f}(\xi + R_{\eta_\mu}^{-1}A_{\alpha,t_\mu}^{-1}T_\mu^{-T}M_c^{-1}m)|^2 |\hat{g}_\mu(A_{\alpha,t_\mu}R_{\eta_\mu}\xi)\hat{g}_\mu(A_{\alpha,t_\mu}R_{\eta_\mu}\xi + T_\mu^{-T}M_c^{-1}m)|\,d\xi\Big)^{1/2}.$$

Finally, we substitute $\xi + R_{\eta_\mu}^{-1}A_{\alpha,t_\mu}^{-1}T_\mu^{-T}M_c^{-1}n \mapsto \xi$ in the second integral and arrive at (2.39), namely

$$|T_2| \leq |\det M_c|^{-1} \sum_{m\in\mathbb{Z}^2\setminus\{0\}} \Big( \int_{\mathbb{R}^2} |\hat{f}(\xi)|^2 \sum_{\mu\in\Delta} |\hat{g}_\mu(A_{\alpha,t_\mu}R_{\eta_\mu}\xi)\hat{g}_\mu(A_{\alpha,t_\mu}R_{\eta_\mu}\xi + T_\mu^{-T}M_c^{-1}m)|\,d\xi\Big)^{1/2}$$
$$\cdot \Big( \int_{\mathbb{R}^2} |\hat{f}(\xi)|^2 \sum_{\mu\in\Delta} |\hat{g}_\mu(A_{\alpha,t_\mu}R_{\eta_\mu}\xi)\hat{g}_\mu(A_{\alpha,t_\mu}R_{\eta_\mu}\xi - T_\mu^{-T}M_c^{-1}m)|\,d\xi\Big)^{1/2}$$
$$\leq |\det M_c|^{-1}\|\hat{f}\|^2 \sum_{m\in\mathbb{Z}^2\setminus\{0\}} \Big[\Gamma(M_c^{-1}m)\Gamma(-M_c^{-1}m)\Big]^{1/2} = \frac{R(c)}{|\det M_c|}\|f\|^2.$$

Altogether, the estimates (2.38) and (2.39) imply

$$\frac{L_{\inf} - R(c)}{|\det(M_c)|}\|f\|^2 \leq \sum_{\mu\in\Delta} S_\mu \leq \frac{L_{\sup} + R(c)}{|\det(M_c)|}\|f\|^2,$$

which finishes the proof. □

Concerning the application of this theorem, let us remark that, intuitively, when $c$ gets smaller also $R(c)$ becomes smaller. A good strategy to fulfill the condition (2.37) in Theorem 2.4.1 is thus to choose the parameter $c$ sufficiently small. This strategy was used for example in [76] to confirm the frame property of the constructed compactly supported shearlets.

In the next chapter, where we will investigate specific instances of $\alpha$-molecule systems, we will use this theorem to deduce sufficient frame criteria for $\alpha$-curvelet and $\alpha$-shearlet molecules.

## 2.5 Appendix: *Proof of Theorem 2.2.2*

In the following, we reproduce the proof of Theorem 2.2.2 given in [59], whereby the notation is adapted to our setting and some minor inaccuracies are corrected. We start by collecting some useful lemmata in Subsections 2.5.1–2.5.3, followed by the actual proof in Subsection 2.5.4.

### 2.5.1 General Estimates

Let us recall the projective bracket introduced in (2.9). An important property of this bracket is given by the following lemma.

**Lemma 2.5.1.** *For $\theta \in \mathbb{R}$ let $\{\theta\}$ denote its 'projection' onto the interval $\mathbf{T} := [-\frac{\pi}{2}, \frac{\pi}{2})$ as introduced in* (2.9)*. It then holds $|\{\theta\}| \asymp |\sin(\theta)|$.*

*Proof.* Due to $\pi$-periodicity it suffices to verify the relation for $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2})$. In this range we have $\frac{2}{\pi}|\theta| \leq |\sin(\theta)| \leq |\theta|$. □

The following lemma can be found in [51, Appendix K.1].

**Lemma 2.5.2.** *For $N > 1$ and $a, a' \in \mathbb{R}_+$, we have the inequality*

$$\int_{\mathbb{R}} (1 + a|x|)^{-N} \left(1 + a'|x - y|\right)^{-N} dx \lesssim \max\{a, a'\}^{-1}(1 + \min\{a, a'\}|y|)^{-N}.$$

*The implicit constant is independent of $a, a'$ and $y \in \mathbb{R}$.*

The following result can be regarded as a corollary of the previous lemma.

**Lemma 2.5.3.** *Assume that $\eta \in \mathbb{R}$ and $N > 1$. Then we have for $a, a' \in \mathbb{R}_+$ the inequality*

$$\int_{\mathbb{T}} (1 + a|\sin(\varphi)|)^{-N} \left(1 + a'|\sin(\varphi + \eta)|\right)^{-N} d\varphi \lesssim \max\{a, a'\}^{-1}(1 + \min\{a, a'\}|\{\eta\}|)^{-N}. \tag{2.40}$$

*The implicit constant is independent of $a, a'$ and $\eta \in \mathbb{R}$.*

*Proof.* Let $\mathbf{T} := [-\pi/2, \pi/2)$. Using the $\pi$-periodicity of the integrand and the equivalence $|\sin(\varphi)| \asymp |\{\varphi\}|$, where $\varphi \in \mathbb{R}$ and $\{\cdot\} = \{\cdot\}_{\mathbf{T}}$ is the projective bracket defined in (2.9), we obtain

$$\begin{aligned}
&\int_{\mathbb{T}} (1 + a|\sin(\varphi)|)^{-N} \left(1 + a'|\sin(\varphi + \eta)|\right)^{-N} d\varphi \\
&\qquad\qquad \asymp \int_{\mathbf{T}} (1 + a|\sin(\varphi)|)^{-N} \left(1 + a'|\sin(\varphi + \eta)|\right)^{-N} d\varphi \\
&\qquad\qquad \asymp \int_{\mathbf{T}} (1 + a|\varphi|)^{-N} \left(1 + a'|\{\varphi + \eta\}|\right)^{-N} d\varphi.
\end{aligned}$$

The left-hand side of (2.40) can thus be estimated by a constant times

$$\sum_{\theta \in \{0, \pm\pi\}} \int_{\mathbb{R}} (1 + a|\varphi|)^{-N} \left(1 + a'|\varphi + \{\eta\} + \theta|\right)^{-N} d\varphi,$$

where we used $\{\varphi+\eta\} = \{\{\varphi\}+\{\eta\}\}$. Further, by Lemma 2.5.2 and since $|\{\eta+\theta\}| \leq |\{\eta\}+\theta|$, this sum is bounded by a constant times

$$\sum_{\theta \in \{0, \pm\pi\}} \max\{a, a'\}^{-1}(1 + \min\{a, a'\}|\{\eta + \theta\}|)^{-N}.$$

It remains to note that for $\theta \in \{0, \pm\pi\}$ we have $|\{\eta+\theta\}| = |\{\eta\}|$. This proves the lemma. □

### 2.5.2 Basic Estimates of $S_{\lambda,M,N_1,N_2}$

We now consider the function $S_{\lambda,M,N_1,N_2} : [0,\infty)\times[0,2\pi) \to \mathbb{R}$ for $\lambda \in \Lambda$ and $M, N_1, N_2 \in \mathbb{N}_0$ which is defined in polar coordinates by

$$S_{\lambda,M,N_1,N_2}(r,\varphi) := \min\Big\{1, t_\lambda(1+r)\Big\}^M \Big(1 + t_\lambda^{-(1-\alpha)}|\sin(\varphi+\eta_\lambda)|\Big)^{-N_2}\Big(1+t_\lambda r\Big)^{-N_1}.$$

The reader might want to compare this definition with (2.6).

The following lemma will be used in order to decouple the angular and the radial variables.

**Lemma 2.5.4.** *For every* $0 \le K \le N_2$,

$$\min\Big\{1, t_\lambda(1+r)\Big\}^M \Big(1+t_\lambda r\Big)^{-N_1}\Big(1+t_\lambda^\alpha r|\sin(\varphi+\eta_\lambda)|\Big)^{-N_2} \lesssim S_{\lambda,M-K,N_1,K}(r,\varphi).$$

*Proof.* After choosing $K$, we can estimate the quantity on the left hand side by

$$\min\{1, t_\lambda(1+r)\}^{M-K}(1+t_\lambda r)^{-N_1}\left(\frac{\min\{1,t_\lambda(1+r)\}}{1+t_\lambda^\alpha r|\sin(\varphi+\eta_\lambda)|}\right)^K.$$

We need to show that

$$\frac{\min\{1,t_\lambda(1+r)\}}{1+t_\lambda^\alpha r|\sin(\varphi+\eta_\lambda)|} \lesssim \Big(1+t_\lambda^{-(1-\alpha)}|\sin(\varphi+\eta_\lambda)|\Big)^{-1}. \tag{2.41}$$

In order to prove (2.41), we distinguish three cases:

- $\mathbf{r \le 1}$**:** For $r \le 1$ we have

$$\frac{\min\{1,t_\lambda(1+r)\}}{1+t_\lambda^\alpha r|\sin(\varphi+\eta_\lambda)|} \lesssim \min\{1,t_\lambda\} \lesssim \Big(1+t_\lambda^{-(1-\alpha)}|\sin(\varphi+\eta_\lambda)|\Big)^{-1}.$$

- $\mathbf{t_\lambda^{-1} \le r}$**:** In this case we derive

$$\begin{aligned}\frac{\min\{1,t_\lambda(1+r)\}}{1+t_\lambda^\alpha r|\sin(\varphi+\eta_\lambda)|} &= \frac{1}{1+t_\lambda^\alpha r|\sin(\varphi+\eta_\lambda)|} \le \frac{1}{1+t_\lambda^\alpha t_\lambda^{-1}|\sin(\varphi+\eta_\lambda)|}\\ &= \Big(1+t_\lambda^{-(1-\alpha)}|\sin(\varphi+\eta_\lambda)|\Big)^{-1}.\end{aligned}$$

If $t_\lambda^{-1} > 1$ we have to examine a third case.

- $\mathbf{1 < r < t_\lambda^{-1}}$**:** In this case we have

$$\frac{\min\{1,t_\lambda(1+r)\}}{1+t_\lambda^\alpha r|\sin(\varphi+\eta_\lambda)|} \le \frac{t_\lambda(1+r)}{1+t_\lambda^\alpha r|\sin(\varphi+\eta_\lambda)|} = \frac{1+r}{r}\,\frac{1}{r^{-1}t_\lambda^{-1}+t_\lambda^{-(1-\alpha)}|\sin(\varphi+\eta_\lambda)|}.$$

  Since $r > 1$, we have $\frac{1+r}{r} < 2$, and since $r < t_\lambda^{-1}$, also $r^{-1}t_\lambda^{-1} > 1$ holds.

This proves the statement. $\square$

The next lemma provides estimates for the inner product of two functions of the form $S_{\lambda,M,N_1,N_2}$.

**Lemma 2.5.5.** *We assume $t_\lambda, t_\mu \le C < \infty$ for all $\lambda \in \Lambda$ and $\mu \in \Delta$. For $A, B \ge 1$ and*

$$N_1 \ge A + \frac{1+\alpha}{2}, \quad N_2 \ge B, \quad M > N_1 - 2,$$

*we have if $N_1 > 1$*

$$\begin{aligned}(t_\lambda t_\mu)^{\frac{1+\alpha}{2}} &\int_{\mathbb{R}_+}\int_{\mathbb{T}} S_{\lambda,M,N_1,N_2}(r,\varphi) S_{\mu,M,N_1,N_2}(r,\varphi) r d\varphi dr \\ &\lesssim \max\left\{\frac{t_\lambda}{t_\mu}, \frac{t_\mu}{t_\lambda}\right\}^{-A} \left(1 + \max\{t_\lambda, t_\mu\}^{-(1-\alpha)} |\{\eta_\lambda - \eta_\mu\}|\right)^{-B}.\end{aligned}$$

*Proof.* We assume without loss of generality that $t_\lambda \ge t_\mu$ and start by proving the angular decay. An application of Lemma 2.5.3 yields

$$\begin{aligned}\int_{\mathbb{T}} \left(1 + t_\lambda^{-(1-\alpha)} |\sin(\varphi + \eta_\lambda)|\right)^{-N_2} \left(1 + t_\mu^{-(1-\alpha)} |\sin(\varphi + \eta_\mu)|\right)^{-N_2} d\varphi \\ \lesssim t_\mu^{1-\alpha} \left(1 + t_\lambda^{-(1-\alpha)} |\{\eta_\lambda - \eta_\mu\}|\right)^{-N_2}.\end{aligned}$$

Taking into account $N_2 \ge B \ge 1$, we thus obtain

$$\begin{aligned}(t_\lambda t_\mu)^{\frac{1+\alpha}{2}} &\int_{\mathbb{R}_+}\int_{\mathbb{T}} S_{\lambda,M,N_1,N_2}(r,\varphi) S_{\mu,M,N_1,N_2}(r,\varphi) r d\varphi dr \\ &\lesssim \mathcal{S} \cdot \left(\frac{t_\lambda}{t_\mu}\right)^{\frac{1+\alpha}{2}} \left(1 + t_\lambda^{-(1-\alpha)} |\{\eta_\lambda - \eta_\mu\}|\right)^{-B},\end{aligned}$$

where

$$\mathcal{S} := t_\mu^2 \int_{\mathbb{R}_+} \min\{1, t_\lambda(1+r)\}^M \min\{1, t_\mu(1+r)\}^M (1 + t_\lambda r)^{-N_1} (1 + t_\mu r)^{-N_1} r dr.$$

The remaining estimate

$$\mathcal{S} \lesssim (t_\lambda / t_\mu)^{-(A + \frac{1+\alpha}{2})} \tag{2.42}$$

is proved by splitting up the integral into the three parts $\mathcal{S}_i$, $i = 1, 2, 3$, where the integration ranges over $0 < r < 1$, $1 \le r \le \max\{1, t_\mu^{-1}\}$ and $\max\{1, t_\mu^{-1}\} < r$, respectively.

*Case 1* ($0 < r < 1$): For $\mathcal{S}_1$ we integrate over $0 < r < 1$. Here we use the moment property and $t_\lambda^{-1} \ge 1/C > 0$ to estimate

$$\begin{aligned}\mathcal{S}_1 &\lesssim t_\mu^2 \int_0^1 t_\lambda^M t_\mu^M dr \\ &= t_\mu^{2+M} t_\lambda^M \\ &\lesssim t_\mu^{2+M} t_\lambda^{-(M+2)} \\ &= (t_\mu / t_\lambda)^{M+2} \\ &\le (t_\mu / t_\lambda)^{A + \frac{1+\alpha}{2}}.\end{aligned}$$

*Case 2* $(1 \leq r \leq \max\{1, t_\mu^{-1}\})$: If $t_\mu^{-1} \leq 1$ then $\mathcal{S}_2 = 0$. For $t_\mu^{-1} > 1$ we estimate

$$\begin{aligned}
\mathcal{S}_2 &\lesssim t_\mu^2 \int_1^{t_\mu^{-1}} (t_\mu r)^M (t_\lambda r)^{-N_1} r dr \\
&\leq t_\mu^{2+M} t_\lambda^{-N_1} \int_0^{t_\mu^{-1}} r^{M+1-N_1} dr \\
&\lesssim t_\mu^{2+M} t_\lambda^{-N_1} t_\mu^{-(M+2-N_1)} \\
&= (t_\mu / t_\lambda)^{N_1} \\
&\leq (t_\mu / t_\lambda)^{A + \frac{1+\alpha}{2}}.
\end{aligned}$$

*Case 3* $(\max\{1, t_\mu^{-1}\} < r)$: For $\mathcal{S}_3$ we estimate

$$\begin{aligned}
\mathcal{S}_3 &\lesssim t_\mu^2 \int_{t_\mu^{-1}}^{\infty} (t_\lambda r)^{-N_1} (t_\mu r)^{-N_1} r dr \\
&= t_\mu^2 t_\mu^{-N_1} t_\lambda^{-N_1} \int_{t_\mu^{-1}}^{\infty} r^{-2N_1+1} dr \\
&\lesssim t_\mu^2 t_\mu^{-N_1} t_\lambda^{-N_1} t_\mu^{2N_1-2} \\
&= (t_\mu / t_\lambda)^{N_1} \\
&\leq (t_\mu / t_\lambda)^{A + \frac{1+\alpha}{2}}.
\end{aligned}$$

Altogether, this establishes (2.42). □

### 2.5.3 Estimates with Differential Operator

Finally, we require some estimates involving the symmetric differential operator $\mathcal{L}_{\lambda,\mu}$ (acting on the frequency variable $\xi \in \mathbb{R}^2$) defined for $\lambda \in \Lambda$ and $\mu \in \Delta$ by

$$\mathcal{L}_{\lambda,\mu} := I - t_0^{-2\alpha} \Delta_\xi - \frac{t_0^{-2}}{1 + t_0^{-2(1-\alpha)} |\{\delta\eta\}|^2} \langle e_\lambda, \nabla \rangle^2, \tag{2.43}$$

where $\delta\eta := \eta_\lambda - \eta_\mu$ and $t_0 := \max\{t_\lambda, t_\mu\}$. The first lemma is an auxiliary result.

**Lemma 2.5.6.** *Assume that the assumptions of Theorem 2.2.2 hold true for two systems of $\alpha$-molecules of order $(L, M, N_1, N_2)$ with respective generators $\{a^{(\lambda)}\}_\lambda$ and $\{b^{(\mu)}\}_\mu$. Given any two of those generators $a^{(\lambda)}, b^{(\mu)}$, the expression*

$$\mathcal{L}_{\lambda,\mu} \left( \hat{a}^{(\lambda)} (A_{\alpha,t_\lambda} R_{\eta_\lambda} \xi) \overline{\hat{b}^{(\mu)} (A_{\alpha,t_\mu} R_{\eta_\mu} \xi)} \right)$$

*can be written as a finite linear combination of terms of the form*

$$\hat{c}^{(\lambda)} (A_{\alpha,t_\lambda} R_{\eta_\lambda} \xi) \overline{\hat{d}^{(\mu)} (A_{\alpha,t_\mu} R_{\eta_\mu} \xi)},$$

*with $c^{(\lambda)}, d^{(\mu)}$ satisfying (2.5) for $L-2, M, N_1, N_2$ with an implicit constant independent of $\lambda$ and $\mu$. Thereby, the number of terms in the linear combinations and the corresponding coefficients are also uniformly bounded in $\lambda$ and $\mu$.*

*Proof.* To prove the claim we treat the three summands of the operator $\mathcal{L}_{\lambda,\mu}$ separately. The first part is the identity, and therefore the statement is trivial. To handle the second part, the frequency Laplacian $t_0^{-2\alpha}\Delta$, we use the product rule

$$\Delta(fg) = 2\left(\partial^{(1,0)} f \partial^{(1,0)} g + \partial^{(0,1)} f \partial^{(0,1)} g\right) + (\Delta f)g + f(\Delta g).$$

Therefore we need to estimate the derivatives of degree 1 and the Laplacians of the two factors in the product

$$\hat{a}^{(\lambda)}\left(A_{\alpha,t_\lambda} R_{\eta_\lambda}\xi\right) \overline{\hat{b}^{(\mu)}\left(A_{\alpha,t_\mu} R_{\eta_\mu}\xi\right)} =: A(\xi)B(\xi).$$

For this, we start with the first factor,

$$A(\xi) = \hat{a}^{(\lambda)}\left(t_\lambda \cos(\eta_\lambda)\xi_1 - t_\lambda \sin(\eta_\lambda)\xi_2, t_\lambda^\alpha \sin(\eta_\lambda)\xi_1 + t_\lambda^\alpha \cos(\eta_\lambda)\xi_2\right).$$

Let us set

$$A_1(\xi) := \left(\partial^{(1,0)}\hat{a}^{(\lambda)}\right)\left(A_{\alpha,t_\lambda} R_{\eta_\lambda}\xi\right) \quad \text{and} \quad A_2(\xi) := \left(\partial^{(0,1)}\hat{a}^{(\lambda)}\right)\left(A_{\alpha,t_\lambda} R_{\eta_\lambda}\xi\right).$$

By definition, the functions $\partial^{(1,0)}\hat{a}^{(\lambda)}$, $\partial^{(0,1)}\hat{a}^{(\lambda)}$ satisfy (2.5) with $L$ replaced by $L-1$. An application of the chain rule shows that

$$\partial^{(1,0)} A(\xi) = t_\lambda \cos(\eta_\lambda) A_1(\xi) + t_\lambda^\alpha \sin(\eta_\lambda) A_2(\xi).$$

Analogously, one can compute

$$\partial^{(0,1)} A(\xi) = -t_\lambda \sin(\eta_\lambda) A_1(\xi) + t_\lambda^\alpha \cos(\eta_\lambda) A_2(\xi),$$

and the exact same expressions for $B$ using the obvious definitions for $B_1$ and $B_2$. We get

$$\begin{aligned} \partial^{(1,0)} A \partial^{(1,0)} B \;=\; & t_\lambda t_\mu \cos(\eta_\lambda)\cos(\eta_\mu) A_1 B_1 + t_\lambda^\alpha t_\mu \sin(\eta_\lambda)\cos(\eta_\mu) A_2 B_1 \\ & + t_\mu^\alpha t_\lambda \sin(\eta_\mu)\cos(\eta_\lambda) A_1 B_2 + (t_\lambda t_\mu)^\alpha \sin(\eta_\lambda)\sin(\eta_\mu) A_2 B_2. \end{aligned}$$

It follows that $t_0^{-2\alpha}\partial^{(1,0)} A \partial^{(1,0)} B$ can be written as a linear combination as claimed (recall that $t_0 = \max\{t_\lambda, t_\mu\}$). The same argument applies to the product $t_0^{-2\alpha}\partial^{(0,1)} A \partial^{(0,1)} B$.

It remains to consider the factor

$$(\Delta A)B + A(\Delta B),$$

where, for symmetry reasons, we only treat the summand $(\Delta A)B$. In fact, it suffices to only consider

$$(\partial^{(2,0)} A)B = \left(t_\lambda^2 \cos(\eta_\lambda)^2 A_{11} + 2t_\lambda^{1+\alpha}\sin(\eta_\lambda)\cos(\eta_\lambda) A_{12} + t_\lambda^{2\alpha}\sin(\eta_\lambda)^2 A_{22}\right) B \tag{2.44}$$

with $A_{ij}$, $i,j \in \{1,2\}$, defined in the obvious way, and where $\partial^{(2,0)}\hat{a}^{(\lambda)}$, $\partial^{(1,1)}\hat{a}^{(\lambda)}$, and $\partial^{(0,2)}\hat{a}^{(\lambda)}$ satisfy (2.5) with $L$ replaced by $L-2$. The term $(\partial^{(0,2)} A)B$, and hence $(\Delta A)B$, can be handled in the same way, as can $A(\Delta B)$. This takes care of the term $t_0^{-2\alpha}\Delta$ in the definition of $\mathcal{L}_{\lambda,\mu}$.

Finally, we need to handle the last term in the definition of $\mathcal{L}_{\lambda,\mu}$, namely

$$\frac{t_0^{-2}}{1+t_0^{-2(1-\alpha)}|\{\eta_\mu\}|^2}\frac{\partial^2}{\partial\xi_1^2}$$

for $\eta_\lambda = 0$ (otherwise the second order derivative would be in the direction of the unit vector with angle $\eta_\lambda$ with obvious modifications in the proof). With our notation and using the product rule we need to consider terms of the form

$$(\partial^{(2,0)}A)B, \quad (\partial^{(1,0)}A)(\partial^{(1,0)}B), \quad A(\partial^{(2,0)}B),$$

and show that each of them, multiplied by the factor $t_0^{-2}/(1+t_0^{-2(1-\alpha)}|\{\eta_\mu\}|^2)$, possesses the desired representation.

Let us start with $(\partial^{(2,0)}A)B$, which, using (2.44) and the fact that $\eta_\lambda = 0$, can be written as

$$(\partial^{(2,0)}A)B = t_\lambda^2 A_{11}B,$$

and which clearly satisfies the desired assertion.

Now consider the expression $(\partial^{(1,0)}A)(\partial^{(1,0)}B)$, which can be written as

$$(\partial^{(1,0)}A)(\partial^{(1,0)}B) = t_\lambda t_\mu \cos(\eta_\mu)A_1B_1 + t_\lambda t_\mu^\alpha \sin(\eta_\mu)A_1B_2.$$

The first summand in this expression clearly causes no problems. To handle the second term we need to show that

$$\frac{t_0^{-2}}{1+t_0^{-2(1-\alpha)}|\{\eta_\mu\}|^2}t_\lambda t_\mu^\alpha|\sin(\eta_\mu)| \lesssim 1. \tag{2.45}$$

Here we have to distinguish two cases. First, assume that $|\{\eta_\mu\}| \le t_0^{1-\alpha}$. Then we can estimate $|\sin(\eta_\mu)| \le t_0^{1-\alpha}$, which readily yields the desired bound for (2.45). For the case $|\{\eta_\mu\}| \ge t_0^{1-\alpha}$ we estimate

$$\frac{t_0^{-2}}{1+t_0^{-2(1-\alpha)}|\{\eta_\mu\}|^2}t_\lambda t_\mu^\alpha|\sin(\eta_\mu)| \le \frac{t_0^{-2}}{1+t_0^{-(1-\alpha)}|\{\eta_\mu\}|}t_0t_0^\alpha|\{\eta_\mu\}| \le \frac{t_0^{-2}}{t_0^{-(1-\alpha)}|\{\eta_\mu\}|}t_0t_0^\alpha|\{\eta_\mu\}| = 1$$

which proves (2.45) also for this case.

We are left with estimating the term $A\left(\partial^{(2,0)}B\right)$, which, similar to (2.44), can be written as

$$t_\mu^2\cos(\eta_\mu)^2AB_{11} + 2t_\mu^{1+\alpha}\sin(\eta_\mu)\cos(\eta_\mu)AB_{12} + t_\mu^{2\alpha}\sin(\eta_\mu)^2AB_{22}.$$

The first two terms are of a form already treated, and the last term can be handled using the fact that $|\sin(\eta_\mu)|^2 \le |\{\eta_\mu\}|^2$. □

**Lemma 2.5.7.** *Assume that the assumptions of Theorem 2.2.2 hold for two systems of $\alpha$-molecules of order $(L,M,N_1,N_2)$ with respective generating functions $\{a^{(\lambda)}\}_\lambda$ and $\{b^{(\mu)}\}_\mu$. Then we have*

$$\mathcal{L}_{\lambda,\mu}^k\left(\hat{a}^{(\lambda)}\left(A_{\alpha,t_\lambda}R_{\eta_\lambda}\xi\right)\overline{\hat{b}^{(\mu)}\left(A_{\alpha,t_\mu}R_{\eta_\mu}\xi\right)}\right) \lesssim S_{\lambda,M-N_2,N_1,N_2}(\xi)S_{\mu,M-N_2,N_1,N_2}(\xi)$$

*for all $k \le L/2$.*

*Proof.* We show that

$$\begin{aligned}
&\left|\mathcal{L}^k_{\lambda,\mu}\left(\hat{a}^{(\lambda)}\left(A_{\alpha,t_\lambda}R_{\eta_\lambda}\xi\right)\overline{\hat{b}^{(\mu)}\left(A_{\alpha,t_\mu}R_{\eta_\mu}\xi\right)}\right)\right| \\
&\quad\lesssim \min\left\{1,t_\lambda(1+r)\right\}^M\left(1+t_\lambda r\right)^{-N_1}\left(1+t_\lambda^\alpha r|\sin(\varphi+\eta_\lambda)|\right)^{-N_2} \\
&\qquad\cdot\min\left\{1,t_\mu(1+r)\right\}^M\left(1+t_\mu r\right)^{-N_1}\left(1+t_\mu^\alpha r|\sin(\varphi+\eta_\mu)|\right)^{-N_2}
\end{aligned} \tag{2.46}$$

which, using Lemma 2.5.4 with $K=N_2$, implies the desired statement.

To prove (2.46), we use induction in $k$, namely we show that if we have two functions $a^{(\lambda)},b^{(\mu)}$ satisfying (2.5) for $L,M,N_1,N_2$, then the expression

$$\mathcal{L}_{\lambda,\mu}\left(\hat{a}^{(\lambda)}\left(A_{\alpha,t_\lambda}R_{\eta_\lambda}\xi\right)\overline{\hat{b}^{(\mu)}\left(A_{\alpha,t_\mu}R_{\eta_\mu}\xi\right)}\right)$$

can be written as a finite linear combination of terms of the form

$$\hat{c}^{(\lambda)}\left(A_{\alpha,t_\lambda}R_{\eta_\lambda}\xi\right)\overline{\hat{d}^{(\mu)}\left(A_{\alpha,t_\mu}R_{\eta_\mu}\xi\right)}$$

with $c^{(\lambda)},d^{(\mu)}$ satisfying (2.5) and $L$ replaced by $L-2$, see Lemma 2.5.6. Iterating this argument we can establish that for $k\le L/2$

$$\mathcal{L}^k_{\lambda,\mu}\left(\hat{a}^{(\lambda)}\left(A_{\alpha,t_\lambda}R_{\eta_\lambda}\xi\right)\overline{\hat{b}^{(\mu)}\left(A_{\alpha,t_\mu}R_{\eta_\mu}\xi\right)}\right) \tag{2.47}$$

can be expressed as a finite linear combination of terms of the form

$$\hat{c}^{(\lambda)}\left(A_{\alpha,t_\lambda}R_{\eta_\lambda}\xi\right)\overline{\hat{d}^{(\mu)}\left(A_{\alpha,t_\mu}R_{\eta_\mu}\xi\right)} \tag{2.48}$$

with

$$\left|\hat{c}^{(\lambda)}(\xi)\right|\lesssim\min\left\{1,t_\lambda+|\xi_1|+t_\lambda^{1-\alpha}|\xi_2|\right\}^M\langle|\xi|\rangle^{-N_1}\langle\xi_2\rangle^{-N_2}, \tag{2.49}$$

and an analogous estimate for $\hat{d}^{(\mu)}$. Combining (2.48) and (2.49), we obtain that $|(2.47)|$ can – up to a constant – be upperbounded by the product of

$$\min\left\{1,t_\lambda+\left|\left[A_{\alpha,t_\lambda}R_{\eta_\lambda}\xi\right]_1\right|+t_\lambda^{1-\alpha}\left|\left[A_{\alpha,t_\lambda}R_{\eta_\lambda}\xi\right]_2\right|\right\}^M\left\langle\left|A_{\alpha,t_\lambda}R_{\eta_\lambda}\xi\right|\right\rangle^{-N_1}\left\langle\left[A_{\alpha,t_\lambda}R_{\eta_\lambda}\xi\right]_2\right\rangle^{-N_2}$$

and

$$\min\left\{1,t_\mu+\left|\left[A_{\alpha,t_\mu}R_{\eta_\mu}\xi\right]_1\right|+t_\mu^{1-\alpha}\left|\left[A_{\alpha,t_\mu}R_{\eta_\mu}\xi\right]_2\right|\right\}^M\left\langle\left|A_{\alpha,t_\mu}R_{\eta_\mu}\xi\right|\right\rangle^{-N_1}\left\langle\left[A_{\alpha,t_\mu}R_{\eta_\mu}\xi\right]_2\right\rangle^{-N_2}.$$

Transforming this inequality into polar coordinates as in (2.6) yields (2.46). This finishes the proof. □

### 2.5.4 Actual Proof of Theorem 2.2.2

We now have all the ingredients to prove Theorem 2.2.2. By our assumptions on the order $(L,M,N_1,N_2)$, there exist $\widetilde{N}_1$ and $\widetilde{N}_2$ such that $N_1\ge\widetilde{N}_1\ge N+\frac{1+\alpha}{2}$ and $N_2\ge\widetilde{N}_2\ge N+\frac{1+\alpha}{2}$ and $M>\widetilde{N}_1+\widetilde{N}_2-2$. The systems $\{m_\lambda\}_\lambda$ and $\{p_\mu\}_\mu$ are also $\alpha$-molecules of order $(L,M,\widetilde{N}_1,\widetilde{N}_2)$, satisfying the assumptions of the theorem. Thus, we can without loss of generality assume the additional condition $M>N_1+N_2-2$.

To keep the notation simple, we assume that $\eta_\lambda = 0$ and define $t_0 := \max\{t_\lambda, t_\mu\}$. Further, we set

$$\delta x := x_\lambda - x_\mu, \quad \delta\eta := \eta_\lambda - \eta_\mu.$$

By definition, we can write

$$m_\lambda(\cdot) = t_\lambda^{-\frac{1+\alpha}{2}} a^{(\lambda)} \left(A^{-1}_{\alpha,t_\lambda} R_{\eta_\lambda}(\cdot - x_\lambda)\right), \quad p_\mu(\cdot) = t_\mu^{-\frac{1+\alpha}{2}} b^{(\mu)} \left(A^{-1}_{\alpha,t_\mu} R_{\eta_\mu}(\cdot - x_\mu)\right),$$

where both $a^{(\lambda)}$ and $b^{(\mu)}$ satisfy (2.5). We have the equality

$$\begin{aligned}
\langle m_\lambda, p_\mu\rangle &= \langle \hat{m}_\lambda, \hat{p}_\mu\rangle \\
&= (t_\lambda t_\mu)^{\frac{1+\alpha}{2}} \int_{\mathbb{R}^2} \hat{a}^{(\lambda)} \left(A_{\alpha,t_\lambda} R_{\eta_\lambda}\xi\right) \overline{\hat{b}^{(\mu)} \left(A_{\alpha,t_\mu} R_{\eta_\mu}\xi\right)} \exp\left(-2\pi i\xi\cdot\delta x\right) d\xi \qquad (2.50) \\
&= (t_\lambda t_\mu)^{\frac{1+\alpha}{2}} \int_{\mathbb{R}^2} \mathcal{L}^k_{\lambda,\mu} \left(\hat{a}^{(\lambda)} \left(A_{\alpha,t_\lambda} R_{\eta_\lambda}\xi\right) \overline{\hat{b}^{(\mu)} \left(A_{\alpha,t_\mu} R_{\eta_\mu}\xi\right)}\right) \mathcal{L}^{-k}_{\lambda,\mu} \left(\exp\left(-2\pi i\xi\cdot\delta x\right)\right) d\xi,
\end{aligned}$$

where $\mathcal{L}_{\lambda,\mu}$ is the symmetric differential operator (acting on the frequency variable) defined in (2.43). Note that by assumption $N_1 > 1$ and thus the boundary terms vanish due to the decay properties of $\hat{a}^{(\lambda)}$ and $\hat{b}^{(\mu)}$ as well as their derivatives.

We have

$$\mathcal{L}^{-k}_{\lambda,\mu} \left(\exp\left(-2\pi i\xi\cdot\delta x\right)\right) = \left(1 + t_0^{-2\alpha}|\delta x|^2 + \frac{t_0^{-2}}{1 + t_0^{-2(1-\alpha)}|\delta\eta|}\langle e_\lambda, \delta x\rangle^2\right)^{-k} \exp\left(-2\pi i\xi\cdot\delta x\right), \qquad (2.51)$$

where $e_\lambda = (\cos(\eta_\lambda), -\sin(\eta_\lambda))$ denotes the unit vector pointing in the direction described by the angle $\eta_\lambda$. By Lemma 2.5.7 and for $k \le \frac{L}{2}$, we have the inequality

$$\mathcal{L}^k_{\lambda,\mu} \left(\hat{a}^{(\lambda)} \left(A_{\alpha,t_\lambda} R_{\eta_\lambda}\xi\right) \overline{\hat{b}^{(\mu)} \left(A_{\alpha,t_\mu} R_{\eta_\mu}\xi\right)}\right) \lesssim S_{\lambda,M-N_2,N_1,N_2}(\xi) S_{\mu,M-N_2,N_1,N_2}(\xi).$$

Then, by (2.50) and (2.51) it follows that

$$\begin{aligned}
|\langle m_\lambda, p_\mu\rangle| \;&\lesssim\; (t_\lambda t_\mu)^{\frac{1+\alpha}{2}} \int_{\mathbb{R}^2} S_{\lambda,M-N_2,N_1,N_2}(\xi) S_{\mu,M-N_2,N_1,N_2}(\xi) d\xi \\
&\quad \cdot \left(1 + t_0^{-2\alpha}|\delta x|^2 + \frac{t_0^{-2}}{1 + t_0^{-2(1-\alpha)}|\delta\eta|}\langle e_\lambda, \delta x\rangle^2\right)^{-k}
\end{aligned}$$

for all $k \le \frac{L}{2}$. Now we can use Lemma 2.5.5 and the fact that $L \ge 2N$ to establish that

$$\begin{aligned}
|\langle m_\lambda, p_\mu\rangle| &\lesssim \max\left\{\frac{t_\lambda}{t_\mu}, \frac{t_\mu}{t_\lambda}\right\}^{-N} \left(1 + t_0^{-2(1-\alpha)}|\delta\eta|^2\right)^{-N} \\
&\qquad \cdot \left(1 + t_0^{-2\alpha}|\delta x|^2 + \frac{t_0^{-2}}{1 + t_0^{-2(1-\alpha)}|\delta\eta|}\langle e_\lambda, \delta x\rangle^2\right)^{-N} \\
&\le \max\left\{\frac{t_\lambda}{t_\mu}, \frac{t_\mu}{t_\lambda}\right\}^{-N} \left(1 + t_0^{-2(1-\alpha)}|\delta\eta|^2 + t_0^{-2\alpha}|\delta x|^2 + \frac{t_0^{-2}}{1 + t_0^{-2(1-\alpha)}|\delta\eta|}\langle e_\lambda, \delta x\rangle^2\right)^{-N} \\
&= \omega_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu))^{-N}.
\end{aligned}$$

This proves the desired statement. $\square$

# Chapter 3

# Examples of $\alpha$-Molecules in $L_2(\mathbb{R}^2)$

In this chapter we will fill the abstract notion of $\alpha$-molecules from Definition 2.1.3 with life by presenting a few prominent examples of specific $\alpha$-molecule systems. Undoubtedly, the most natural examples are given by curvelet systems since the construction of those served as a guiding principle for the $\alpha$-molecule definition. But, as we will see in this chapter, also wavelet systems, ridgelet systems and even shear-based constructions fit into the framework. Two important subclasses of $\alpha$-molecule systems, namely $\alpha$-curvelet and $\alpha$-shearlet molecules, are distinguished, and we will show that their associated parametrizations are consistent in a suitable sense.

## 3.1 Continuous $\alpha$-Curvelets

We begin our exposition with a prototypical instance of a continuous $\alpha$-molecule frame for $L_2(\mathbb{R}^2)$. Guided by the construction of the parabolically scaled curvelets in [11], we will construct a continuous Parseval frame of $\alpha$-curvelets for every $\alpha \in [0,1]$. The obtained systems, denoted by $\mathfrak{C}_\alpha$, are band-limited and based on a specific tiling of the frequency domain.

To realize this tiling, let us first define two radial functions $U, U_1 \in C_c^\infty([0,\infty))$ which shall be nonnegative and satisfy

$$\operatorname{supp} U \subset (\tfrac{1}{2}, 2) \quad \text{and} \quad \operatorname{supp} U_1 \subset [0,2)\,.$$

Further, they shall fulfill the continuous Calderón condition

$$U_1^2(r) + \int_0^1 U(rt)^2 \frac{dt}{t} = 1 \quad \text{for all } r \geq 0\,. \tag{3.1}$$

Next, let us take a non-negative angular function $V \in C_c^\infty([-\pi,\pi])$ with the property

$$\operatorname{supp} V \subset (-1,1) \quad \text{and} \quad \int_{-1}^1 V(\eta)^2 \, d\eta = 1\,.$$

Further, for convenience, let us also introduce the constant function $V_1 : [-\pi,\pi] \to \{\frac{1}{\sqrt{2\pi}}\}$.

Now we are ready to define the functions $W_{\eta,t} \in C_c^\infty(\mathbb{R}^2)$, where $\eta \in \mathbb{T} = [0,2\pi)$ and $t \in (0,1]$, which correspond to the desired frequency tiling. Using polar coordinates $\xi(r,\phi) = (r\cos(\phi), r\sin(\phi)) \in \mathbb{R}^2$ with $r \in \mathbb{R}_0^+ := [0,\infty)$ and $\phi \in [0,2\pi)$, we put

$$\begin{aligned} W_{\eta,1}(\xi(r,\phi)) &:= U_1(r)V_1(\{\phi\}_{2\mathbf{T}}) = \frac{1}{\sqrt{2\pi}}U_1(r) &&, \eta \in \mathbb{T},\, t=1, \\ W_{\eta,t}(\xi(r,\phi)) &:= U(tr)V(t^{\alpha-1}\{\phi+\eta\}_{2\mathbf{T}}) &&, \eta \in \mathbb{T},\, 0<t<1. \end{aligned} \tag{3.2}$$

Here $\{\cdot\}_{2\mathbf{T}}$ denotes the projective bracket (2.9) for the interval $2\mathbf{T} = [-\pi, \pi)$.

For $t = 1$ the functions $W_{\eta,t}$ are supported in a closed ball around the origin. If $0 < t < 1$ their support is a wedge-like tile whose position is determined by $\eta$ and $t$. The induced tiling resembles that of a dicrete $\alpha$-curvelet frame depicted in Figure 3.1.

From $W_{\eta,t}$ we obtain the *$\alpha$-curvelets* $\psi_{x,\eta,t} \in L_2(\mathbb{R}^2)$ by defining on the Fourier side

$$\widehat{\psi}_{x,\eta,t}(\cdot) := t^{(1+\alpha)/2} W_{\eta,t}(\cdot) \exp(-2\pi i \langle x, \cdot \rangle), \quad (x, \eta, t) \in \mathbb{X},$$

with indices from the *curvelet domain*

$$\mathbb{X} := \mathbb{R}^2 \times \mathbb{T} \times (0, 1].$$

Note that, due to the normalization factor, the $\alpha$-curvelets $\psi_{x,\eta,t}$ are $L_2$-normalized. Further, they are Schwartz functions, i.e., $\psi_{x,\eta,t} \in \mathcal{S}(\mathbb{R}^2)$. Due to the lack of symmetry of $W_{\eta,t}$, however, they are not real-valued as the discrete parabolic curvelets in [15].

On the spatial side, the $\alpha$-curvelets can be represented in the form

$$\psi_{x,\eta,t}(\cdot) = \psi_{0,0,t}(R_\eta(\cdot - x)) \quad \text{with} \quad \widehat{\psi}_{0,0,t} = t^{(1+\alpha)/2} W_{0,t}, \quad (x, \eta, t) \in \mathbb{X}, \tag{3.3}$$

where $\mathbb{R}_\eta$ is the rotation matrix (2.3) given by

$$R_\eta = \begin{pmatrix} \cos(\eta) & -\sin(\eta) \\ \sin(\eta) & \cos(\eta) \end{pmatrix}, \quad \eta \in \mathbb{R}. \tag{3.4}$$

Altogether, the constructed $\alpha$-curvelets constitute a continuous multi-scale system in $L_2(\mathbb{R}^2)$ for which we subsequently use the notation

$$\mathfrak{C}_\alpha := \Big\{ \psi_{\mathbf{x}} : \mathbf{x} \in \mathbb{X} \Big\}.$$

This system is inhomogeneous, consisting of a high-scale and a low-scale part corresponding to parameters $0 < t < 1$ and $t = 1$, respectively. The index set $\mathbb{X}$ is split accordingly into a *homogeneous* and an *inhomogeneous component*, namely

$$\mathbb{X}_0 := \mathbb{R}^2 \times \mathbb{T} \times (0, 1) \quad \text{and} \quad \mathbb{X}_1 := \mathbb{R}^2 \times \mathbb{T} \times \{1\}.$$

Each of the components $\mathbb{X}_0$ and $\mathbb{X}_1$ is equipped with the usual product topology, whereas $\mathbb{X}$ is topologized as the disconnected union of the two. With this topology, $\mathbb{X}$ becomes a locally compact Hausdorff space.

We next search for a Radon measure, i.e., a regular Borel measure, $\mu$ on $\mathbb{X}$ such that $x \mapsto \psi_x$ is weakly measurable, i.e., the assignment $x \mapsto \langle f, \psi_x \rangle$ is measurable for every $f \in L_2(\mathbb{R}^2)$, and such that the continuous Parseval identity is true, i.e.,

$$\|f|L_2\|^2 = \int\limits_{\mathbb{X}} |\langle f, \psi_{\mathbf{x}} \rangle|^2 \, d\mu(\mathbf{x}) \quad \text{for all } f \in L_2(\mathbb{R}^2). \tag{3.5}$$

Note that the weak measurability is needed for the integral in (3.5) to be well-defined.

With respect to $\mu$, the system $\mathfrak{C}_\alpha$ is then a continuous Parseval frame for $L_2(\mathbb{R}^2)$ (see [1]). In particular, the reconstruction formula

$$f = \int_{\mathbb{X}} \langle f, \psi_{\mathbf{x}} \rangle \psi_{\mathbf{x}} \, d\mu(\mathbf{x}) \tag{3.6}$$

holds in a weak sense. For signal analysis this is relevant since it means that from the data $F(\mathbf{x}) = \langle f, \psi_{\mathbf{x}} \rangle$ we can reconstruct the signal $f \in L_2(\mathbb{R}^2)$ via (3.6).

To find $\mu$, since the curvelet domain $\mathbb{X}$ does not carry a group structure, we cannot resort to a Haar measure as the canonical choice. Instead, we introduce a measure $\mu$ given by the integration

$$\int_{\mathbb{X}} F(\mathbf{x})\, d\mu(\mathbf{x}) = \int_0^1 \int_{\mathbb{T}} \int_{\mathbb{R}^2} F(x,\eta,t) \frac{dx\, d\eta\, dt}{t^3} + \int_{\mathbb{T}} \int_{\mathbb{R}^2} F(x,\eta,1)\, dx\, d\eta \tag{3.7}$$

for every $F \in C_c(\mathbb{X})$. Its restriction to the homogeneous and inhomogeneous component $\mathbb{X}_0$ and $\mathbb{X}_1$ shall be denoted by $\mu_0$ and $\mu_1$, respectively. For this choice of $\mu$ the map $\mathbf{x} \mapsto \psi_{\mathbf{x}}$ is indeed weakly measurable and the system $\mathfrak{C}_\alpha$ is a continuous Parseval frame for $L_2(\mathbb{R}^2)$, i.e., the Parseval identity (3.5) holds true.

This is proved by the following proposition.

**Proposition 3.1.1.** *The continuous $\alpha$-curvelet system $\mathfrak{C}_\alpha$ is a continuous Parseval frame for $L_2(\mathbb{R}^2)$ with respect to the measure $\mu$ given by* (3.7)*.*

*Proof.* Let $f \in L_2(\mathbb{R}^2)$. We have $\langle f, \psi_{\mathbf{x}} \rangle = \langle \hat{f}, \hat{\psi}_{\mathbf{x}} \rangle = t^{(1+\alpha)/2} (\hat{f} W_{\eta,t})^{\vee}(x)$. Further,

$$\int_{\mathbb{R}^2} |(\hat{f} W_{\eta,t})^{\vee}(x)|^2\, dx = \int_{\mathbb{R}^2} |(\hat{f} W_{\eta,t})(\xi)|^2\, d\xi.$$

We deduce

$$\begin{aligned} \int_{\mathbb{X}} |\langle f, \psi_{\mathbf{x}} \rangle|^2\, d\mu(\mathbf{x}) &= \int_0^1 \int_{\mathbb{T}} \int_{\mathbb{R}^2} |(\hat{f} W_{\eta,t})(\xi)|^2\, d\xi \frac{d\eta\, dt}{t^{2-\alpha}} + \int_{\mathbb{T}} \int_{\mathbb{R}^2} |(\hat{f} W_{\eta,1})(\xi)|^2\, d\xi\, d\eta \\ &= \int_{\mathbb{R}^2} |\hat{f}(\xi)|^2 \Big( \int_0^1 \int_{\mathbb{T}} W_{\eta,t}(\xi)^2 \frac{d\eta\, dt}{t^{2-\alpha}} + \int_{\mathbb{T}} W_{\eta,1}(\xi)^2\, d\eta \Big)\, d\xi\,. \end{aligned}$$

At this point, the Calderón condition (3.1) comes into play. For all $\xi \in \mathbb{R}^2$

$$\begin{aligned} &\int_0^1 \int_{\mathbb{T}} W_{\eta,t}(\xi)^2 \frac{d\eta\, dt}{t^{2-\alpha}} + \int_{\mathbb{T}} W_{\eta,1}(\xi)^2\, d\eta \\ &\quad = \int_0^1 \int_{\mathbb{T}} U(t|\xi|)^2 V(t^{\alpha-1}\{\phi(\xi)+\eta\}_{2\mathbf{T}})^2 \frac{d\eta\, dt}{t^{2-\alpha}} + \int_{\mathbb{T}} \frac{1}{2\pi} U_1(|\xi|)^2\, d\eta \\ &\quad = \int_0^1 t^{-(\alpha-1)} U(t|\xi|)^2 \frac{dt}{t^{2-\alpha}} + U_1(|\xi|)^2 = \int_0^1 U(t|\xi|)^2 \frac{dt}{t} + U_1(|\xi|)^2 = 1\,. \end{aligned}$$

The proof is finished. □

In the next subsection, we will see that $\mathfrak{C}_\alpha$ is a special instance of an $\alpha$-molecule system.

### 3.1.1 The Canonical Parametrization

Although the $\alpha$-curvelet frame $\mathfrak{C}_\alpha$ is not an affine construction, it is useful to consider an affine-like representation utilizing the $\alpha$-scaling matrix (2.4), namely

$$A_{\alpha,t} := \begin{pmatrix} t & 0 \\ 0 & t^\alpha \end{pmatrix}, \quad t > 0. \tag{3.8}$$

In view of (3.3), we can write

$$\psi_{x,\eta,t}(\cdot) = t^{-(1+\alpha)/2} g_t(A_{\alpha,t}^{-1} R_\eta(\cdot - x))$$

with a scale-dependent generator $g_t \in L_2(\mathbb{R}^2)$ whose Fourier transform is given by

$$\hat{g}_t(\cdot) = t^{-(1+\alpha)/2} \widehat{\psi}_{0,0,t}(A_{\alpha,t}^{-1}\cdot) = W_{0,t}(A_{\alpha,t}^{-1}\cdot). \tag{3.9}$$

For $t = 1$ we clearly have $\hat{g}_1 = W_{0,1}$. Further, for $0 < t < 1$, note that $W_{0,t}$ is obtained from the function $\widetilde{W}_{0,1}(\xi(r,\phi)) := U(r)V(\{\phi\}_{2\mathbf{T}})$ by $\alpha$-scaling in polar coordinates. Since this polar operation is closely related to the affine $\alpha$-scaling operator the generators $\hat{g}_t$ are all close to each other, namely small deviations from $\widetilde{W}_{0,1}$.

From this heuristic consideration, it is plausible that the continuous $\alpha$-curvelet frame $\mathfrak{C}_\alpha$ constitutes a system of $\alpha$-molecules. This will indeed be shown in Proposition 3.1.3 below. The associated parametrization takes a particularly simple form and is called the *canonical parametrization.*

**Definition 3.1.2.** The *canonical parametrization* $(\mathbb{X}, \Phi_\iota)$ is the pair consisting of the curvelet domain $\mathbb{X}$ given by

$$\mathbb{X} := \mathbb{R}^2 \times \mathbb{T} \times (0,1]$$

and the canonical embedding $\Phi_\iota : \mathbb{X} \to \mathbb{P}$ into the parameter space $\mathbb{P} = \mathbb{R}^2 \times \mathbb{T} \times \mathbb{R}_+$, i.e., the map

$$\Phi_\iota \ : \ (x,\eta,t) \mapsto (x,\eta,t).$$

With this definition, we can now prove that the continuous $\alpha$-curvelet frame $\mathfrak{C}_\alpha$ is a system of $\alpha$-molecules.

**Proposition 3.1.3.** *Let $\alpha \in [0,1]$. The continuous $\alpha$-curvelet frame $\mathfrak{C}_\alpha$ is a system of $\alpha$-molecules of order $(\infty,\infty,\infty,\infty)$ with respect to the canonical parametrization $(\mathbb{X}, \Phi_\iota)$.*

*Proof.* We need to show that the generators $g_t$ defined in (3.9) satisfy (2.5) for arbitrary orders $(L, M, N_1, N_2) \in \mathbb{N}_0^4$. From

$$\operatorname{supp} \hat{\psi}_{0,0,t} \subset [-2t^{-1}, 2t^{-1}] \times [-2t^{-\alpha}, 2t^{\alpha}], \quad t \in (0,1],$$

it follows that

$$\operatorname{supp} \hat{g}_t \subset [-2,2]^2 \quad \text{for all } t \in (0,1].$$

Next, if $t \in (0,1)$, we observe that the functions $\hat{\psi}_{0,0,t}$ vanish on the squares $[-\frac{1}{4}t^{-1}, \frac{1}{4}t^{-1}]^2$, which implies

$$\hat{g}_t(\xi) = 0 \quad \text{for } \xi \in [-\tfrac{1}{4}, \tfrac{1}{4}] \times [-\tfrac{1}{4}t^{\alpha-1}, \tfrac{1}{4}t^{\alpha-1}].$$

The derivatives $\partial^\rho \hat{g}_t$ are well-defined for all $\rho = (\rho_1, \rho_2) \in \mathbb{N}_0^2$ since $\hat{g}_t \in C^\infty(\mathbb{R}^2)$, and they are subject to the same support conditions as the functions $\hat{g}_t$. Further, we have

$$\sup_{\xi \in \mathbb{R}^2} |\partial_1^{\rho_1} \partial_2^{\rho_2} W_{0,t}(\xi)| \lesssim t^{\rho_1} t^{\alpha\rho_2} \quad \text{uniformly in } t \in (0,1].$$

With the chain rule we deduce

$$\sup_{\xi \in \mathbb{R}^2} |\partial^\rho \hat{g}_t(\xi)| = \sup_{\xi \in \mathbb{R}^2} |\partial_1^{\rho_1} \partial_2^{\rho_2} (W_{0,t}(A_{\alpha,t}^{-1}\cdot))(\xi)| = t^{-\rho_1} t^{-\alpha\rho_2} \sup_{\xi \in \mathbb{R}^2} |(\partial_1^{\rho_1} \partial_2^{\rho_2} W_{0,t})(A_{\alpha,t}^{-1}\xi)| \lesssim 1,$$

where the implicit constant is independent of $t \in (0,1]$. Together with the support properties of $\partial^\rho \hat{g}_t$, this uniform bound implies condition (2.5) for arbitrary orders $(L, M, N_1, N_2)$. □

Note, that the canonical parametrization $(\mathbb{X}, \Phi_\iota)$ does not depend on the parameter $\alpha \in [0,1]$ and is thus the same for all systems $\mathfrak{C}_\alpha$. Considering $\Phi_\iota : \mathbb{X} \to \mathbb{P}$ as an embedding, the curvelet domain $\mathbb{X}$ can further be viewed as a subset of the parameter space $\mathbb{P}$. Since the topology on $\mathbb{X}$ differs from the usual subspace topology imposed by $\mathbb{P}$, one has to be a little careful though with this perspective.

Let us end this subsection with the remark that the domain $\mathbb{X}$ can be considered as a natural parameter space for inhomogeneous $\alpha$-molecule systems. Given such an inhomogeneous system $\mathfrak{M}_\alpha$ with a parametrization $(\Lambda, \Phi)$ such that $\Phi(\lambda) = (x_\lambda, \eta_\lambda, t_\lambda) \in \mathbb{P}$ and $t_\lambda \le C$ for all $\lambda \in \Lambda$ and some constant $C > 0$, it is always possible to reparameterize $\mathfrak{M}_\alpha$ with a base scale not larger than 1 by using the modified parametrization map $\Phi'(\lambda) := (x_\lambda, \eta_\lambda, t_\lambda/C)$. One can prove that $\mathfrak{M}_\alpha$ is then still a system of $\alpha$-molecules with respect to the new parametrization $(\Lambda, \Phi')$.

## 3.2 $\alpha$-Curvelet Molecules

As we have already seen, the idea for the construction of curvelets is inspired by wavelets which are obtained by isotropically scaling and translating a set of generating functions. For a curvelet system, this construction principle is slightly modified to improve the directional adaptivity of the system elements at high scales. Instead of isotropic scaling, an anisotropic form of scaling is used and, as a means to adjust the orientation, rotations come into play.

This basic idea is cast into a concrete form by the notion of *$\alpha$-curvelet molecules*, a concept which unites many different curvelet-like constructions under one common roof and allows a unified treatment of such systems. The definition given below is a direct generalization of the earlier introduced curvelet molecules from [13]. As the name suggests, the anisotropic scaling is realized via the $\alpha$-scaling matrix (3.8) and, as was the case for $\alpha$-molecules, the generators have the freedom to vary as long as they obey a certain time-frequency localization.

We will only consider discrete systems which correspond to certain regular sampling grids of the continuous curvelet domain $\mathbb{X}$. Thereby the scales shall be numbered by $j \in \mathbb{N}_0$ and the distance between the different scales, specified by a real number $\sigma > 1$, shall be fixed. Further, the translational grid is assumed to be a transformation of $\mathbb{Z}^2$ via the matrix (2.35) defined by

$$M_c = \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix} \quad \text{for some fixed vector } c = (c_1, c_2) \in \mathbb{R}^2_+. \tag{3.10}$$

The resolution of the angular sampling is scale-dependent and determined at each scale $j \in \mathbb{N}_0$ by a positive real number $\omega_j \in \mathbb{R}_+$, whereby we require the sequence $(\omega_j)_{j\in\mathbb{N}_0}$ to fulfill $\omega_j \asymp \sigma^{-j(1-\alpha)}$.

For convenience, let us display the set of parameters associated to the resolution of the sampling grid in the following box,

$$\boxed{\sigma > 1\,, \quad c = (c_1, c_2) \in \mathbb{R}^2_+\,, \quad (\omega_j)_{j\in\mathbb{N}_0} \in \mathbb{R}_+^{\mathbb{N}_0} \text{ with } \omega_j \asymp \sigma^{-j(1-\alpha)}.} \tag{3.11}$$

A typical index set of a system of $\alpha$-curvelet molecules is of the form

$$\Lambda^c := \Big\{ (j, \ell, k) \ : \ j \in \mathbb{N}_0,\, \ell \in \mathbb{L}_j,\, k \in \mathbb{Z}^2 \Big\}, \tag{3.12}$$

where the index $j$ corresponds to the scale, $\ell$ to the orientation, and $k$ to the spatial position of the molecules. The number of different orientations at each scale $j$ is determined by the set $\mathbb{L}_j \neq \emptyset$ which is of the form

$$\mathbb{L}_j \subset \mathbb{Z} \quad \text{such that} \quad \max\Big\{|\ell| : \ell \in \mathbb{L}_j\Big\} \leq L_j \tag{3.13}$$

for a sequence $(L_j)_{j\in\mathbb{N}_0}$ of nonnegative integers satisfying $L_j \lesssim \sigma^{j(1-\alpha)}$.

The definition of a system of $\alpha$-curvelet molecules is then as follows, whereby we need the matrices (3.4), (3.8), and (3.10).

**Definition 3.2.1.** Let $\alpha \in [0,1]$ and let $L, M, N_1, N_2 \in \mathbb{N}_0 \cup \{\infty\}$. Further let the sampling parameters (3.11) be fixed. A family of functions

$$\mathfrak{M}_\alpha^c := \{m_\lambda \in L_2(\mathbb{R}^2) : \lambda \in \Lambda^c\},$$

indexed by a set $\Lambda^c$ of the form (3.12), is called a system of *$\alpha$-curvelet molecules of order* $(L, M, N_1, N_2)$ if all functions $m_\lambda = m_{j,\ell,k}$ are obtained via

$$m_{j,\ell,k}(\cdot) := \sigma^{(1+\alpha)j/2} g_{j,\ell,k}\left(A_{\alpha,\sigma^j} R_{\ell\omega_j} \cdot - M_c k\right)$$

from corresponding generators $g_{j,\ell,k} \in L_2(\mathbb{R}^2)$ which satisfy for every $\rho \in \mathbb{N}_0^2$, $|\rho| \leq L$,

$$|\partial^\rho \hat{g}_{j,\ell,k}(\xi_1,\xi_2)| \lesssim \min\Big\{1, \sigma^{-j} + |\xi_1| + \sigma^{-j(1-\alpha)}|\xi_2|\Big\}^M \cdot \langle|\xi|\rangle^{-N_1} \cdot \langle \xi_2\rangle^{-N_2}. \tag{3.14}$$

The implicit constant is required to be uniform over all $\lambda \in \Lambda^c$ and $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$. If a control parameter equals $\infty$, the respective quantity can be chosen arbitrarily large in (3.14).

As for $\alpha$-molecules, the quantities $M, L, N_1, N_2$ specify almost vanishing moment, localization, and frequency decay properties of the respective functions. The following proposition shows that Definition 3.2.1 is compatible with an earlier notion of curvelet molecules considered in [13].

**Proposition 3.2.2** ([62])**.** *Curvelet molecules of regularity $R \in \mathbb{N}_0$, as defined in [13], are $\frac{1}{2}$-curvelet molecules of order $(\infty, \infty, R/2, R/2)$.*

*Proof.* For the proof we refer to [62]. □

The concept of $\alpha$-curvelet molecules comprises many curvelet-like constructions. In particular, the classic curvelets are included, as we will see in Subsection 3.2.3. Before we turn to concrete examples though, let us show that $\alpha$-curvelet molecules are a special class of $\alpha$-molecules.

### 3.2.1 The $\alpha$-Curvelet Parametrization

Systems of $\alpha$-curvelet molecules constitute a special class of discrete $\alpha$-molecule systems. They are characterized by a corresponding class of parametrizations, called *$\alpha$-curvelet parametrizations.*

**Definition 3.2.3** (compare [59, Def. 3.2])**.** An *$\alpha$-curvelet parametrization* $(\Lambda^c, \Phi^c)$ consists of an $\alpha$-curvelet index set $\Lambda^c$ of the form (3.12), i.e.,

$$\Lambda^c := \Big\{(j, \ell, k) \ : \ j \in \mathbb{N}_0,\, \ell \in \mathbb{L}_j,\, k \in \mathbb{Z}^2\Big\},$$

and a mapping $\Phi^c : \Lambda^c \to \mathbb{P}$ given by

$$\Phi^c \ : \ (j, \ell, k) \mapsto \big(R^{-1}_{\ell\omega_j} A^{-1}_{\alpha,\sigma^j} M_c k, (\ell\omega_j)_{2\pi}, \sigma^{-j}\big)$$

with fixed sampling parameters $\sigma > 1$, $c = (c_1, c_2) \in \mathbb{R}^2_+$, $(\omega_j)_{j\in\mathbb{N}_0} \subset \mathbb{R}_+$ as in (3.11).

Comparing the definition of $\alpha$-molecules and the definition of $\alpha$-curvelet molecules, the following proposition is self-evident.

**Proposition 3.2.4.** *Every system of $\alpha$-curvelet molecules of order $(L, M, N_1, N_2)$ constitutes a system of $\alpha$-molecules of the same order with respect to a corresponding $\alpha$-curvelet parametrization, and vice versa.*

*Proof.* This is obvious from the definitions. □

Whereas Proposition 3.2.4 may not come as a surprise, we will see later in Section 3.3 that the concept of $\alpha$-molecules also comprises shear-based constructions. In particular, $\alpha$-shearlet systems and their more general siblings *$\alpha$-shearlet molecules* are included, as proved in Proposition 3.3.6.

### 3.2.2 A Sufficient Frame Condition for $\alpha$-Curvelet Molecules

Observe that the frame criterion from Theorem 2.4.1 can be applied to systems of $\alpha$-curvelet molecules, provided that the corresponding generators do not vary with the translation index. Indeed, the $\alpha$-curvelet index set $\Lambda^c$ has the required structure

$$\Lambda^c = \Delta^c \times \mathbb{Z}^2 \quad \text{with } \Delta^c := \Big\{(j, \ell) \ : \ j \in \mathbb{N}_0, \ell \in \mathbb{L}_j\Big\}.$$

Further, the parametrization map

$$\Phi^c : \Lambda^c \to \mathbb{P} \ , \ (j, \ell, k) \mapsto \big(R^{-1}_{\ell\omega_j} A^{-1}_{\alpha,\sigma^j} M_c k, (\ell\omega_j)_{2\pi}, \sigma^{-j}\big)$$

is of the required form (2.34) with the matrix $T_\mu$ chosen as the identity for each $\mu \in \Delta^c$.

Let us now assume that $\{m_\lambda\}_{\lambda\in\Lambda^c}$ is a system of $\alpha$-curvelet molecules whose generators $\{g_\lambda\}_{\lambda\in\Lambda^c}$ satisfy $g_\lambda = g_\mu$ for all $\lambda = (\mu, k) \in \Lambda^c$. Then we can define associated correlation functions $\Phi^c : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ and $\Gamma^c : \mathbb{R}^2 \to \mathbb{R}$ as

$$\begin{aligned} \Phi^c(\xi, \omega) &:= \sum_{\mu\in\Delta^c} |\hat{g}_\mu(A_{\alpha,t_\mu} R_{\eta_\mu}\xi)||\hat{g}_\mu(A_{\alpha,t_\mu} R_{\eta_\mu}\xi + \omega)|, \\ \Gamma^c(\omega) &:= \operatorname*{ess\,sup}_{\xi\in\mathbb{R}^2} \Phi^c(\xi, \omega). \end{aligned}$$

For $c \in \mathbb{R}^2_+$, we further define

$$R^c(c) := \sum_{m\in\mathbb{Z}^2\setminus\{0\}} \big[\Gamma^c(M_c^{-1}m)\Gamma^c(-M_c^{-1}m)\big]^{1/2}.$$

Now we can formulate the following spin-off of Theorem 2.4.1 in the curvelet setting.

**Theorem 3.2.5.** *Let $\{m_\lambda\}_{\lambda\in\Lambda^c}$ be a system of $\alpha$-curvelet molecules with generators $\{g_\lambda\}_{\lambda\in\Lambda^c}$ such that $g_\lambda = g_{\mu,k}$ do not vary with $k \in \mathbb{Z}^2$. Further, let*

$$L^c_{\inf} := \operatorname*{ess\,inf}_{\xi\in\mathbb{R}^2} \Phi^c(\xi,0) \quad \textit{and} \quad L^c_{\sup} := \Gamma^c(0) = \operatorname*{ess\,sup}_{\xi\in\mathbb{R}^2} \Phi^c(\xi,0)$$

*and assume that*

$$R^c(c) < L^c_{\inf} \le L^c_{\sup} < \infty,$$

*where $c \in \mathbb{R}^2_+$ denotes the parameter in (3.11) associated with the density of the translation grid. Then $\{m_\lambda\}_{\lambda\in\Lambda^c}$ constitutes a frame for $L_2(\mathbb{R}^2)$ with frame bounds $A, B > 0$ satisfying*

$$\frac{L^c_{\inf} - R^c(c)}{|\det M_c|} \le A \le B \le \frac{L^c_{\sup} + R^c(c)}{|\det M_c|}.$$

*Proof.* This is a direct corollary of Theorem 2.4.1. □

Finally, after all these considerations concerning $\alpha$-curvelet molecules on a general level, let us now provide a concrete $\alpha$-curvelet construction.

### 3.2.3 Discrete $\alpha$-Curvelet Systems

Similar to the construction of the continuous $\alpha$-curvelet frame $\mathfrak{C}_\alpha$ in Section 3.1, we now construct a prototypical example of a discrete $\alpha$-curvelet frame, subsequently denoted by $\mathfrak{C}^\bullet_\alpha$. The easiest path towards $\mathfrak{C}^\bullet_\alpha$ seems to be to just sample the continuous frame $\mathfrak{C}_\alpha$. This approach has the disadvantage, however, that it is not clear how dense the sampling must be for the resulting subsystem to still form a frame. Therefore we choose a direct approach which offers more control over the construction and even allows to obtain a Parseval frame.

The subsequent construction of $\mathfrak{C}^\bullet_\alpha$ is a slight modification of the construction of $\mathfrak{C}_\alpha$. As for $\mathfrak{C}_\alpha$, it is band-limited and again it starts by defining radial and angular components of a suitable partition of the frequency domain. To obtain the Parseval property in the discrete setting the Calderón condition (3.1) needs to be adapted, however (see (3.16) below).

For the construction of the radial functions, we use two $C^\infty$-functions $U_0 : \mathbb{R}^+_0 \to [0,1]$ and $U : \mathbb{R}^+_0 \to [0,1]$ on the ray $\mathbb{R}^+_0 := [0,\infty)$ with the properties

$$\begin{aligned} \operatorname{supp} U_0 \subseteq C\cdot[0,\tau_2], &\quad U_0 \equiv 1 \text{ on } C\cdot[0,\tau_1], \\ \operatorname{supp} U \subseteq C\cdot[2^{-1}\tau_1,\tau_2], &\quad U \equiv 1 \text{ on } C\cdot[2^{-1}\tau_2,\tau_1], \end{aligned} \tag{3.15}$$

where $1 < \tau_1 < \tau_2 < 2$ and $C > 0$ are fixed parameters. We then generate functions $U_j$, $j \ge 1$, from $U$ via

$$U_j(\cdot) := U(2^{-j}\cdot).$$

Altogether, we thus obtain a family $\{U_j\}_{j\in\mathbb{N}_0}$ of radial functions $U_j \in C^\infty(\mathbb{R}^+_0,[0,1])$. By suitably choosing $U_0$ and $U$ we can further ensure the discrete Calderón condition

$$\sum_{j\ge 0} U_j^2(r) = 1 \quad \text{for all } r \in \mathbb{R}^+_0. \tag{3.16}$$

The details of such a construction have been carried out for example in [103, Lem. 2.2].

Next, we define angular functions $V_{j,\ell} : \mathbb{S}^1 \to [0,1]$ on the unit circle $\mathbb{S}^1 \subset \mathbb{R}^2$, where $j \in \mathbb{N}_0$ and the index $\ell$ runs through $0, \dots, L_j - 1$ with

$$L_j := 2^{\lfloor j(1-\alpha) \rfloor}, \quad j \in \mathbb{N}_0.$$

The corresponding index set is denoted by

$$\mathbb{J} := \Big\{ J = (j,\ell) \ : \ j \in \mathbb{N}_0, \ell \in \{0, \dots, L_j - 1\} \Big\}.$$

Its elements $(j,\ell) \in \mathbb{J}$ can be interpreted as *scale-angle pairs*, with the index $j$ standing for a scale and $\ell$ for an orientation. To simplify the notation, we will often use the capital letter $J$ for a pair $(j,\ell) \in \mathbb{J}$. In this context, $|J|$ shall then denote the corresponding scale variable $j$, i.e., $|J| = j$ for $J = (j,\ell)$.

To construct the family $\{V_J\}_{J \in \mathbb{J}}$, we begin with a single function $\widetilde{V} \in C^\infty(\mathbb{R},[0,1])$ satisfying

$$\operatorname{supp} \widetilde{V} \subseteq [-\tfrac{3}{4}\pi, \tfrac{3}{4}\pi], \quad \widetilde{V} \equiv 1 \text{ on } [-\tfrac{\pi}{4}, \tfrac{\pi}{4}], \quad \textstyle\sum_{k \in \mathbb{Z}} \widetilde{V}^2(\cdot - k\pi) \equiv 1.$$

By rescaling, we then obtain functions $\widetilde{V}_j(\cdot) := \widetilde{V}(L_j \cdot) \in C^\infty(\mathbb{R},[0,1])$ for every $j \in \mathbb{N}_0$ which in turn, via the bijection $t \mapsto e^{it}$, give rise to corresponding functions $\widetilde{V}_{j,0} \in C^\infty(\mathbb{S}^1,[0,1])$ on the unit circle. After a symmetrization, we arrive at the functions $V_{j,0} \in C^\infty(\mathbb{S}^1,[0,1])$ given by

$$V_{j,0}(\xi) := \widetilde{V}_{j,0}(\xi) + \widetilde{V}_{j,0}(-\xi), \quad \xi \in \mathbb{S}^1.$$

At last, we rotate the functions $V_{j,0}$ by integer multiples of the angle

$$\omega_j := \pi L_j^{-1} = \pi 2^{-\lfloor j(1-\alpha) \rfloor}, \quad j \in \mathbb{N}_0.$$

Using the rotation matrix $R_{j,\ell} := R_{\ell\omega_j}$, this yields functions $V_{j,\ell} \in C^\infty(\mathbb{S}^1,[0,1])$ given by

$$V_{j,\ell}(\xi) := V_{j,0}(R_{j,\ell}\xi), \quad \xi \in \mathbb{S}^1.$$

The constructed family $\{V_J\}_{J \in \mathbb{J}}$ then clearly satisfies

$$\sum_{|J|=j} V_J^2(\xi) = 1 \quad \text{for all } \xi \in \mathbb{S}^1 \text{ and all } j \in \mathbb{N}_0.$$

In conjunction with (3.16), this property yields

$$\sum_{J \in \mathbb{J}} W_J^2 \equiv 1 \tag{3.17}$$

for the bivariate functions $W_J = W_{j,\ell} \in C^\infty(\mathbb{R}^2,[0,1])$, which are defined as the polar tensor products

$$W_{j,\ell}(\xi) := U_j(|\xi|_2) V_{j,\ell}(\xi/|\xi|_2), \quad \xi \in \mathbb{R}^2. \tag{3.18}$$

These functions are symmetric, i.e., $W_{j,\ell}(\xi) = W_{j,\ell}(-\xi)$ for $\xi \in \mathbb{R}^2$, and they are supported in corresponding wedges $\mathcal{W}^+_{j,\ell}[C]$, depending on the constant $C$ in (3.15). We have

$$\begin{aligned} \mathcal{W}^+_{0,0}[C] &:= \Big\{\xi \in \mathbb{R}^2 \ : \ |\xi|_2 \le C2^{j+1}\Big\} \quad \text{ and for } (j,\ell) \in \mathbb{J}\backslash\{(0,0)\} \\ \mathcal{W}^+_{j,\ell}[C] &:= \Big\{\xi \in \mathbb{R}^2 \ : \ C2^{j-1} \le |\xi|_2 \le C2^{j+1}, \quad |\langle \xi, R_{j,\ell}e_1\rangle| \ge \cos(3\omega_j/4)|\xi|_2\Big\}, \end{aligned} \tag{3.19}$$

where $e_1 = (1,0) \in \mathbb{R}^2$ denotes the first unit vector of $\mathbb{R}^2$.

Now we fix $C := \frac{1}{6\pi}$ in (3.15) such that $\mathcal{W}^+_{j,\ell}[C]$ is contained in the rectangle

$$\Xi_J := R_J^{-1}\Xi_{j,0} \quad \text{with} \quad \Xi_{j,0} := [-2^{j-1}, 2^{j-1}] \times [-2^{j\alpha-1}, 2^{j\alpha-1}]. \tag{3.20}$$

The rectangles $\Xi_{j,0}$ are of size $2^j \times 2^{j\alpha}$ and hence the Fourier system $\{u_{j,0,k}\}_{k\in\mathbb{Z}^2}$ with

$$u_{j,0,k}(\xi) := 2^{-j(1+\alpha)/2} \exp\left(2\pi i(2^{-j}k_1\xi_1 + 2^{-j\alpha}k_2\xi_2)\right), \quad \xi = (\xi_1,\xi_2) \in \mathbb{R}^2,$$

constitutes an orthonormal basis for $L_2(\Xi_{j,0})$ . Consequently, also the rotated system $\{u_{j,\ell,k}\}_{k\in\mathbb{Z}^2}$ consisting of the functions

$$u_{j,\ell,k}(\xi) := u_{j,0,k}(R_{j,\ell}\xi), \quad \xi \in \mathbb{R}^2, \tag{3.21}$$

is an orthonormal basis for $L_2(\Xi_J)$.

After this preparation, we are now ready to define the $\alpha$-curvelet system $\mathfrak{C}^\bullet_\alpha$.

**Definition 3.2.6.** Let $\alpha \in [0,1]$, and let $\{W_J\}_{J\in\mathbb{J}}$ be the family of wedge functions constructed in (3.18). Further, let $u_{j,\ell,k}$ be the functions defined in (3.21). The discrete $\alpha$-curvelet system $\mathfrak{C}^\bullet_\alpha := \{\psi_\mu\}_{\mu\in M}$ with associated index set $M := \mathbb{J} \times \mathbb{Z}^2$ shall consist of the functions $\psi_\mu = \psi_{j,\ell,k}$ given by

$$\widehat{\psi}_{j,\ell,k}(\xi) := W_{j,\ell}(\xi)u_{j,\ell,k}(\xi)\,, \quad \xi \in \mathbb{R}^2. \tag{3.22}$$

Note that $\mathfrak{C}^\bullet_\alpha$ depends on the utilized family $\{W_J\}_{J\in\mathbb{J}}$, which is not accounted for in the notation.

In contrast to the continuous $\alpha$-curvelets in $\mathfrak{C}_\alpha$, the $\alpha$-curvelets $\psi_\mu \in \mathfrak{C}^\bullet_\alpha$ are real-valued due to the symmetry of $W_{j,\ell}$. They are not strictly $L_2$-normalized, however. Their $L_2$-norms may vary slightly with the scale, but there exist fixed constants $0 < C_1 \le C_2 < \infty$ such that $C_1 \le \|\psi_\mu\|_2 \le C_2$ holds true for all $\mu \in M$.

Concerning the frame property, we have the following result.

**Lemma 3.2.7.** *Let $\alpha \in [0,1]$. The system $\mathfrak{C}^\bullet_\alpha$ given by* (3.22) *is a Parseval frame for $L_2(\mathbb{R}^2)$.*

*Proof.* The functions $W_J$ satisfy condition (3.17) wherefore

$$\|f\|_2^2 = \|\widehat{f}\|_2^2 = \sum_{J\in\mathbb{J}} \|\widehat{f}W_J\|_2^2 \quad \text{for every } f \in L_2(\mathbb{R}^2).$$

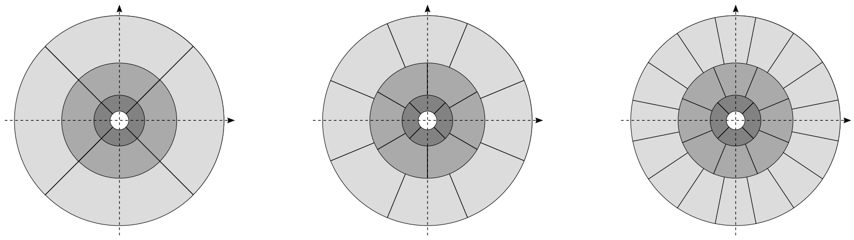
Since $\operatorname{supp}(\widehat{f}W_J) \subseteq \Xi_J$ and since $\{u_{J,k}\}_{k\in\mathbb{Z}^2}$ is an orthonormal basis of $L_2(\Xi_J)$ we have the orthogonal expansion $\widehat{f}W_J = \sum_k \langle \widehat{f}W_J, u_{J,k}\rangle u_{J,k}\mathcal{X}_{\Xi_J}$. The proof is finished by the following equality,

$$\|\widehat{f}W_J\|_2^2 = \sum_{k\in\mathbb{Z}^2} |\langle \widehat{f}W_J, u_{J,k}\rangle|^2 = \sum_{k\in\mathbb{Z}^2} |\langle \widehat{f}, W_J u_{J,k}\rangle|^2 = \sum_{k\in\mathbb{Z}^2} |\langle \widehat{f}, \widehat{\psi}_{J,k}\rangle|^2 = \sum_{k\in\mathbb{Z}^2} |\langle f, \psi_{J,k}\rangle|^2.$$

$\square$

Lemma 3.2.7 shows that $\mathfrak{C}_\alpha^\bullet$ constitutes a Parseval frame for $L_2(\mathbb{R}^2)$ for every $\alpha \in [0,1]$. Hence, we now have a whole scale of discrete Parseval frames of $\alpha$-curvelets available for $L_2(\mathbb{R}^2)$ which interpolates between wavelet systems for $\alpha = 1$ and ridgelet systems for $\alpha = 0$. The induced frequency tiling for different $\alpha \in [0,1]$ is schematically depicted in Figure 3.1.

Note that for $\alpha = \frac{1}{2}$ one obtains a variation of the classic second generation curvelet frame introduced by Candès and Donoho in [15]. Historically, this frame can be considered as the first true curvelet construction. Introduced in 2002, it provably provides quasi-optimal approximation for a model class of cartoon-like functions. Its invention triggered the development of many other directional representation system, in particular the generalization to $\alpha$-curvelets.



(a) (b) (c)

Figure 3.1: Partition of the Fourier domain induced by $\alpha$-curvelets for (a): $\alpha = 1$, (b): $\alpha = 1/2$, and (c): $\alpha = 0$.

Whereas the frame of second generation curvelets slightly differs from the frame $\mathfrak{C}_{1/2}^\bullet$, they are both instances of $\frac{1}{2}$-curvelet molecules. We have the following proposition.

**Proposition 3.2.8** (compare [59, Prop. 3.3]). *The following statements hold.*

*(i) Second generation curvelets are $\frac{1}{2}$-curvelet molecules of order $(\infty,\infty,\infty,\infty)$ with parameters $\sigma = 4$ and $c = (1,1)$, $\omega_j = \pi 4^{-j/2}$, and $L_j = 2^j$.*

*(ii) For each $\alpha \in [0,1]$, the discrete $\alpha$-curvelet frame $\mathfrak{C}_\alpha^\bullet$ is a system of $\alpha$-curvelet molecules of order $(\infty,\infty,\infty,\infty)$ with parameters $\sigma = 2$, $c = (1,1)$, $\omega_j = \pi 2^{-\lfloor j(1-\alpha)\rfloor}$, and $L_j = 2^{\lfloor j(1-\alpha)\rfloor}$.*

*Proof.* (i) was proved in [62].
(ii) In spatial domain, the $\alpha$-curvelets $\psi_\mu = \psi_{j,\ell,k} \in \mathfrak{C}_\alpha^\bullet$ have the representation

$$\psi_{j,\ell,k}(x) = \psi_{j,0,0}\left(R_{j,\ell}\left(x - x_{j,\ell,k}\right)\right) \quad \text{with} \quad x_{j,\ell,k} := R_{j,\ell}^{-1} A_j^{-1} k, \tag{3.23}$$

where $R_{j,\ell} := R_{\ell\omega_j}$ and $A_j := A_{\alpha,2^j}$ is a dyadic $\alpha$-scaling matrix, i.e.,

$$R_{j,\ell} = \begin{pmatrix} \cos(\ell\omega_j) & -\sin(\ell\omega_j) \\ \sin(\ell\omega_j) & \cos(\ell\omega_j) \end{pmatrix} \quad \text{and} \quad A_j = \begin{pmatrix} 2^j & 0 \\ 0 & 2^{j\alpha} \end{pmatrix}. \tag{3.24}$$

Further, introducing the functions

$$a_j := 2^{-j(1+\alpha)/2} \psi_{j,0,0}(A_j^{-1} \cdot), \quad j \in \mathbb{N}_0, \tag{3.25}$$

we can write $\psi_{j,\ell,k}$ in the form

$$\psi_{j,\ell,k}(x) = 2^{j(1+\alpha)/2} a_j \left(A_j R_{j,\ell}\left(x - x_{j,\ell,k}\right)\right) = 2^{j(1+\alpha)/2} a_j \left(A_j R_{j,\ell} x - k\right). \tag{3.26}$$

On the Fourier side the functions (3.25) have the form

$$\widehat{a}_j = 2^{j(1+\alpha)/2} \widehat{\psi}_{j,0,0}(A_j \cdot) = W_{j,0}(A_j \cdot).$$

Since $\operatorname{supp} W_{j,0} \subseteq \mathcal{W}^+_{j,0}[1/6\pi]$ (see (3.19)) and

$$\mathcal{W}^+_{j,0}[1/6\pi] \subseteq [-2^{j-1}, 2^{j-1}] \times [-2^{j\alpha-1}, 2^{j\alpha-1}] = \Xi_{j,0},$$

this implies

$$\operatorname{supp} \widehat{a}_j \subseteq [-2^{-1}, 2^{-1}] \times [-2^{-1}, 2^{-1}] = \Xi_{0,0}. \tag{3.27}$$

Further, if $j > 0$ the function $\widehat{\psi}_{j,0,0}$ vanishes on the square $[-2^{j-7}, 2^{j-7}]^2$. Consequently, the associated function $\widehat{a}_j$ vanishes on $[-2^{-7}, 2^{-7}] \times \left(2^{j(1-\alpha)} \cdot [-2^{-7}, 2^{-7}]\right)$.

Next, we analyze the derivatives of $\widehat{a}_j$. First observe that for fixed $\rho = (\rho_1, \rho_2) \in \mathbb{N}_0^2$ the mixed derivatives $\partial_1^{\rho_1} \partial_2^{\rho_2} W_{j,0}$ obey uniformly in $j \in \mathbb{N}_0$

$$\|\partial_1^{\rho_1} \partial_2^{\rho_2} W_{j,0}\|_\infty \lesssim 2^{-j\rho_1} 2^{-j\alpha\rho_2}.$$

With the chain rule we deduce

$$\|\partial^\rho \widehat{a}_j\|_\infty = \|\partial_1^{\rho_1} \partial_2^{\rho_2} W_{j,0}(A_j \cdot)\|_\infty = 2^{j\rho_1} 2^{j\alpha\rho_2} \|(\partial_1^{\rho_1} \partial_2^{\rho_2} W_{j,0})(A_j \cdot)\|_\infty \lesssim 1.$$

Due to $\operatorname{supp} \partial^\rho \widehat{a}_j \subseteq \operatorname{supp} \widehat{a}_j$, this estimate together with the support properties of $\widehat{a}_j$ implies (2.5). $\square$

As a consequence of this proposition, the $\alpha$-curvelet frame $\mathfrak{C}^\bullet_\alpha$ is in particular a system of $\alpha$-molecules of order $(\infty, \infty, \infty, \infty)$. Its parametrization $(M, \Phi_M)$ consists of the $\alpha$-curvelet index set $M = \mathbb{J} \times \mathbb{Z}^2$ and the parametrization map $\Phi_M$ from $M$ into the phase-space $\mathbb{P} = \mathbb{R}^2 \times \mathbb{T} \times \mathbb{R}_+$ which is given as follows

$$\Phi_M : M \to \mathbb{P}, \quad (j, \ell, k) \mapsto (x_{j,\ell,k}, \ell\omega_j, 2^{-j}) = (R^{-1}_{\ell\omega_j} A_j^{-1} k, \ell\omega_j, 2^{-j}). \tag{3.28}$$

The frame $\mathfrak{C}^\bullet_\alpha$ is a suitable anchor system for the application of the transfer principle formulated in Theorem 2.3.6. It will be used in Chapters 5 and 6 to study cartoon approximation properties of $\alpha$-molecules. Via the transfer principle, the results obtained for $\mathfrak{C}^\bullet_\alpha$ have consequences for many other $\alpha$-molecule systems. Among these are $\alpha$-curvelet constructions [15, 60], but also band-limited [81, 66, 70] as well as compactly supported [76, 73, 83] $\alpha$-shearlet systems. A general framework for discrete $\alpha$-shearlet systems is the topic of the next section.

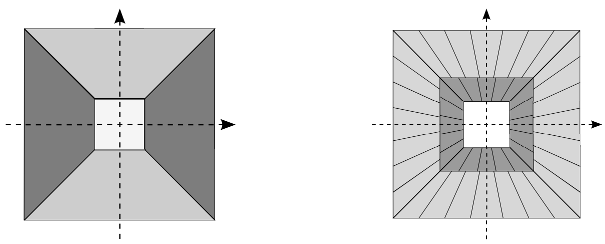
## 3.3 $\alpha$-Shearlet Molecules

The concept of *$\alpha$-shearlet molecules* can be viewed as the analogue of the concept of $\alpha$-curvelet molecules in the shearlet setting. To motivate this concept, let us first recall the basic construction principles of a cone-adapted shearlet system.

The idea for the construction of a shearlet system in general is to apply anisotropic scalings, shearings, and translations to a set of generating functions [81]. This is similar to curvelet constructions, the essential difference is the utilization of shearings instead of rotations as a means to adjust the orientation. On the one hand, this provides advantages in the discrete setting and for numerical implementations, on the other hand, the shearing operation leads to a directional bias with respect to the vertical or the horizontal coordinate axis. This bias is a disadvantage when one requires a uniform treatment of all spatial directions.

To compensate for this drawback, the concept of cone-adapted shearlet systems [66] emerged. Those are systems assembled from different shearlet subsystems, each taking care of a different coordinate direction. In the frequency domain, each subsystem correlates with a double cone aligned with one of the coordinate axes. In case of an inhomogeneous system, there is in addition a distinguished subsystem of base-scale functions corresponding to a low-frequency box.

A typical tiling of the frequency domain induced by a cone-adapted shearlet system is depicted in Figure 3.2. The cones associated with the $\varepsilon$-direction, $\varepsilon \in \{1, 2\}$, are denoted by $\mathcal{C}_\varepsilon$, and the symbol $\mathcal{R}_0$ is used for the low-frequency box.



(a) (b)

Figure 3.2: (a): The Fourier domain is partitioned into a horizontal and vertical double cone and a low-frequency box. (b): Partition of the Fourier domain induced by a cone-adapted shearlet system.

In the classic case [66], each shearlet subsystem is generated via affine transformations from a single generator. For many constructions, however, this building principle is relaxed to obtain more flexibility in the design, see for instance [70, 67]. A shearlet framework which also comprises such more general constructions is provided by the concept of shearlet molecules.

The notion first appeared in [68] and was later generalized in [62]. Like curvelet molecules, shearlet molecules incorporate the molecule idea, i.e., to allow variable generators for each function as long as those satisfy a uniform time-frequency localization. The definition of $\alpha$-shearlet molecules given here generalizes both earlier definitions.

As for $\alpha$-curvelet molecules, we only consider the discrete setting. For flexibility, we allow different samplings of the shearlet domain, however restricted to regular sampling grids. Those are specified by a set of parameters, similar to (3.11), namely

$$\boxed{\sigma > 1\,, \quad c=(c_1,c_2)\in\mathbb{R}^2_+\,, \quad (\upsilon_j)_{j\in\mathbb{N}_0}\in\mathbb{R}_+^{\mathbb{N}_0} \text{ with } \upsilon_j \asymp \sigma^{-j(1-\alpha)}.} \tag{3.29}$$

The parameters $\sigma$ and $c$ determine the scale and translational resolution of the sampling. The numbers $\upsilon_j$ determine the resolution of the shear sampling.

The subsystems associated to the $\varepsilon$-direction, $\varepsilon\in\{1,2\}$, are indexed by a set $\Lambda^s_\varepsilon$ of the form

$$\Lambda^s_\varepsilon := \Big\{(\varepsilon,j,\ell,k) \;:\; j\in\mathbb{N}_0,\, \ell\in\mathbb{L}^\varepsilon_j,\, k\in\mathbb{Z}^2\Big\}, \tag{3.30}$$

where $j\in\mathbb{N}_0$ corresponds to the scale, $\ell\in\mathbb{L}^\varepsilon_j$ to the orientation, and $k\in\mathbb{Z}^2$ to the spatial position of the elements. The possible shears at each scale $j$ are restricted by a nonempty set

$$\mathbb{L}^\varepsilon_j \subset \mathbb{Z} \quad \text{such that } \max\Big\{|\ell| \;:\; \ell\in\mathbb{L}^\varepsilon_j\Big\} \le L^\varepsilon_j,$$

where $(L^\varepsilon_j)_{j\in\mathbb{N}_0}$ is a sequence of nonnegative integers with $L^\varepsilon_j \lesssim \sigma^{j(1-\alpha)}$.

For the subsequent definitions we need the following matrices, where the different versions correspond to different regions of the frequency domain, i.e., either the cones $\mathcal{C}_\varepsilon$, $\varepsilon\in\{1,2\}$, or the low-frequency box $\mathcal{R}_0$. For the scaling we utilize the *$\alpha$-scaling matrices*

$$A^{(0)}_{\alpha,t} := A^{(1)}_{\alpha,t} := \begin{pmatrix} t & 0\\ 0 & t^\alpha\end{pmatrix} \quad\text{and}\quad A^{(2)}_{\alpha,t} = \begin{pmatrix} t^\alpha & 0\\ 0 & t\end{pmatrix}, \quad t>0,$$

for the shearing we utilize the *shear matrices*

$$S^{(0)}_\upsilon := S^{(1)}_\upsilon := \begin{pmatrix} 1 & \upsilon\\ 0 & 1\end{pmatrix} \quad\text{and}\quad S^{(2)}_\upsilon = \begin{pmatrix} 1 & 0\\ \upsilon & 1\end{pmatrix}, \quad \upsilon\in\mathbb{R}.$$

Further, the translational grid is obtained from $\mathbb{Z}^2$ using matrices of the form

$$M^{(0)}_c := M^{(1)}_c := \begin{pmatrix} c_1 & 0\\ 0 & c_2\end{pmatrix}, \quad M^{(2)}_c = \begin{pmatrix} c_2 & 0\\ 0 & c_1\end{pmatrix}, \quad c=(c_1,c_2)\in\mathbb{R}^2_+.$$

Now we are ready to introduce the subsystems $\mathfrak{M}^s_{\alpha,\varepsilon}$ corresponding to the frequency cones $\mathcal{C}_\varepsilon$ in $\varepsilon$-direction, $\varepsilon\in\{1,2\}$.

**Definition 3.3.1.** Let $\alpha\in[0,1]$, $\varepsilon\in\{1,2\}$, and $L,M,N_1,N_2\in\mathbb{N}_0\cup\{\infty\}$. Further, let the sampling parameters (3.29) be fixed, and let $\Lambda^s_\varepsilon$ be an index set of the form (3.30). We call a system

$$\mathfrak{M}^s_{\alpha,\varepsilon} := \Big\{m_\lambda\in L_2(\mathbb{R}^2) : \lambda\in\Lambda^s_\varepsilon\Big\}$$

a *system of $\alpha$-shearlet molecules of order $(L,M,N_1,N_2)$ associated with the $\varepsilon$-direction* if the functions $m_\lambda = m_{\varepsilon,j,\ell,k}$ can be represented in the form

$$m_{\varepsilon,j,\ell,k}(\cdot) := \sigma^{j(1+\alpha)/2}\gamma_{\varepsilon,j,\ell,k}\left(A^{(\varepsilon)}_{\alpha,\sigma^j}S^{(\varepsilon)}_{\ell\upsilon_j}\cdot - M^{(\varepsilon)}_c k\right)$$

with generators $\gamma_{\varepsilon,j,\ell,k}\in L_2(\mathbb{R}^2)$ which satisfy for every $\rho\in\mathbb{N}^2_0$ with $|\rho|\le L$,

$$|\partial^\rho\hat{\gamma}_{\varepsilon,j,\ell,k}(\xi_1,\xi_2)| \lesssim \min\Big\{1, \sigma^{-j}+|\xi_\varepsilon|+\sigma^{-j(1-\alpha)}|\xi_{3-\varepsilon}|\Big\}^M \cdot \langle|\xi|\rangle^{-N_1}\cdot\langle\xi_{3-\varepsilon}\rangle^{-N_2}. \tag{3.31}$$

Hereby the implicit constant shall be independent of the indices $(\varepsilon, j, \ell, k) \in \Lambda^s_\varepsilon$ and $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$. The value $\infty$ of a control parameter indicates that the respective quantity can be chosen arbitrarily large.

Next, we introduce the system $\mathfrak{M}^s_{\alpha,0}$ corresponding to the low-frequency box $\mathcal{R}_0$. As index set we use

$$\Lambda^s_0 := \Big\{(0,0,0,k) \ : \ k \in \mathbb{Z}^2\Big\}. \tag{3.32}$$

**Definition 3.3.2.** Let $\alpha \in [0,1]$ and $L, N_1, N_2 \in \mathbb{N}_0 \cup \{\infty\}$. The system $\mathfrak{M}^s_{\alpha,0}$ of *base-scale functions of order* $(L, N_1, N_2)$ is given by

$$\mathfrak{M}^s_{\alpha,0} := \Big\{m_\lambda : \lambda \in \Lambda^s_0\Big\}$$

with functions $m_\lambda \in L_2(\mathbb{R}^2)$ of the form

$$m_\lambda(\cdot) := \gamma_\lambda(\cdot - M_c^{(0)} k) \quad \text{for } \lambda = (0,0,0,k) \in \Lambda^s_0.$$

The generators $\gamma_\lambda \in L_2(\mathbb{R}^2)$ are assumed to satisfy

$$|\partial^\rho \hat{\gamma}_\lambda(\xi_1, \xi_2)| \lesssim \langle |\xi| \rangle^{-N_1} \cdot \langle \xi_2 \rangle^{-N_2}, \quad \xi = (\xi_1, \xi_2) \in \mathbb{R}^2,$$

for every $\rho \in \mathbb{N}_0^2$ with $|\rho| \le L$ and with an implicit constant independent of the index $\lambda \in \Lambda^s_0$ and $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$.

A full system of $\alpha$-shearlet molecules for $L_2(\mathbb{R}^2)$ is obtained by combining the systems $\mathfrak{M}^s_{\alpha,\varepsilon}$, $\varepsilon \in \{0,1,2\}$.

**Definition 3.3.3.** An *$\alpha$-shearlet index set* $\Lambda^s$ is defined as the union

$$\Lambda^s := \bigcup_{\varepsilon \in \{0,1,2\}} \Lambda^s_\varepsilon \tag{3.33}$$

of sets $\Lambda^s_\varepsilon$ of the form (3.30) and (3.32). A system $\mathfrak{M}^s_\alpha := \{m_\lambda \in L_2(\mathbb{R}^2) : \lambda \in \Lambda^s\}$, indexed by an $\alpha$-shearlet index set $\Lambda^s$, constitutes a *system of $\alpha$-shearlet molecules of order* $(L, M, N_1, N_2)$ if the subsystems $\mathfrak{M}^s_{\alpha,\varepsilon} := \{m_\lambda : \lambda \in \Lambda^s_\varepsilon\}$, $\varepsilon \in \{1,2\}$, are systems of $\alpha$-shearlet molecules of order $(L, M, N_1, N_2)$ in the sense of Definition 3.3.1 and if $\mathfrak{M}^s_{\alpha,0} := \{m_\lambda : \lambda \in \Lambda^s_0\}$ is a system of base-scale functions of order $(L, N_1, N_2)$ as in Definition 3.3.2. For convenience, we write this as

$$\mathfrak{M}^s_\alpha = \bigcup_{\varepsilon \in \{0,1,2\}} \mathfrak{M}^s_{\alpha,\varepsilon}.$$

The definition of $\alpha$-shearlet molecules is compatible with other notions of shearlet molecules, in particular those defined in [68, Def. 4.1].

**Proposition 3.3.4** (compare [62, Prop. 3.14])**.** *Shearlet molecules of regularity $R \in \mathbb{N}_0$, as defined in [68], are $\frac{1}{2}$-shearlet molecules of order* $(\infty, \infty, R/2, R/2)$.

*Proof.* An argument can be found right before [62, Prop. 3.14]. □

Next we will prove that, like $\alpha$-curvelet molecules, systems of $\alpha$-shearlet molecules constitute their own subclass of discrete $\alpha$-molecule systems.

### 3.3.1 The $\alpha$-Shearlet Parametrization

Since the construction of $\alpha$-shearlet molecules is based on shearings instead of rotations, the associated parametrizations need to adequately translate the shearing parameters to corresponding orientation angles. This complicates their definition compared to the definition of the $\alpha$-curvelet parametrizations in Definition 3.2.3.

**Definition 3.3.5.** With parameters given as in (3.29), an *$\alpha$-shearlet parametrization* $(\Lambda^s, \Phi^s)$ consists of an index set $\Lambda^s$ of the form (3.33) and a map $\Phi^s : \Lambda^s \to \mathbb{P}$ from $\Lambda^s$ into $\mathbb{P} = \mathbb{R}^2 \times \mathbb{T} \times \mathbb{R}_+$ defined by

$$\Phi^s \ : \ (\varepsilon, j, \ell, k) \mapsto \Big( S^{(\varepsilon)}_{-\ell \upsilon_j} A^{(\varepsilon)}_{\alpha,\sigma^{-j}} M^{(\varepsilon)}_c k, \big(\max\{0, \varepsilon - 1\}\pi/2 + \arctan(-\ell\upsilon_j)\big)_{2\pi}, \sigma^{-j} \Big).$$

Now we are ready to prove the essential result that $\alpha$-shearlet molecules are instances of $\alpha$-molecules. In fact, they can even be characterized as precisely those systems of $\alpha$-molecules which correspond to an $\alpha$-shearlet parametrization.

**Proposition 3.3.6** (compare [59, Prop. 3.9])**.** *Every system of $\alpha$-shearlet molecules of order $(L, M, N_1, N_2)$ constitutes a system of $\alpha$-molecules of the same order with respect to a corresponding $\alpha$-shearlet parametrization, and vice versa.*

*Proof.* The main ingredients of the proof can be found in [59, Subsec. 6.1.1], where [59, Prop. 3.9] is proved.

Let $\mathfrak{M}^s_\alpha := \{m_\lambda\}_{\lambda\in\Lambda^s}$ be a function system in $L_2(\mathbb{R}^2)$ indexed by a set $\Lambda^s = \Lambda^s_0 \cup \Lambda^s_1 \cup \Lambda^s_2$ of the form (3.33). Further, let $\Phi^s : \lambda \mapsto (x_\lambda, \eta_\lambda, t_\lambda)$ denote an $\alpha$-shearlet parametrization subject to parameters $\sigma$, $c = (c_1, c_2)$, $(\upsilon_j)_{j\in\mathbb{N}_0}$ as in (3.29).

Clearly, for each $\lambda \in \Lambda^s$ there exist unique functions $g_\lambda, \gamma_\lambda \in L_2(\mathbb{R}^2)$ such that

$$t_\lambda^{-(1+\alpha)/2} \gamma_\lambda\big(A^{(\varepsilon)}_{\alpha,t_\lambda^{-1}} S^{(\varepsilon)}_{\upsilon_\lambda} \cdot - M^{(\varepsilon)}_c k\big) = m_\lambda = t_\lambda^{-(1+\alpha)/2} g_\lambda\big(A^{-1}_{\alpha,t_\lambda} R_{\eta_\lambda}(\cdot - x_\lambda)\big),$$

where $\upsilon_\lambda := \ell\upsilon_j$. We need to show that the Fourier transform $\hat{g}_\lambda$ of $g_\lambda$ satisfies (2.5) if and only if $\hat{\gamma}_\lambda$ satisfies (3.31). For this investigation we decompose

$$\mathfrak{M}^s_\alpha = \bigcup_{\varepsilon\in\{0,1,2\}} \mathfrak{M}^s_{\alpha,\varepsilon}$$

and handle the subsystems $\mathfrak{M}^s_{\alpha,\varepsilon} := \{m_\lambda\}_{\lambda\in\Lambda^s_\varepsilon}$, $\varepsilon \in \{0, 1, 2\}$, separately.

Let us begin with the case $\varepsilon = 0$. Then $g_\lambda = \gamma_\lambda$ and the assertion is obvious.
Of the cases $\varepsilon \in \{1, 2\}$, we only handle $\varepsilon = 1$ since the arguments for the case $\varepsilon = 2$ are essentially the same. For $\mathfrak{M}^s_{\alpha,1}$ the relation between $g_\lambda$ and $\gamma_\lambda$ can be expressed by

$$\gamma_\lambda(\cdot) = g_\lambda(T_{j,\ell}\, \cdot\,) \quad \text{and} \quad g_\lambda(\cdot) = \gamma_\lambda(T^{-1}_{j,\ell}\, \cdot\,) \tag{3.34}$$

with a matrix

$$T_{j,\ell} := A^{-1}_{\alpha,t_\lambda} R_{\eta_\lambda} S^{-1}_{\ell\upsilon_j} A_{\alpha,t_\lambda}, \tag{3.35}$$

which describes the transfer from the rotation-based to the shear-based generators. Let us investigate the properties of this ‘transfer matrix’ $T_{j,\ell}$.

For this purpose, it is useful to first examine the matrix $\widetilde{T}_{j,\ell} := R_{\eta_\lambda} S^{-1}_{\ell\upsilon_j}$. Since $\eta_\lambda = (\arctan(-\ell\upsilon_j))_{2\pi}$, we have

$$S_{\ell\upsilon_j} = \begin{pmatrix} 1 & \ell\upsilon_j \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -\tan(\eta_\lambda) \\ 0 & 1 \end{pmatrix}.$$

Using

$$0 = \tan(\eta_\lambda)\cos(\eta_\lambda) - \sin(\eta_\lambda) \quad \text{and} \quad \cos(\eta_\lambda)^{-1} = \tan(\eta_\lambda)\sin(\eta_\lambda) + \cos(\eta_\lambda),$$

we calculate

$$\widetilde{T}_{j,\ell} = \begin{pmatrix} \cos(\eta_\lambda) & 0 \\ \sin(\eta_\lambda) & \cos(\eta_\lambda)^{-1} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{1+(\ell\upsilon_j)^2}} & 0 \\ \frac{-\ell\upsilon_j}{\sqrt{1+(\ell\upsilon_j)^2}} & \sqrt{1+(\ell\upsilon_j)^2} \end{pmatrix} =: \begin{pmatrix} a(j,\ell) & 0 \\ b(j,\ell) & c(j,\ell) \end{pmatrix}.$$

Taking into account $a(j,\ell) = c(j,\ell)^{-1}$, we obtain for the inverse

$$\widetilde{T}^{-1}_{j,\ell} = \begin{pmatrix} c(j,\ell) & 0 \\ -b(j,\ell) & a(j,\ell) \end{pmatrix} = \begin{pmatrix} \sqrt{1+(\ell\upsilon_j)^2} & 0 \\ \frac{\ell\upsilon_j}{\sqrt{1+(\ell\upsilon_j)^2}} & \frac{1}{\sqrt{1+(\ell\upsilon_j)^2}} \end{pmatrix}.$$

By assumption $|\ell| \lesssim \sigma^{j(1-\alpha)}$ and $\upsilon_j \asymp \sigma^{-j(1-\alpha)}$, which implies the existence of a bound $B > 0$ such that

$$|\ell\upsilon_j| \le B \quad \text{for all } j \in \mathbb{N}_0,\ \ell \in \mathbb{L}^1_j.$$

Hence, uniformly for all $j \in \mathbb{N}_0$ and $\ell \in \mathbb{L}^1_j$

$$1/\sqrt{1+B^2} \le a(j,\ell) \le 1 \le c(j,\ell) \le \sqrt{1+B^2}\,, \quad |b(j,\ell)| \le B/\sqrt{1+B^2}. \tag{3.36}$$

Turning to the matrix $T_{j,\ell}$, we calculate

$$T_{j,\ell} = \begin{pmatrix} a(j,\ell) & 0 \\ \sigma^{-j(1-\alpha)} b(j,\ell) & c(j,\ell) \end{pmatrix} \quad \text{and} \quad T^{-1}_{j,\ell} = \begin{pmatrix} c(j,\ell) & 0 \\ -\sigma^{-j(1-\alpha)} b(j,\ell) & a(j,\ell) \end{pmatrix}.$$

Since $|\sigma^{-j(1-\alpha)}| \le 1$ for every $j \in \mathbb{N}_0$, we obtain for the Frobenius norm

$$\|T^T_{j,\ell}\|_F = \|T_{j,\ell}\|_F \le \sqrt{2+B^2} \quad \text{and} \quad \|T^{-T}_{j,\ell}\|_F = \|T^{-1}_{j,\ell}\|_F \le \sqrt{2+B^2}.$$

As a consequence, we have uniformly in $j \in \mathbb{N}_0$, $\ell \in \mathbb{L}^1_j$,

$$\frac{1}{\sqrt{2+B^2}}|\xi| \le |T^T_{j,\ell}\xi| \le \sqrt{2+B^2}|\xi| \quad \text{for all } \xi \in \mathbb{R}^2\,, \tag{3.37}$$

and the respective relation for $|T^{-T}_{j,\ell}\xi|$. Since $\det T_{j,\ell} = \det T^{-1}_{j,\ell} = 1$, we obtain from (3.34)

$$\hat{\gamma}_\lambda(\cdot) = \hat{g}_\lambda(T^{-T}_{j,\ell}\,\cdot\,) \quad \text{and} \quad \hat{g}_\lambda(\cdot) = \hat{\gamma}_\lambda(T^T_{j,\ell}\,\cdot\,).$$

Now we are ready to show that the assumption (3.31) on $\hat{\gamma}_\lambda$ implies property (2.5) for $\hat{g}_\lambda$, and vice versa. Using $|\sigma^{-j(1-\alpha)}| \le 1$, the uniform boundedness of $|a(j,\ell)|, |b(j,\ell)|, |c(j,\ell)|$, and the chain rule, we can estimate for any $\rho \in \mathbb{N}_0^2$ with $|\rho| \le L$,

$$|\partial^\rho \hat{g}_\lambda(\xi)| \lesssim \sup_{|\nu|\le L} \left| \partial^\nu \hat{\gamma}_\lambda \left( \begin{pmatrix} a(j,\ell) & \sigma^{-j(1-\alpha)} b(j,\ell) \\ 0 & c(j,\ell) \end{pmatrix} \xi \right) \right|. \tag{3.38}$$

Utilizing the moment estimate in (3.31) for $\hat{\gamma}_\lambda$ gives the moment property required in (2.5), namely

$$\begin{aligned} |\partial^\rho \hat{g}_\lambda(\xi)| &\lesssim \Big( \sigma^{-j} + |a(j,\ell)\xi_1 + \sigma^{-j(1-\alpha)} b(j,\ell)\xi_2| + \sigma^{-j(1-\alpha)} |c(j,\ell)\xi_2| \Big)^M \\ &\lesssim \Big( t_\lambda + |\xi_1| + t_\lambda^{1-\alpha} |\xi_2| \Big)^M . \end{aligned}$$

It remains to show the decay of $\partial^\rho \hat{g}_\lambda$ for large frequencies $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$. We obtain from (3.38) and the decay estimate in (3.31),

$$|\partial^\rho \hat{g}_\lambda(\xi)| \lesssim \left\langle \left| \begin{pmatrix} a(j,\ell) & \sigma^{-j(1-\alpha)} b(j,\ell) \\ 0 & c(j,\ell) \end{pmatrix} \xi \right| \right\rangle^{-N_1} \langle c(j,\ell)\xi_2 \rangle^{-N_2} \lesssim \langle |\xi| \rangle^{-N_1} \langle \xi_2 \rangle^{-N_2},$$

where the last estimate is a consequence of $|T_{j,\ell}^T \xi| \asymp |\xi|$ due to (3.37) and $c(j,\ell)\xi_2 \asymp \xi_2$.

We finish the proof by noting that an analogous argumentation, with the matrix $T_{j,\ell}^{-T}$ taking the role of $T_{j,\ell}^T$, yields (3.31) for $\hat{\gamma}_\lambda$ under the assumption (2.5) on $\hat{g}_\lambda$. □

### 3.3.2 A Sufficient Frame Condition for $\alpha$-Shearlet Molecules

We will now use Theorem 2.4.1 to derive a sufficient frame criterion for $\alpha$-shearlet molecules similar to Theorem 3.2.5. First, recall that an $\alpha$-shearlet parametrization $(\Lambda^s, \Phi^s)$ is determined by an $\alpha$-shearlet index set, which can be decomposed in the form

$$\Lambda^s = \Delta^s \times \mathbb{Z}^2 \quad \text{with } \Delta^s := \Big\{ (0,0,0) \Big\} \cup \Big\{ (\varepsilon, j, \ell) \ : \ \varepsilon \in \{1,2\}, j \in \mathbb{N}_0, \ell \in \mathbb{L}_j^\varepsilon \Big\},$$

and a corresponding parametrization map

$$\Phi^s : \Lambda^s \to \mathbb{P} \ : \ (\varepsilon, j, \ell, k) \mapsto \Big( S^{(\varepsilon)}_{-\ell \upsilon_j} A^{(\varepsilon)}_{\alpha, \sigma^{-j}} M_c^{(\varepsilon)} k, \big( \max\{0, \varepsilon - 1\} \frac{\pi}{2} + \arctan(-\ell \upsilon_j) \big)_{2\pi}, \sigma^{-j} \Big).$$

Hence, $\alpha$-shearlet parametrizations have the structure (2.34) if for each $\mu = (\varepsilon, j, \ell) \in \Delta^s$ we choose $T_\mu := T_{j,\ell}^{(\varepsilon)} M_c^{(\varepsilon)} M_c^{-1}$ with the matrix

$$T_{j,\ell}^{(\varepsilon)} := A_{\alpha, t_\mu}^{-1} R_{\eta_\mu} S^{(\varepsilon)}_{-\ell \upsilon_j} A^{(\varepsilon)}_{\alpha, t_\mu}. \tag{3.39}$$

Note that indeed $|\det(T_\mu)| = 1$, as required. Note further that the matrices $T_{j,\ell}^{(\varepsilon)}$ are transfer matrices of the type (3.35) which were analyzed in the proof of Proposition 3.3.6.

Now let $\{m_\lambda\}_{\lambda \in \Lambda^s}$ be a system of $\alpha$-shearlet molecules with associated shear-based generators $\{\gamma_\lambda\}_{\lambda \in \Lambda^s}$ and assume that the functions $\gamma_\lambda = \gamma_{\mu,k}$ do not depend on $k \in \mathbb{Z}^2$. In this situation the frame criterion for $\alpha$-molecules, Theorem 2.4.1, can be applied to positively decide whether this system forms a frame for $L_2(\mathbb{R}^2)$.

In the case of $\alpha$-shearlet molecules, it is more convenient to formulate the correlation functions in terms of the shear-based generators $\gamma_\lambda = \gamma_\mu$, $\lambda = (\mu, k) \in \Lambda^s$. Hence, we introduce the functions $\Phi^s : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ and $\Gamma^s : \mathbb{R}^2 \to \mathbb{R}$ as

$$\begin{aligned}\Phi^s(\xi,\omega) &:= \sum_{\mu\in\Delta^s} |\hat{\gamma}_\mu(A^{(\varepsilon)}_{\alpha,t_\mu}(S^{(\varepsilon)}_{\ell\upsilon_j})^{-T}\xi)||\hat{\gamma}_\mu(A^{(\varepsilon)}_{\alpha,t_\mu}(S^{(\varepsilon)}_{\ell\upsilon_j})^{-T}\xi + (M^{(\varepsilon)}_c)^{-1}M_c\omega)|,\\ \Gamma^s(\omega) &:= \operatorname*{ess\,sup}_{\xi\in\mathbb{R}^2} \Phi^s(\xi,\omega).\end{aligned} \tag{3.40}$$

For $c = (c_1, c_2) \in \mathbb{R}^2_+$, we further define

$$R^s(c) := \sum_{m\in\mathbb{Z}^2\setminus\{0\}} \left[\Gamma^s(M_c^{-1}m)\Gamma^s(-M_c^{-1}m)\right]^{1/2}.$$

Then we have the following result.

**Theorem 3.3.7.** *Let $\{m_\lambda\}_{\lambda\in\Lambda^s}$ be a system of $\alpha$-shearlet molecules with a corresponding family of generators $\{\gamma_\lambda\}_{\lambda\in\Lambda^s}$, and assume that the generators $\gamma_\lambda = \gamma_{\mu,k}$ do not vary with $k \in \mathbb{Z}^2$. Further, let $c = (c_1, c_2) \in \mathbb{R}^2_+$ be the parameter in* (3.29) *associated with the density of the translation grid. Then the condition*

$$R^s(c) < L^s_{\inf} \le L^s_{\sup} < \infty$$

*for the quantities*

$$L^s_{\inf} := \operatorname*{ess\,inf}_{\xi\in\mathbb{R}^2} \Phi^s(\xi,0) \quad \text{and} \quad L^s_{\sup} := \Gamma^s(0) = \operatorname*{ess\,sup}_{\xi\in\mathbb{R}^2} \Phi^s(\xi,0)$$

*ensures that $\{m_\lambda\}_{\lambda\in\Lambda^s}$ constitutes a frame for $L_2(\mathbb{R}^2)$ with frame bounds $A, B > 0$ satisfying*

$$\frac{L^s_{\inf} - R^s(c)}{|\det M_c|} \le A \le B \le \frac{L^s_{\sup} + R^s(c)}{|\det M_c|}.$$

*Proof.* The system $\{m_\lambda\}_{\lambda\in\Lambda^s}$ is a system of $\alpha$-molecules with respect to an $\alpha$-shearlet parametrization. The associated rotation-based generators are given by $g_\mu = \gamma_\mu((T^{(\varepsilon)}_{j,\ell})^{-1}\cdot)$, where $T^{(\varepsilon)}_{j,\ell}$ is the transfer matrix (3.39). It follows

$$\hat{\gamma}_\mu = \hat{g}_\mu((T^{(\varepsilon)}_{j,\ell})^{-T}\cdot) = \hat{g}_\mu(A_{\alpha,t_\mu}R_{\eta_\mu}(S^{(\varepsilon)}_{\ell\upsilon_j})^T(A^{(\varepsilon)}_{\alpha,t_\mu})^{-1}\cdot).$$

Plugging this into (3.40) yields for the correlation function

$$\Phi^s(\xi,\omega) = \sum_{\mu\in\Delta^s} |\hat{g}_\mu(A_{\alpha,t_\mu}R_{\eta_\mu}\xi)||\hat{g}_\mu(A_{\alpha,t_\mu}R_{\eta_\mu}\xi + T_\mu^{-T}\omega)|,$$

where $T_\mu^{-T} = (T^{(\varepsilon)}_{j,\ell})^{-T}(M^{(\varepsilon)}_c)^{-1}M_c$. Hence we can apply Theorem 2.4.1 and the assertion follows. □

Finally note that $S_\ell A_{\alpha,\sigma^j} = A_{\alpha,\sigma^j}S_{\ell\upsilon_j}$ if $\upsilon_j = \sigma^{-j(1-\alpha)}$. Hence for a strict $\alpha$-shearlet system as in Definition 3.3.8 this is precisely the criterion proved in [76, Thm. 3.4].

### 3.3.3 Discrete $\alpha$-Shearlet Systems

The concept of $\alpha$-shearlet molecules comprises many common cone-adapted shearlet constructions. These include band-limited as well as compactly supported systems, as we will see below.

The general structure of a regular cone-adapted discrete $\alpha$-shearlet system is recalled in Definition 3.3.8. It is a generalization of [80, Def. 11] to $\alpha$-scaling. Note, that sometimes the parameter $\beta = \alpha^{-1}$ is used in the definition instead, as for example in [59, Def. 3.10].

**Definition 3.3.8** (compare [59, Def. 3.10])**.** For $c = (c_1, c_2) \in \mathbb{R}^2_+$ and $\alpha \in (0,1)$, the *cone-adapted $\alpha$-shearlet system* $SH(\phi, \psi, \tilde{\psi}; c, \alpha)$ generated by $\phi, \psi, \tilde{\psi} \in L_2(\mathbb{R}^2)$ is defined by

$$SH(\phi, \psi, \tilde{\psi}; c, \alpha) := \Phi(\phi; c, \alpha) \cup \Psi(\psi; c, \alpha) \cup \tilde{\Psi}(\tilde{\psi}; c, \alpha),$$

where, with $\beta = \alpha^{-1}$,

$$\Phi(\phi; c, \alpha) := \{\phi_k = \phi(\cdot - M_c^{(0)} k) : k \in \mathbb{Z}^2\},$$
$$\Psi(\psi; c, \alpha) := \{\psi_{j,\ell,k} = 2^{j(\beta+1)/4} \psi(S_\ell^{(1)} A^{(1)}_{\alpha, 2^{j\beta/2}} \cdot - M_c^{(1)} k) : j \in \mathbb{N}_0, |\ell| \le \lceil 2^{j(\beta-1)/2} \rceil, k \in \mathbb{Z}^2\},$$
$$\tilde{\Psi}(\tilde{\psi}; c, \alpha) := \{\tilde{\psi}_{j,\ell,k} = 2^{j(\beta+1)/4} \tilde{\psi}(S_\ell^{(2)} A^{(2)}_{\alpha, 2^{j\beta/2}} \cdot - M_c^{(2)} k) : j \in \mathbb{N}_0, |\ell| \le \lceil 2^{j(\beta-1)/2} \rceil, k \in \mathbb{Z}^2\}.$$

In the following, we present some examples of cone-adapted $\alpha$-shearlet systems as in Definition 3.3.8. Thereby the generators $\phi, \psi, \tilde{\psi} \in L_2(\mathbb{R}^2)$ are assumed to be either band-limited or compactly supported.

In the band-limited case, we require $\hat{\phi}, \hat{\psi}, \hat{\tilde{\psi}} \in C_0^\infty(\mathbb{R}^2)$ and a frequency support of the form

$$\operatorname{supp} \hat{\phi} \subseteq Q, \qquad \operatorname{supp} \hat{\psi} \subseteq W, \qquad \operatorname{supp} \hat{\tilde{\psi}} \subseteq \tilde{W},$$

with $Q \subset \mathbb{R}^2$ being a cube centered at the origin and $W, \tilde{W} \subset \mathbb{R}^2$ being sets of the form

$$W = [-a, a] \times ([-c, -b] \cup [b, c]), \qquad \tilde{W} = ([-c, -b] \cup [b, c]) \times [-a, a]$$

with $0 < b < c$ and $0 < a$.

In the compact case, the coarse-scale generator $\phi$ shall satisfy

$$\phi \in C_0^{N_1 + N_2}(\mathbb{R}^2).$$

Furthermore, we assume the separability of $\psi \in L_2(\mathbb{R}^2)$, i.e. $\psi(x_1, x_2) = \psi_1(x_1)\psi_2(x_2)$, and let $\tilde{\psi}$ be its rotation by $\pi/2$. Finally, the functions $\psi_1$, $\psi_2$ shall satisfy

$$\psi_1 \in C_0^{N_1}(\mathbb{R}) \quad \text{and} \quad \psi_2 \in C_0^{N_1+N_2}(\mathbb{R}),$$

and for $\psi_1$ we assume $M \in \mathbb{N}_0$ vanishing moments.

The following proposition shows that under these assumptions the systems $SH(\phi, \psi, \tilde{\psi}; c, \alpha)$ are instances of $\alpha$-shearlet molecules.

**Proposition 3.3.9** ([59, Prop. 3.11])**.** *Let $\alpha \in (0,1)$, $\beta = \alpha^{-1}$, and $c \in \mathbb{R}^2_+$ be fixed parameters. The following statements hold.*

*(i) For band-limited generators $\phi, \psi$, and $\tilde{\psi}$ subject to the conditions above, the cone-adapted $\alpha$-shearlet system $SH(\phi, \psi, \tilde{\psi}; c, \alpha)$ is a system of $\alpha$-shearlet molecules of order $(\infty, \infty, \infty, \infty)$ with $\sigma = 2^{\beta/2}$, $\upsilon_j = \sigma^{-j(1-\alpha)}$, and $L_j = \lceil \sigma^{j(1-\alpha)} \rceil$.*

(ii) *For compactly supported generators $\phi, \psi$, and $\tilde{\psi}$ subject to the conditions above, the cone-adapted $\alpha$-shearlet system $SH(\phi, \psi, \tilde{\psi}; c, \alpha)$ is a system of $\alpha$-shearlet molecules of order $(L, M - L, N_1, N_2)$, where $L \in \{0, \ldots, M\}$ is arbitrary, with $\sigma = 2^{\beta/2}$, $\upsilon_j = \sigma^{-j(1-\alpha)}$, and $L_j = \lceil \sigma^{j(1-\alpha)} \rceil$.*

*Proof.* For the proof we also refer to [59, Subsec. 6.1.2].

Let us first rename the functions of the system $SH(\phi, \psi, \tilde{\psi}; c, \alpha)$. For $j \in \mathbb{N}_0$, $\ell \in \mathbb{Z}$ with $|\ell| \le \lceil 2^{j(\beta-1)/2} \rceil$, and $k \in \mathbb{Z}^2$ we put

$$\begin{aligned}
\psi_{0,0,0,k} &:= \phi_k = \phi(\cdot - M_c^{(0)} k),\\
\psi_{1,j,\ell,k} &:= \psi_{j,\ell,k} = 2^{j(\beta+1)/4} \psi(S_\ell^{(1)} A_{\alpha, 2^{j\beta/2}}^{(1)} \cdot - M_c^{(1)} k),\\
\psi_{2,j,\ell,k} &:= \tilde{\psi}_{j,\ell,k} = 2^{j(\beta+1)/4} \tilde{\psi}(S_\ell^{(2)} A_{\alpha, 2^{j\beta/2}}^{(2)} \cdot - M_c^{(2)} k).
\end{aligned}$$

With $\sigma = 2^{\beta/2}$ we can rewrite $A_{\alpha,2^{j\beta/2}}^{(1)} = A_{\alpha,\sigma^j}^{(1)}$ and $A_{\alpha,2^{j\beta/2}}^{(2)} = A_{\alpha,\sigma^j}^{(2)}$. Further, using $\upsilon_j = \sigma^{-j(1-\alpha)}$, we obtain

$$S_\ell^{(1)} A_{\alpha,\sigma^j}^{(1)} = A_{\alpha,\sigma^j}^{(1)} S_{\ell\sigma^{-j(1-\alpha)}}^{(1)} = A_{\alpha,\sigma^j}^{(1)} S_{\ell\upsilon_j}^{(1)}$$

and

$$S_\ell^{(2)} A_{\alpha,\sigma^j}^{(2)} = A_{\alpha,\sigma^j}^{(2)} S_{\ell\sigma^{-j(1-\alpha)}}^{(2)} = A_{\alpha,\sigma^j}^{(2)} S_{\ell\upsilon_j}^{(2)}.$$

Finally, due to $2^{j(\beta+1)/4} = \sigma^{j(1+\alpha)/2}$, we arrive at the representation

$$\begin{aligned}
\psi_{0,0,0,k} &= \phi(\cdot - M_c^{(0)} k),\\
\psi_{1,j,\ell,k} &= \sigma^{j(1+\alpha)/2} \psi(A_{\alpha,\sigma^j}^{(1)} S_{\ell\upsilon_j}^{(1)} \cdot - M_c^{(1)} k),\\
\psi_{2,j,\ell,k} &= \sigma^{j(1+\alpha)/2} \tilde{\psi}(A_{\alpha,\sigma^j}^{(2)} S_{\ell\upsilon_j}^{(2)} \cdot - M_c^{(2)} k).
\end{aligned}$$

Hence, the system $SH(\phi, \psi, \tilde{\psi}; c, \alpha) = \{\psi_\lambda\}_{\lambda \in \Lambda^s}$ has the structure of a system of $\alpha$-shearlet molecules generated by $\gamma_{1,j,\ell,k} := \psi$, $\gamma_{2,j,\ell,k} := \tilde{\psi}$, and $\gamma_{0,0,0,k} := \phi$.

It remains to prove that these generators satisfy (3.31). In case of band-limited functions the proof is analogous to the proof of Proposition 3.2.8(ii). The only interesting part is thus the case of generators with compact support. Here we restrict our considerations to the functions $\gamma_{1,j,\ell,k} = \psi$.

The inverse Fourier transform of $\partial^\rho \hat{\psi}$, where $\rho \in \mathbb{N}_0^2$, is up to a constant given by $x \mapsto x^\rho \psi(x)$. By smoothness and compact support of $\psi_1, \psi_2$, we find that for any $|\rho| \le L$ the functions

$$x \mapsto \partial^{(N_1, N_1+N_2)}(x^\rho \psi(x)) \quad \text{and} \quad x \mapsto x^\rho \psi(x)$$

belong to $L_1(\mathbb{R}^2)$. Hence, on the Fourier side

$$\xi \mapsto \xi_1^{N_1} \xi_2^{N_1+N_2} \partial^\rho \hat{\psi}(\xi) \quad \text{and} \quad \xi \mapsto \partial^\rho \hat{\psi}(\xi)$$

are continuous and contained in $L_\infty(\mathbb{R}^2)$. It follows that

$$\langle \xi_1 \rangle^{N_1} \langle \xi_2 \rangle^{N_1+N_2} \partial^\rho \hat{\psi}(\xi)$$

is bounded in modulus. Using $\langle x\rangle\langle y\rangle \geq \langle\sqrt{x^2+y^2}\rangle$ we get the decay estimate for large frequencies

$$|\partial^\rho \hat{\psi}(\xi)| \lesssim \langle|\xi|\rangle^{-N_1}\langle\xi_2\rangle^{-N_2}.$$

Let us turn to the moment conditions. Let $\rho = (\rho_1, \rho_2) \in \mathbb{N}_0^2$ with $|\rho_1| \leq L$ for some $L = 0, \ldots, M$. Then

$$x^\rho \psi(x) = x_1^{\rho_1}\psi_1(x_1)x_2^{\rho_2}\psi_2(x_2)$$

restricted to the variable $x_1$ possesses at least $M-L$ vanishing moments, since $\psi_1$ is assumed to possess $M$ vanishing moments. This yields a decay of order $\min\{1, |\xi_1|^{M-L}\}$ for the derivatives up to order $L$ of $\hat{\psi}$ by the following lemma, whose proof can be found, e.g., in [62].

**Lemma 3.3.10** ([62])**.** *Suppose that $g : \mathbb{R} \to \mathbb{C}$ is continuous, compactly supported and possesses $M$ vanishing moments. Then*

$$|\hat{g}(\xi)| \lesssim \min\{1, |\xi|\}^M.$$

The proof is finished. □

Proposition 3.3.9 shows that various versions of cone-adapted shearlet systems are united under the roof of $\alpha$-shearlet molecules. Furthermore, also non-affine constructions such as the smooth Parseval frame of shearlets by Guo and Labate [70, 67] fall into this general framework.

Since, by Proposition 3.3.6, $\alpha$-shearlet molecules are particular instances of $\alpha$-molecules, these examples further show that the concept of $\alpha$-molecules is general enough to include both shear-based and rotation-based constructions.

## 3.4 Consistency of $\alpha$-Curvelet and $\alpha$-Shearlet Parametrizations

Despite their different constructions, shearlet and curvelet systems are closely related and in many respects exhibit a similar behavior. For example, the same approximation rates with respect to cartoon-like data have been observed for various systems [15, 60, 67, 78, 73]. An explanation for this similar behavior can be given by the framework of $\alpha$-molecules.

Both, $\alpha$-curvelets and $\alpha$-shearlets are instances of discrete $\alpha$-molecule systems and, as we will prove in this section, the corresponding parametrizations are consistent in the sense of Definition 2.3.5. Hence, the same approximation rates are a direct consequence of the transfer principle for $\alpha$-molecules, Theorem 2.3.6.

In the main result of this section, Theorem 3.4.3, we first compare the $\alpha$-curvelet and $\alpha$-shearlet parametrizations with the canonical parametrization from Definition 3.1.2. As a corollary, we can then easily deduce the consistency of the different $\alpha$-curvelet and $\alpha$-shearlet parametrizations among themselves, stated in Corollary 3.4.4.

For the proof of Theorem 3.4.3, we need two auxiliary lemmas. The first one is given below.

**Lemma 3.4.1.** *Let $0 \leq \delta < \frac{\pi}{2}$ be fixed. Then uniformly for $|\eta| \leq \delta$ and $|\theta| \leq \frac{\pi}{2}$*

$$|\{\eta - \theta\}| \asymp |\eta - \theta|.$$

*Proof.* Since $|\eta - \theta| \leq \frac{\pi}{2} + \delta < \pi$, there exists $\iota \in \{-1, 0, 1\}$ with

$$|\{\eta - \theta\}| = |\eta - \theta - \iota\pi|.$$

In case $\iota = 0$ it even holds $|\{\eta - \theta\}| = |\eta - \theta|$. In case $|\iota| = 1$ we estimate

$$|\{\eta - \theta\}| = |\eta - \theta - \iota\pi| \geq |\pi - |\eta - \theta|| \geq \frac{\pi}{2} - \delta.$$

Further, $|\eta - \theta| \leq \frac{\pi}{2} + \delta$ and hence $|\{\eta - \theta\}| \geq \frac{\pi - 2\delta}{\pi + 2\delta} |\eta - \theta|$. The other direction, i.e., the estimate $|\{\eta - \theta\}| \leq |\eta - \theta|$, is always true. □

The second auxiliary lemma is as follows.

**Lemma 3.4.2** ([59, Lem. 6.8])**.** *For all $x$, $y \in \mathbb{R}$, absolutely bounded by some fixed bound $B \geq 0$, i.e., $|x|, |y| \leq B$, we have*

$$|\arctan x - \arctan y| \asymp |x - y|.$$

*Proof.* For $x \neq y$ we have for some $\xi$ between $x$ and $y$ by the mean value theorem

$$\frac{|\arctan x - \arctan y|}{|x - y|} = \arctan'(\xi) = \frac{1}{1 + \xi^2}.$$

This yields

$$\frac{1}{1 + B^2} |x - y| \leq |\arctan x - \arctan y| \leq |x - y|.$$

The case $x = y$ is trivial. □

Note, that, as a consequence of this lemma, if $\theta, \theta' \in c[-\pi/2, \pi/2]$ with $0 < c < 1$ then

$$|\theta - \theta'| \asymp |\tan(\theta) - \tan(\theta')|.$$

Now we are ready to prove Theorem 3.4.3. The proof is analogous to the proof of [59, Lem. 5.8] in [59, Subsec. 6.3.2].

**Theorem 3.4.3** (compare [59, Lem. 5.8])**.** *Let $\alpha \in [0, 1]$, and let $(\Lambda, \Phi_\Lambda)$ be either an $\alpha$-curvelet or an $\alpha$-shearlet parametrization. Then we have for all $N > 2$*

$$\sup_{\mathbf{y} \in \mathbb{P}} \sum_{\lambda \in \Lambda} \omega_\alpha(\Phi_\Lambda(\lambda), \mathbf{y})^{-N} < \infty.$$

*Proof.* Let us write $\mathbf{x}_\lambda = (x_\lambda, \eta_\lambda, t_\lambda) := \Phi_\Lambda(\lambda)$ for $\lambda \in \Lambda$. By the definition of $\omega_\alpha$, we need to consider

$$S_{\mathbf{y}} := \sum_{j \in \mathbb{N}_0} \sum_{\substack{\lambda \in \Lambda \\ t_\lambda = \sigma^{-j}}} \max\left\{\frac{t_\lambda}{u}, \frac{u}{t_\lambda}\right\}^{-N} \left(1 + d_\alpha(\mathbf{x}_\lambda, \mathbf{y})\right)^{-N}$$

for every $\mathbf{y} = (y, \theta, u) \in \mathbb{P}$. We will prove below that

$$S_{j,\mathbf{y}} := \sum_{\substack{\lambda \in \Lambda \\ t_\lambda = \sigma^{-j}}} \left(1 + d_\alpha(\mathbf{x}_\lambda, \mathbf{y})\right)^{-N} \lesssim \max\left\{\frac{u}{t_\lambda}, 1\right\}^2, \tag{3.41}$$

where the implicit constant is independent of $\mathbf{y} = (y, \theta, u) \in \mathbb{P}$ and $j \in \mathbb{N}_0$. Now let $j' \in \mathbb{R}$ denote the unique real number with $u = \sigma^{-j'}$. Then we can deduce

$$S_{\mathbf{y}} \lesssim \sum_{j \in \mathbb{N}_0} \max\left\{\frac{\sigma^{-j}}{u}, \frac{u}{\sigma^{-j}}\right\}^{2-N} = \sum_{j \in \mathbb{N}_0} \sigma^{|j-j'|(2-N)} \le 2 \sum_{j \in \mathbb{N}_0} \sigma^{j(2-N)} =: C < \infty,$$

with a positive constant $C$ independent of $\mathbf{y} \in \mathbb{P}$. The assertion of the theorem follows.

It remains to establish (3.41) for the $\alpha$-curvelet as well as the $\alpha$-shearlet parametrization. In view of Lemma 2.2.4, it suffices to estimate the sum

$$\widetilde{S}_{j,\mathbf{y}} := \sum_{\substack{\lambda \in \Lambda \\ t_\lambda = \sigma^{-j}}} \left(1 + \widetilde{d}_\alpha(\Phi_\Lambda(\lambda), \mathbf{y})\right)^{-N}, \tag{3.42}$$

where $\widetilde{d}_\alpha := d_\alpha^{\mathrm{sim}}$ is the simplified version of $d_\alpha$ from Definition 2.2.3 given by

$$\widetilde{d}_\alpha(\Phi_\Lambda(\lambda), \mathbf{y}) := t_0^{-2(1-\alpha)} |\{\eta_\lambda - \theta\}|^2 + t_0^{-2\alpha} |x_\lambda. - y|^2 + t_0^{-1} |\langle e_\lambda, x_\lambda - y\rangle|.$$

Here $e_\lambda := e_{\eta_\lambda}$ and $t_0 := \max\{t_\lambda, u\}$ for $\mathbf{y} = (y, \theta, u) \in \mathbb{P}$.

We subsequently handle the cases of $(\Lambda, \Phi_\Lambda)$ being either $\alpha$-curvelet or $\alpha$-shearlet parametrization separately.

Part (i): Let us first assume that $(\Lambda, \Phi_\Lambda) = (\Lambda^c, \Phi^c)$ is an $\alpha$-curvelet parametrization with parameters $\sigma > 1$, $c = (c_1, c_2) \in \mathbb{R}^2_+$, $(\omega_j)_j$, and $(L_j)_j$ as in (3.11).

For the subsequent arguments, let us denote the first component of a vector $z \in \mathbb{R}^2$ by $[z]_1$ and the second by $[z]_2$. Further, let $e_1$ denote the first unit vector of $\mathbb{R}^2$.

We obtain the estimates

$$|\langle e_\lambda, x_\lambda - y\rangle| = |\langle R_{\eta_\lambda}^{-1} e_1, R_{\eta_\lambda}^{-1} A_{\alpha, t_\lambda} M_c k - y\rangle| = |t_\lambda c_1 k_1 - [R_{\eta_\lambda} y]_1|, \tag{3.43}$$

and

$$|x_\lambda - y| = |R_{\eta_\lambda}^{-1} A_{\alpha, t_\lambda} M_c k - y| \ge |t_\lambda^\alpha c_2 k_2 - [R_{\eta_\lambda} y]_2|. \tag{3.44}$$

To deal with the term $|\{\eta_\lambda - \theta\}|$, note that $\{\eta_\lambda - \theta\} = \{\ell\omega_j - \theta\}$ and define

$$\mathbb{Z}_j^{(m)} := \left\{\ell \in \mathbb{Z} : \ell\omega_j - \theta - m\pi \in [-\pi/2, \pi/2)\right\}, \qquad m \in \mathbb{Z}.$$

Since $L_j \lesssim \sigma^{j(1-\alpha)}$ and $\omega_j \asymp \sigma^{-j(1-\alpha)}$ we have $|\ell\omega_j| \lesssim 1$ for $\ell \in \mathbb{L}_j$. Hence, there is a bound $B > 0$ such that

$$|\ell\omega_j| \le B \quad \text{for all } j \in \mathbb{N}_0,\ \ell \in \mathbb{L}_j. \tag{3.45}$$

Hence, there exists $M \in \mathbb{N}_0$, independent of $j$, such that $\mathbb{L}_j^{(m)} := \mathbb{L}_j \cap \mathbb{Z}_j^{(m)} = \emptyset$ for all $j$ and $|m| > M$. We can thus decompose

$$\widetilde{S}_{j,\mathbf{y}} = \sum_{m=-M}^{M} \widetilde{S}_{j,\mathbf{y}}^{(m)}$$

into a sum consisting of the terms

$$\widetilde{S}_{j,\mathbf{y}}^{(m)} := \sum_{\ell \in \mathbb{L}_j^{(m)}} \sum_{k \in \mathbb{Z}^2} \Big( 1 + \frac{|\ell \omega_j - \theta - m\pi|^2}{t_0^{2(1-\alpha)}} + \frac{|x_\lambda - y|^2}{t_0^{2\alpha}} + \frac{|\langle e_\lambda, x_\lambda - y\rangle|}{t_0} \Big)^{-N} .$$

Using (3.43), (3.44), and $\omega_j \asymp \sigma^{-j(1-\alpha)} = t_\lambda^{1-\alpha}$, we obtain

$$\widetilde{S}_{j,\mathbf{y}}^{(m)} \lesssim \sum_{\ell \in \mathbb{L}_j^{(m)}} \sum_{k \in \mathbb{Z}^2} \Big( 1 + \Big| \ell \Big(\frac{t_\lambda}{t_0}\Big)^{1-\alpha} - a_1(m) \Big|^2 + \Big| k_2 \Big(\frac{t_\lambda}{t_0}\Big)^{\alpha} - a_2(\ell) \Big|^2 + \Big| k_1 \Big(\frac{t_\lambda}{t_0}\Big) - a_3(\ell) \Big| \Big)^{-N} \tag{3.46}$$

with the quantities

$$a_1(m) := t_0^{-(1-\alpha)}(\theta + m\pi), \qquad a_2(\ell) := t_0^{-\alpha} c_2^{-1} [R_{\eta_\lambda} y]_2, \qquad a_3(\ell) := t_0^{-1} c_1^{-1} [R_{\eta_\lambda} y]_1.$$

All these quantities vary with $j \in \mathbb{N}_0$ and $\mathbf{y} = (y, \theta, u) \in \mathbb{P}$. Furthermore, as indicated by the notation, $a_1$ is also dependent on $m$, whereas $a_2$ and $a_3$ also depend on $\ell$.

To proceed, we interpret the sum on the right as a Riemann sum, which is bounded up to a multiplicative constant by the corresponding integral. We obtain

$$\begin{aligned} \Big(\frac{t_\lambda}{t_0}\Big)^2 \widetilde{S}_{j,\mathbf{y}}^{(m)} &\lesssim \sum_{\ell \in \mathbb{Z}} \Big(\frac{t_\lambda}{t_0}\Big)^{1-\alpha} \sum_{k_1 \in \mathbb{Z}} \Big(\frac{t_\lambda}{t_0}\Big) \sum_{k_2 \in \mathbb{Z}} \Big(\frac{t_\lambda}{t_0}\Big)^{\alpha} \Big( 1 + \Big| \ell \Big(\frac{t_\lambda}{t_0}\Big)^{1-\alpha} - a_1(m) \Big|^2 \\ &\qquad + \Big| k_2 \Big(\frac{t_\lambda}{t_0}\Big)^{\alpha} - a_2(\ell) \Big|^2 + \Big| k_1 \Big(\frac{t_\lambda}{t_0}\Big) - a_3(\ell) \Big| \Big)^{-N} \\ &\lesssim \int_{\mathbb{R}} dy \int_{\mathbb{R}^2} dx \big(1 + |y|^2 + |x_2|^2 + |x_1|\big)^{-N}, \end{aligned}$$

where the integral is finite precisely if $N > 2$ (see Lemma 4.5.11 applied with $r = 1$, $\gamma = 0$). Hence, we arrive at

$$\widetilde{S}_{j,\mathbf{y}}^{(m)} \lesssim \Big(\frac{t_\lambda}{t_0}\Big)^{-2} = \max\Big\{ \frac{u}{t_\lambda}, 1 \Big\}^2,$$

and due to $t_\lambda / t_0 \le 1$ the implicit constant is independent of $j \in \mathbb{N}_0$ and $\mathbf{y} = (y, \theta, u) \in \mathbb{P}$. Since the number of summands $\widetilde{S}_{j,\mathbf{y}}^{(m)}$ does not exceed $2M+1$ the proof of part (i) is finished.

Part (ii): Let us now turn to the case when $(\Lambda, \Phi_\Lambda) = (\Lambda^s, \Phi^s)$ is an $\alpha$-shearlet parametrization, specified by a set of parameters $\sigma > 1$, $c = (c_1, c_2) \in \mathbb{R}_+^2$, $(\upsilon_j)_j$, and $(L_j^\varepsilon)_j$ as in (3.29). In this case, the sum $\widetilde{S}_{j,\mathbf{y}}$ from (3.42) can be split into three parts

$$\widetilde{S}_{j,\mathbf{y}}^{(\varepsilon)} := \sum_{\substack{\lambda \in \Lambda_\varepsilon^s \\ t_\lambda = \sigma^{-j}}} \big(1 + \widetilde{d}_\alpha(\Phi^s(\lambda), \mathbf{y})\big)^{-N}, \quad \varepsilon \in \{0, 1, 2\},$$

corresponding to the respective regions of the frequency domain. In the following, we handle these sums separately.

$\underline{\Lambda_0^s}$**:** We first treat $\varepsilon = 0$ and see that $\widetilde{S}^{(0)}_{j,\mathbf{y}}$, due to the definition of $\Lambda_0^s$, is an empty sum if $j > 0$. In case $j = 0$ we need to look at the partial sum

$$\widetilde{S}^{(0)}_{0,\mathbf{y}} = \sum_{\lambda\in\Lambda_0^s} \big(1 + \widetilde{d}_\alpha(\Phi^s(\lambda), \mathbf{y})\big)^{-N} = \sum_{k\in\mathbb{Z}^2} \big(1 + t_0^{-2\alpha}|M_c k - y|^2 + t_0^{-1}|\langle e_1, M_c k - y\rangle|\big)^{-N},$$

where here $t_0 = \max\{1, u\}$. Since $\alpha \le 1$ we obtain the bound

$$\widetilde{S}^{(0)}_{0,\mathbf{y}} \le \sum_{k\in\mathbb{Z}^2} \big(1 + t_0^{-2}|c_2k_2 - y_2|^2 + t_0^{-1}|c_1k_1 - y_1|\big)^{-N}.$$

The sum on the right-hand side can be interpreted as a Riemann sum. This enables the estimate

$$\widetilde{S}^{(0)}_{0,\mathbf{y}} \lesssim t_0^2 \int_{\mathbb{R}^2} \big(1 + |c_2x_2 - y_2|^2 + |c_1x_1 - y_1|\big)^{-N}\, dx \lesssim t_0^2.$$

Since $t_\lambda = 1$ and $t_0^2 = \max\{1, u\}^2$ we are finished.

$\underline{\Lambda_\varepsilon^s,\ \varepsilon \in \{1,2\}}$**:** For symmetry reasons, both partial sums for $\varepsilon \in \{1, 2\}$ can be treated in the same fashion. It therefore suffices to present the proof for the case $\varepsilon = 1$.

Since $L_j^1 \lesssim \sigma^{j(1-\alpha)}$ and $\upsilon_j \asymp \sigma^{-j(1-\alpha)}$, analogous to (3.45), there is a bound $B > 0$ with

$$|\ell\upsilon_j| \le B \text{ for all } j \in \mathbb{N}_0, \ell \in \mathbb{L}_j^1. \tag{3.47}$$

Putting $\delta := \arctan(B)$, we thus have $0 < \delta < \frac{\pi}{2}$ and

$$|\{\eta_\lambda\}| = |\arctan(-\ell\upsilon_j)| \le \delta$$

for $\eta_\lambda = (\arctan(-\ell\upsilon_j))_{2\pi}$. Recall the proof of Proposition 3.3.6, where we have shown that the transfer matrix $\widetilde{T}_\lambda = \widetilde{T}_{j,\ell} = R_{\eta_\lambda} S^{-1}_{\ell\upsilon_j}$ has the form

$$\widetilde{T}_\lambda = \widetilde{T}_{j,\ell} = \begin{pmatrix} \cos(\eta_\lambda) & 0 \\ \sin(\eta_\lambda) & \cos(\eta_\lambda)^{-1} \end{pmatrix}.$$

We know from (3.36) that the diagonal entries of $\widetilde{T}_\lambda$ are bounded by positive constants from above and below. Furthermore, the off-diagonal entry is bounded from above in modulus. This leads to

$$\begin{aligned} |\langle e_\lambda, x_\lambda - y\rangle| &= |\langle R^{-1}_{\eta_\lambda} e_1, S^{-1}_{\ell\upsilon_j} A_{\alpha,t_\lambda} M_c k - y\rangle| = |\langle e_1, \widetilde{T}_\lambda A_{\alpha,t_\lambda} M_c k - R_{\eta_\lambda} y\rangle| \\ &= |t_\lambda c_1 k_1 \cos(\eta_\lambda) - [R_{\eta_\lambda} y]_1| \asymp |t_\lambda k_1 - \cos(\eta_\lambda)^{-1}[c_1^{-1} R_{\eta_\lambda} y]_1|. \end{aligned} \tag{3.48}$$

Next, we estimate the term $|x_\lambda - y|$. We have $|S_{\ell\upsilon_j} x| \asymp |x|$ uniformly for $x \in \mathbb{R}^2$ and $j \in \mathbb{N}_0$, $\ell \in \mathbb{L}_j^1$, and therefore it holds

$$\begin{aligned} |x_\lambda - y| &= |S^{-1}_{\ell\upsilon_j} A_{\alpha,t_\lambda} M_c k - y| \asymp |M_c A_{\alpha,t_\lambda} k - S_{\ell\upsilon_j} y| \\ &\asymp |A_{\alpha,t_\lambda} k - M_c^{-1} S_{\ell\upsilon_j} y| \ge |t_\lambda^\alpha k_2 - [M_c^{-1} S_{\ell\upsilon_j} y]_2|. \end{aligned}$$

At last, we deal with the term $|\{\eta_\lambda - \theta\}|$. First, recall that $|\{\eta_\lambda\}| \le \delta = \arctan(B) < \frac{\pi}{2}$ for all $\lambda \in \Lambda_1^s$, where $B > 0$ is the bound from (3.47). Further, $|\{\theta\}| \le \frac{\pi}{2}$ for every $\theta \in \mathbb{T} = [0, 2\pi)$. Applying Lemma 3.4.1, we hence obtain uniformly for all $\lambda \in \Lambda_1^s$ and $\theta \in \mathbb{T}$

$$|\{\eta_\lambda - \theta\}| = |\{\{\eta_\lambda\} - \{\theta\}\}| \asymp |\{\eta_\lambda\} - \{\theta\}|.$$

Next, we distinguish between those $\theta \in \mathbb{T}$ with $|\{\theta\}| \leq \arctan(B+1)$ and those with $|\{\theta\}| > \arctan(B+1)$.

For $|\{\theta\}| \leq \arctan(B+1)$, we use Lemma 3.4.2 and $\tan(\theta) = \tan(\{\theta\})$ to obtain

$$|\{\eta_\lambda - \theta\}| \asymp |\{\eta_\lambda\} - \{\theta\}| = |\arctan(-\ell\upsilon_j) - \{\theta\}| \asymp |\ell\upsilon_j + \tan(\theta)|.$$

In case $|\{\theta\}| > \arctan(B+1)$ we estimate directly, using $|\{\eta_\lambda\}| \leq \arctan(B)$,

$$|\{\eta_\lambda - \theta\}| \asymp |\{\eta_\lambda\} - \{\theta\}| \geq ||\{\eta_\lambda\}| - |\{\theta\}|| > \arctan(B+1) - \arctan(B) > 0.$$

Since $|\ell\upsilon_j| \leq B$ according to (3.47) this implies $|\{\eta_\lambda - \theta\}| \gtrsim |\ell\upsilon_j|$.

We now introduce the quantity

$$q(\theta) := \begin{cases} \tan(\theta) & , |\{\theta\}| \leq \arctan(B+1), \\ 0 & , |\{\theta\}| > \arctan(B+1). \end{cases}$$

Then we can summarize

$$|\{\eta_\lambda - \theta\}| \gtrsim |\ell\upsilon_j + q(\theta)|. \tag{3.49}$$

In view of the estimates (3.48)-(3.49) and $\upsilon_j \asymp \sigma^{-j(1-\alpha)} \asymp t_\lambda^{1-\alpha}$, we obtain

$$\begin{aligned} \widetilde{S}^{(1)}_{j,\mathbf{y}} &\lesssim \sum_{\ell \in \mathbb{L}_j} \sum_{k \in \mathbb{Z}^2} \Big(1 + \frac{|\ell\upsilon_j + q(\theta)|^2}{t_0^{2(1-\alpha)}} + \frac{|t_\lambda^\alpha k_2 - [c_2^{-1} S_{\ell\upsilon_j} y]_2|^2}{t_0^{2\alpha}} + \frac{|t_\lambda k_1 - [c_1^{-1} R_{\eta_\lambda} y]_1 / \cos(\eta_\lambda)|}{t_0}\Big)^{-N} \\ &\lesssim \sum_{\ell \in \mathbb{Z}} \sum_{k \in \mathbb{Z}^2} \Big(1 + \Big|\ell\Big(\frac{t_\lambda}{t_0}\Big)^{(1-\alpha)} - a_1(\theta)\Big|^2 + \Big|k_2\Big(\frac{t_\lambda}{t_0}\Big)^{\alpha} - a_2(\ell)\Big|^2 + \Big|k_1\Big(\frac{t_\lambda}{t_0}\Big) - a_3(\ell)\Big|\Big)^{-N} \end{aligned}$$

with the quantities

$$a_1(\theta) := -t_0^{-(1-\alpha)} q(\theta), \quad a_2(\ell) := t_0^{-\alpha}[c_2^{-1} S_{\ell\upsilon_j} y]_2, \quad a_3(\ell) := t_0^{-1}[c_1^{-1} R_{\eta_\lambda} y]_1 / \cos(\eta_\lambda),$$

depending on $j$, $\ell$, $\theta$ and $\mathbf{y}$. This expression is similar to (3.46). Therefore, from here we can proceed as in part (i) of the proof. □

As a corollary of Theorem 3.4.3 we obtain the desired consistency of $\alpha$-curvelet and $\alpha$-shearlet parametrizations.

**Corollary 3.4.4** ([59, Thm. 5.7])**.** *Let $\alpha \in [0,1]$, and let $(\Lambda, \Phi_\Lambda)$ be either an $\alpha$-curvelet or an $\alpha$-shearlet parametrization. Then, any other $\alpha$-curvelet or $\alpha$-shearlet parametrization $(\Delta, \Phi_\Delta)$, with possibly different parameters, is $(\alpha, k)$-consistent to $(\Lambda, \Phi_\Lambda)$ for $k > 2$.*

*Proof.* This is a direct consequence of Theorem 3.4.3 and the quasi-symmetry of $\omega_\alpha$. □

The consistency plays an important role for the application of the transfer principle. In particular, Corollary 3.4.4 will be used in Chapters 5 and 6, where we analyze the cartoon approximation capabilities of discrete $\alpha$-molecule frames.

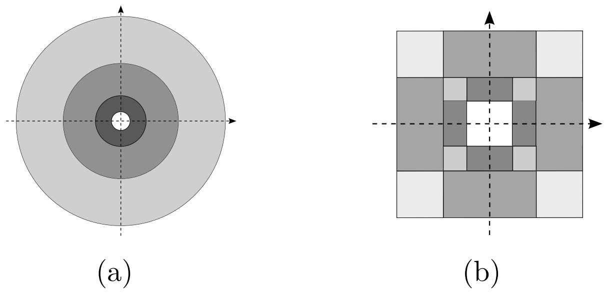


Figure 3.3: (a): Partition of the Fourier domain induced by radial wavelets. (b): Partition of the Fourier domain induced by tensor wavelets.

## 3.5 Wavelet Systems

In a strict sense, wavelet systems do not belong to the class of directional representation systems since they are isotropically scaled. Nevertheless, the framework of $\alpha$-molecules also covers such isotropic systems for the case $\alpha = 1$. It turns out that many wavelet constructions can be subsumed under the notion of 1-curvelet molecules.

This justifies the following definition.

**Definition 3.5.1.** A system of 1-curvelet molecules shall also be referred to as a system of *wavelet molecules*. The associated 1-curvelet parametrization is then simply called a *wavelet parametrization*.

We will subsequently consider two different types of wavelet systems in $L_2(\mathbb{R}^2)$, namely radial wavelet systems and tensor wavelet systems.

### 3.5.1 Radial Wavelets

A typical radial wavelet system in $L_2(\mathbb{R}^2)$ is given as follows.

**Definition 3.5.2.** Let $\psi_0$, $\psi$ be radial functions in $L_2(\mathbb{R}^2)$. Further, let $\sigma > 1$, $c = (c_1, c_2) \in \mathbb{R}^2_+$ be fixed parameters. The associated *radial wavelet system* $W_{\mathrm{rad}}(\psi_0, \psi; \sigma, c)$ is then defined by

$$W_{\mathrm{rad}}(\psi_0, \psi; \sigma, c) := \Big\{ \psi_{j,k} \;:\; j \in \mathbb{N}_0,\, k \in \mathbb{Z}^2 \Big\},$$

where

$$\psi_{0,k}(\cdot) := \psi_0(\cdot - M_c k) \quad \text{and} \quad \psi_{j,k}(\cdot) := \sigma^j \psi(\sigma^j \cdot - M_c k) \quad \text{for } j \geq 1.$$

Here $M_c$ stands for the matrix (3.10).

The index set associated to such a radial wavelet system $W_{\mathrm{rad}}(\psi_0, \psi; \sigma, c)$ shall be denoted by

$$\Lambda^w_{\mathrm{rad}} := \{(j,k) \;:\; j \in \mathbb{N}_0,\, k \in \mathbb{Z}^2\}.$$

We subsequently analyze those systems in more detail where the generators $\psi_0, \psi \in L_2(\mathbb{R}^2)$ are bandlimited and where $\psi$ has infinitely many vanishing moments. Concretely, we assume that the functions $\psi_0, \psi$ satisfy

$$\hat{\psi}_0,\, \hat{\psi} \in C^L(\mathbb{R}^2) \qquad \text{for some } L \in \mathbb{N}_0 \cup \{\infty\},$$

and that there exist $0 < a$ and $0 < b < c$ such that

$$\operatorname{supp}\hat{\psi}_0 \subseteq \mathcal{B}_a := \{\xi \in \mathbb{R}^2 : |\xi| \le a\} \quad \text{and} \quad \operatorname{supp}\hat{\psi} \subseteq \mathcal{C}_{b,c} := \{\xi \in \mathbb{R}^2 : b \le |\xi| \le c\}.$$

According to the following proposition, under these assumptions $W_{\mathrm{rad}}(\psi_0, \psi; \sigma, c)$ is a system of 1-molecules of order $(L, \infty, \infty, \infty)$. The corresponding tiling of the frequency plane is illustrated in Figure 3.3 (a).

**Proposition 3.5.3.** *Suppose that the generators of the radial wavelet system $W_{\mathrm{rad}}(\psi_0, \psi; \sigma, c)$ fulfill the conditions specified above. Then this system is a system of* 1*-molecules of order $(L, \infty, \infty, \infty)$ with respect to the* radial wavelet parametrization $(\Lambda^w_{\mathrm{rad}}, \Phi^w_{\mathrm{rad}})$ *with parametrization map*

$$\Phi^w_{\mathrm{rad}} : \Lambda^w_{\mathrm{rad}} \to \mathbb{P}, \quad (j,k) \mapsto (\sigma^{-j} M_c k, 0, \sigma^{-j}).$$

*Proof.* The proof is analogous to the proof of Proposition 3.2.8 (ii). □

We next observe that the system $W_{\mathrm{rad}}(\psi_0, \psi; \sigma, c)$ can even be interpreted as a system of 1-curvelet molecules of order $(L, \infty, \infty, \infty)$, with associated parametrization $(\Lambda^c, \Phi^c)$ and parameters $\sigma$, $c = (c_1, c_2)$, as well as $L_j := 0$ and $\omega_j := 2\pi$ for every $j \in \mathbb{N}_0$ (see (3.11) and (3.13)). For this, we just need to relabel the elements of $W_{\mathrm{rad}}(\psi_0, \psi; \sigma, c)$ via the bijection

$$\iota_{\mathrm{rad}} : \Lambda^w_{\mathrm{rad}} \to \Lambda^c, \quad (j,k) \mapsto (j, 0, k).$$

The relation between the radial wavelet parametrization $(\Lambda^w_{\mathrm{rad}}, \Phi^w_{\mathrm{rad}})$ and the corresponding 1-curvelet parametrization $(\Lambda^c, \Phi^c)$ is then given by

$$\Phi^w_{\mathrm{rad}} = \Phi^c \circ \iota_{\mathrm{rad}}.$$

As an immediate consequence, we can derive the following consistency result from Theorem 3.4.3.

**Proposition 3.5.4.** *Let $N > 2$. Then, with $(\Lambda^w_{\mathrm{rad}}, \Phi^w_{\mathrm{rad}})$ being the radial wavelet parametrization, we have*

$$\sup_{\mathbf{y} \in \mathbb{P}} \sum_{\lambda \in \Lambda^w_{\mathrm{rad}}} \omega_\alpha(\Phi^w_{\mathrm{rad}}(\lambda), \mathbf{y})^{-N} < \infty.$$

In particular, the parametrization $(\Lambda^w_{\mathrm{rad}}, \Phi^w_{\mathrm{rad}})$ is $(1,k)$-consistent with other 1-curvelet and 1-shearlet parametrizations for $k > 2$ (compare Corollary 3.4.4).

### 3.5.2 Tensor Wavelets

Another important class of wavelets in $L_2(\mathbb{R}^2)$ is obtained via tensoring of univariate wavelets. We subsequently recall the tensor product construction from [114]. It builds upon a given multi-resolution analysis for $L_2(\mathbb{R})$ whose scaling function and associated wavelet shall be denoted by $\phi^0 \in L_2(\mathbb{R})$ and $\phi^1 \in L_2(\mathbb{R})$, respectively. For every index $e = (e_1, e_2) \in E$, where $E = \{0,1\}^2$, one then defines the functions $\psi^e \in L_2(\mathbb{R}^2)$ as the tensor products

$$\psi^e = \phi^{e_1} \otimes \phi^{e_2}. \tag{3.50}$$

These serve as the generators for the tensor wavelet system $W_{\mathrm{ten}}(\phi^0, \phi^1; \sigma, c)$ defined below.

**Definition 3.5.5.** Let $\phi^0$, $\phi^1 \in L_2(\mathbb{R})$ and $\psi^e \in L_2(\mathbb{R}^2)$, $e \in E$, be defined as above. Further, let $\sigma > 1$, $c = (c_1, c_2) \in \mathbb{R}^2_+$ be fixed parameters. The associated *tensor wavelet system* $W_{\text{ten}}(\phi^0, \phi^1; \sigma, c)$ is then defined by

$$W_{\text{ten}}(\phi^0, \phi^1; \sigma, c) := \Big\{\psi^{(0,0)}(\cdot - M_c k) \;:\; k \in \mathbb{Z}^2\Big\} \\ \cup \Big\{\sigma^j \psi^e(\sigma^j \cdot - M_c k) \;:\; e \in E\backslash\{(0,0)\},\, j \in \mathbb{N}_0,\, k \in \mathbb{Z}^2\Big\}.$$

Let us now assume that the generating functions $\phi^0, \phi^1 \in L_2(\mathbb{R})$ satisfy

$$\hat{\phi}^0,\, \hat{\phi}^1 \in C^L(\mathbb{R}) \qquad \text{for some } L \in \mathbb{N}_0 \cup \{\infty\}, \tag{3.51}$$

and that there exist $0 < a$ and $0 < b < c$ such that

$$\operatorname{supp} \hat{\phi}^0 \subset [-a, a] =: J^{(0)} \quad \text{and} \quad \operatorname{supp} \hat{\phi}^1 \subset [-c, c]\backslash[-b, b] =: J^{(1)}. \tag{3.52}$$

Then $W_{\text{ten}}(\phi^0, \phi^1; \sigma, c)$ induces a frequency tiling as in Figure 3.3 (b) and, as shown by the following proposition, this tensor wavelet system is a special instance of a 1-molecule system.

**Proposition 3.5.6.** *Let $\sigma > 1$, $c = (c_1, c_2) \in \mathbb{R}^2_+$ be fixed, and assume that the functions $\phi^0$, $\phi^1$ satisfy* (3.51) *and* (3.52). *Then the tensor wavelet system $W_{\text{ten}}(\phi^0, \phi^1; \sigma, c)$ constitutes a system of* 1*-molecules of order $(L, \infty, \infty, \infty)$ with respect to the* tensor wavelet parametrization $(\Lambda^w_{\text{ten}}, \Phi^w_{\text{ten}})$ *where*

$$\Lambda^w_{\text{ten}} := \{((0,0), 0, k) \;:\; k \in \mathbb{Z}^2\} \cup \{(e, j, k) \;:\; e \in E\backslash\{(0,0)\},\, j \in \mathbb{N}_0,\, k \in \mathbb{Z}^2\}$$

*and*

$$\Phi^w_{\text{ten}} : \Lambda^w_{\text{ten}} \to \mathbb{P}, \quad (e, j, k) \mapsto (\sigma^{-j} M_c k, 0, \sigma^{-j}).$$

*Proof.* For $(e, j, k) \in \Lambda^w_{\text{ten}}$ we define the generators $g_{e,j,k} := \psi^e$, with $\psi^e$ being the functions from (3.50). We then have $\hat{g}_{e,j,k} = \hat{\psi}^e \in C^L(\mathbb{R}^2)$ by (3.51). Further, (3.52) implies that

$$\operatorname{supp} \widehat{\psi}^e \subseteq J^e := J^{(e_1)} \times J^{(e_2)} \qquad \text{for all } e = (e_1, e_2) \in E.$$

Hence $\operatorname{supp}(\partial^\rho \hat{g}_{e,j,k}) \subseteq J^e$ for every $\rho \in \mathbb{N}^2_0$ with $|\rho|_1 \le L$ and for all $(e, j, k) \in \Lambda^w_{\text{ten}}$. Further, the expression $\sup_{\xi \in \mathbb{R}^2} |\partial^\rho \hat{g}_{e,j,k}(\xi)| = \sup_{\xi \in \mathbb{R}^2} |\partial^\rho \hat{\psi}^e(\xi)|$ is bounded uniformly in $(e, j, k) \in \Lambda^w_{\text{ten}}$. Altogether, this proves that the functions $g_{e,j,k}$ satisfy condition (2.5). Since the wavelets can be written in the form

$$\psi^e_{j,k}(\cdot) := \sigma^j \psi^e(\sigma^j \cdot - M_c k) = \sigma^j g_{e,j,k}(\sigma^j(\cdot - \sigma^{-j} M_c k)),$$

the proof is finished. □

We remark that conditions as in (3.51) and (3.52) are fulfilled, for instance, if $\phi^0, \phi^1 \in L_2(\mathbb{R})$ are the generators of a Lemarié-Meyer wavelet system. Moreover, results similar to Proposition 3.5.6 can be proven for other wavelet systems of the form $W_{\text{ten}}(\phi^0, \phi^1; \sigma, c)$, including systems generated by compactly supported functions $\phi^0$ and $\phi^1$.

Finally, let us again interpret $W_{\text{ten}}(\phi^0, \phi^1; \sigma, c)$ as an instance of a 1-curvelet molecule system with parametrization $(\Lambda^c, \Phi^c)$. To this end, we choose $L_j := 2$ and $\omega_j := 2\pi$ for all

$j \in \mathbb{N}_0$. Further, we use $\mathbb{L}_0 := \{-1, 0, 1, 2\}$ and $\mathbb{L}_j := \{-1, 0, 1\}$ for $j \geq 1$ in the definition (3.12) of the 1-curvelet index set $\Lambda^c$. Then

$$\iota_{\text{ten}} : \Lambda^w_{\text{ten}} \to \Lambda^c, \quad (e, j, k) \mapsto \begin{cases} (j, e_1 - e_2, k), & e \in E \backslash \{(0,0)\}, \\ (0, 2, k), & e = (0,0), \end{cases}$$

is a bijection and

$$\Phi^w_{\text{ten}} = \Phi^c \circ \iota_{\text{ten}}.$$

Analogous to Proposition 3.5.4, we can derive the canonical consistency of the tensor wavelet parametrization.

**Proposition 3.5.7.** *Let $N > 2$. Then, with $(\Lambda^w_{\text{ten}}, \Phi^w_{\text{ten}})$ being the tensor wavelet parametrization, we have*

$$\sup_{\mathbf{y} \in \mathbb{P}} \sum_{\lambda \in \Lambda^w_{\text{ten}}} \omega_\alpha(\Phi^w_{\text{ten}}(\lambda), \mathbf{y})^{-N} < \infty.$$

Hence, like the radial wavelet parametrization, the tensor wavelet parametrization is $(1, k)$-consistent for $k > 2$ with other 1-curvelet and 1-shearlet parametrizations. Moreover, this result shows that both wavelet parametrizations are $(1, k)$-consistent, $k > 2$, with each other.

## 3.6 Ridgelet Systems

The last section of this chapter is devoted to ridgelet systems. Whereas there does not exist a common definition of a ridgelet, the different variants of this notion that occur in the literature are all related to the concept of a so-called *ridge function*. In the bivariate setting, this is a function $\Phi : \mathbb{R}^2 \to \mathbb{C}$ which only varies in one coordinate direction and can thus be represented in the form $\Phi = \phi(\langle \nu, \cdot \rangle)$ using a suitable univariate profile $\phi$ and a direction vector $\nu \in \mathbb{R}^2$.

The term 'ridgelet' was first used by Candès in [8] to refer to such bivariate ridge functions forming a system of the type

$$\sqrt{t}\phi(\langle te_\eta, \cdot \rangle - x), \quad e_\eta \in \mathbb{S}^1, x \in \mathbb{R}^2, t \in \mathbb{R}_+, \tag{3.53}$$

where the profile $\phi$ is a univariate wavelet. Systems of this kind can be used for example to analyze functions and give rise to the so-called ridgelet transform. A viable theory for this transform has been developed in [8]. One difficulty, however, when dealing with these 'pure ridgelets' is their lack of integrability.

In order to obtain system in $L_2(\mathbb{R}^2)$, Donoho slightly relaxed the original definition, allowing the ridgelets a slow decay along the ridge. Using this idea, he constructed an orthonormal basis for $L_2(\mathbb{R}^2)$ whose elements he called 'orthonormal ridgelets' [36]. Their relationship to the original 'pure ridgelets' in the sense of Candès has been analyzed in [37].

An alternative approach to define ridgelet systems in $L_2(\mathbb{R}^2)$ goes back to Grohs [57]. He considers function systems of the form

$$y \mapsto \sqrt{t}\psi(A_{0,t}R_\eta y - x), \tag{3.54}$$

obtained by applying dilations $A_{0,t} = \operatorname{diag}(t,1) \in \mathbb{R}^{2\times 2}$ with $t \in \mathbb{R}_+$, rotations $R_\eta$, $\eta \in \mathbb{T}$, and translations to some generator $\psi \in L_2(\mathbb{R}^2)$, which is assumed to be oscillatory in one coordinate direction. The construction principle is thus the same as for 0-curvelets and closely resembles (3.53). In fact, the ridgelet construction in [57] more or less coincides with the 0-curvelet frame $\mathfrak{C}_0^\bullet$ from Subsection 3.2.3.

To ensure the frame property, the scaling is not carried out in Cartesian coordinates, but in polar coordinates. This causes the ridgelet generators to vary with the scale as was the case for $\mathfrak{C}_0^\bullet$ (see (3.25)). By relaxing the rigid construction principle (3.54) and allowing variable generators, one then again arrives at the notion of 0-curvelet molecules.

For convenience, we thus make the following definition.

**Definition 3.6.1.** A system of 0-curvelet molecules is also called a system of *ridgelet molecules*. The associated 0-curvelet parametrization is then accordingly referred to as a *ridgelet parametrization*.

Due to Proposition 3.2.4, ridgelet molecules in the above sense are special instances of 0-molecules. Further, due to Proposition 3.2.8, the 0-curvelet frame $\mathfrak{C}_0^\bullet$ from Definition 3.2.6 is a special case of a ridgelet molecule system.

**Proposition 3.6.2** ([59, Prop. 3.5])**.** *The* 0*-curvelet frame* $\mathfrak{C}_0^\bullet$ *is a system of ridgelet molecules of order* $(\infty,\infty,\infty,\infty)$*.*

Summarizing, we can record that ridgelet-type systems in the sense of (3.54) conveniently fit into the already existing theory. They are covered by the notion of $\alpha$-curvelet molecules for $\alpha = 0$.

# Chapter 4

# $\alpha$-Molecule Coorbit Spaces

In this chapter, we build upon the continuous Parseval frame of $\alpha$-curvelets $\mathfrak{C}_\alpha = \{\psi_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ from Section 3.1 to introduce an associated transform which is a direct generalization of the continuous curvelet transform from [10, Sec. 2]. Subsequently it will be called the continuous $\alpha$-curvelet transform.

Utilizing the coorbit theory put forward in [74], which does not require an underlying group structure of the voice transform, we then define associated $\alpha$-curvelet coorbit spaces. Further, based on the more general concept of a continuous $\alpha$-molecule transform, which in particular comprises the cone-adapted $\alpha$-shearlet transform, we will also introduce so-called $\alpha$-molecule coorbit spaces.

In Theorem 4.3.8 it is shown that those are equivalent to the $\alpha$-curvelet coorbits. In Theorem 4.3.13 we further give a discrete characterization which identifies them with known smoothness spaces, for example from [85]. As an application of the abstract machinery available for coorbit spaces, we deduce two further discretization results, Theorem 4.4.19 and Theorem 4.4.21, yielding atomic decompositions as well as quasi-Banach frames.

## 4.1 The Continuous $\alpha$-Curvelet Transform

In Section 3.1 of Chapter 3 we have constructed the continuous Parseval frame of $\alpha$-curvelets $\mathfrak{C}_\alpha = \{\psi_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$, whose index set

$$\mathbb{X} = \mathbb{R}^2 \times \mathbb{T} \times (0, 1]$$

can be viewed as a subspace of the parameter domain $\mathbb{P}$ defined in (2.1). However, recall that the topology on $\mathbb{X}$ shall not be the subspace topology induced by $\mathbb{P}$. Instead, we will think of $\mathbb{X}$ as being assembled as a disconnected union $\mathbb{X} = \mathbb{X}_0 \cup \mathbb{X}_1$ of two components, namely the *homogeneous component* $\mathbb{X}_0$ and the *inhomogeneous component* $\mathbb{X}_1$ given by

$$\mathbb{X}_0 := \mathbb{R}^2 \times \mathbb{T} \times (0, 1) \quad \text{and} \quad \mathbb{X}_1 := \mathbb{R}^2 \times \mathbb{T} \times \{1\}.$$

Each of these components shall thereby carry the subspace topology inherited from $\mathbb{P}$. As noted in Section 3.1, $\mathbb{X}$ is then a locally compact Hausdorff space.

Further, $\mathbb{X}$ is equipped with the Radon measure $\mu$ defined by (3.7) satisfying $\operatorname{supp}\mu = \mathbb{X}$. Its restrictions to the components $\mathbb{X}_0$ and $\mathbb{X}_1$ are denoted by $\mu_0$ and $\mu_1$. They are given by

$$d\mu_0(x, \eta, t) = \frac{dx\, d\eta\, dt}{t^3} \quad \text{and} \quad d\mu_1(x, \eta, 1) = d\eta\, dt.$$

We have already shown in Section 3.1 that with respect to this measure $\mathfrak{C}_\alpha$ is a continuous Parseval frame satisfying the Parseval identity (3.5) and the reconstruction formula (3.6). These are important relations, since in the following we want to use the frame $\mathfrak{C}_\alpha$ as a tool for signal analysis.

### 4.1.1 Basic Transform on $L_2(\mathbb{R}^2)$

In the context of coorbit theory, the analysis operator of a frame is usually called the *voice transform.* For the frame $\mathfrak{C}_\alpha = \{\psi_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ it takes the form

$$V_{\mathfrak{C}_\alpha} : L_2(\mathbb{R}^2) \to L_2(\mathbb{X})\,, \qquad V_{\mathfrak{C}_\alpha} f(\mathbf{x}) := \langle f, \psi_{\mathbf{x}}\rangle, \quad \mathbf{x} \in \mathbb{X}. \tag{4.1}$$

Subsequently it will be called the *continuous $\alpha$-curvelet transform.* Since $\mathfrak{C}_\alpha$ is a Parseval frame it defines an isometry from $L_2(\mathbb{R}^2)$ to $L_2(\mathbb{X})$.

The corresponding synthesis operator $V^*_{\mathfrak{C}_\alpha}$ is the Hilbert-adjoint of $V_{\mathfrak{C}_\alpha}$ given by

$$V^*_{\mathfrak{C}_\alpha} : L_2(\mathbb{X}) \to L_2(\mathbb{R}^2)\,, \qquad V^*_{\mathfrak{C}_\alpha} F = \int_{\mathbb{X}} F(\mathbf{x})\psi_{\mathbf{x}}\, d\mu(\mathbf{x}).$$

Hereby the integral is in general only defined in a weak sense.

Due to the Parseval property, the associated frame operator, which by definition is the composition of the analysis and the synthesis operator, is the identity. In other words, for all $f \in L_2(\mathbb{R}^2)$ we have the reconstruction formula

$$f = V^*_{\mathfrak{C}_\alpha} V_{\mathfrak{C}_\alpha} f = \int_{\mathbb{X}} V_{\mathfrak{C}_\alpha} f(\mathbf{x})\psi_{\mathbf{x}}\, d\mu(\mathbf{x}), \tag{4.2}$$

where again the integral is usually only a weak integral in $L_2(\mathbb{R}^2)$.

Let us finally take a look at the associated Gramian matrix $\mathcal{G}[\mathfrak{C}_\alpha]$ introduced in (2.7). It has the entries

$$\mathcal{G}[\mathfrak{C}_\alpha](\mathbf{x},\mathbf{y}) = \langle \psi_{\mathbf{y}}, \psi_{\mathbf{x}}\rangle\,, \quad \mathbf{x},\mathbf{y} \in \mathbb{X},$$

and gives rise to the so-called *Gramian operator* $\mathcal{G}[\mathfrak{C}_\alpha] : L_2(\mathbb{X}) \to L_2(\mathbb{X})$ with

$$\mathcal{G}[\mathfrak{C}_\alpha]F(\mathbf{x}) = \int_{\mathbb{X}} \mathcal{G}[\mathfrak{C}_\alpha](\mathbf{x},\mathbf{y})F(\mathbf{y})\, d\mu(\mathbf{y})\,, \quad \mathbf{x} \in \mathbb{X}.$$

Taking the role of an integral kernel, the Gramian matrix is often referred to as the *Gramian kernel.* Also note that, for simplicity, we do not distinguish between the Gramian kernel and the Gramian operator in the notation.

From (4.2) we can derive the following reproducing formula, valid for all $f \in L_2(\mathbb{R}^2)$,

$$V_{\mathfrak{C}_\alpha} f(\mathbf{x}) = \int_{\mathbb{X}} \mathcal{G}[\mathfrak{C}_\alpha](\mathbf{x},\mathbf{y}) V_{\mathfrak{C}_\alpha} f(\mathbf{y})\, d\mu(\mathbf{y})\,, \quad \mathbf{x} \in \mathbb{X}. \tag{4.3}$$

We next turn to an extension of $V_{\mathfrak{C}_\alpha}$ to the space of tempered distributions $\mathcal{S}'(\mathbb{R}^2)$.

### 4.1.2 Extension to $\mathcal{S}'(\mathbb{R}^2)$

As it stands, the $\alpha$-curvelet transform (4.1) is only defined for square-integrable functions $f \in L_2(\mathbb{R}^2)$. For applications in signal analysis, this is a severe limitation since many signals of interest are not square-integrable. Therefore, in order to enhance the applicability of $V_{\mathfrak{C}_\alpha}$, we need to find a way to extend its definition beyond $L_2(\mathbb{R}^2)$.

A larger reservoir of signals is given by the *space of tempered distributions* $\mathcal{S}'(\mathbb{R}^2)$, the topological dual of the *Schwartz space* $\mathcal{S}(\mathbb{R}^2)$. Its elements, the so-called *tempered distributions*, are the continuous linear functionals on $\mathcal{S}(\mathbb{R}^2)$ which is the space of functions

$$\mathcal{S}(\mathbb{R}^2) := \Big\{ f \in C^\infty(\mathbb{R}^2) \;:\; |f|_{\kappa,\nu} < \infty \text{ for all } (\kappa,\nu) \in \mathbb{N}_0^2 \times \mathbb{N}_0^2 \Big\},$$

topologized by the family of semi-norms

$$|f|_{\kappa,\nu} := \sup_{x\in\mathbb{R}^2} \left| x^\kappa \partial^\nu f(x) \right|, \quad \kappa,\nu \in \mathbb{N}_0^2. \tag{4.4}$$

Equipped with the topology induced by the collection $\{|\cdot|_{\kappa,\nu} : \kappa,\nu\in\mathbb{N}_0^2\}$, the space $\mathcal{S}(\mathbb{R}^2)$ becomes a locally convex Hausdorff space.

An alternative way to define the topology of $\mathcal{S}(\mathbb{R}^2)$ is to use the collection of norms $\{\|\cdot\|_N : N\in\mathbb{N}_0\}$ given by

$$\|f\|_N := \sup_{x\in\mathbb{R}^2} (1+|x|)^N \sum_{|\gamma|\le N} |\partial^\gamma f(x)|. \tag{4.5}$$

This collection constitutes a complete filtrating family of norms (see e.g. [72]) generating the topology of $\mathcal{S}(\mathbb{R}^2)$. Using these norms, we can easily derive a metric for $\mathcal{S}(\mathbb{R}^2)$ which is consistent with this topology. It is given by

$$d(f,g) := \sum_{N\in\mathbb{N}_0} 2^{-N} \frac{\|f-g\|_N}{1+\|f-g\|_N} \quad \text{for } f,g\in\mathcal{S}(\mathbb{R}^2).$$

The relation of the spaces $\mathcal{S}(\mathbb{R}^2)$ and $\mathcal{S}'(\mathbb{R}^2)$ to the Hilbert space $L_2(\mathbb{R}^2)$ is illustrated by the following chain of embeddings

$$\mathcal{S}(\mathbb{R}^2) \overset{\iota}{\hookrightarrow} L_2(\mathbb{R}^2) \overset{R}{\hookrightarrow} L_2(\mathbb{R}^2)' \overset{\iota^*}{\hookrightarrow} \mathcal{S}'(\mathbb{R}^2), \tag{4.6}$$

where $\iota : \mathcal{S}(\mathbb{R}^2)\to L_2(\mathbb{R}^2)$ is the canonical injection, $\iota^* : L_2(\mathbb{R}^2)'\to\mathcal{S}'(\mathbb{R}^2)$ its adjoint, and $R : L_2(\mathbb{R}^2)\to L_2(\mathbb{R}^2)'$ denotes the Riesz map between $L_2(\mathbb{R}^2)$ and its dual $L_2(\mathbb{R}^2)'$, i.e., the canonical conjugate-linear isomorphism given by $f\mapsto\langle\cdot,f\rangle$.

The duality product $\langle\cdot,\cdot\rangle_{\mathcal{S}'\times\mathcal{S}}$ on the pair $\mathcal{S}'(\mathbb{R}^2)\times\mathcal{S}(\mathbb{R}^2)$ provides a natural way to extend $V_{\mathfrak{C}_\alpha}$ to $\mathcal{S}'(\mathbb{R}^2)$. Crucial for this is the observation that $\mathfrak{C}_\alpha$ is contained in $\mathcal{S}(\mathbb{R}^2)$. In fact, the $\alpha$-curvelets $\psi_{x,\eta,t}\in\mathfrak{C}_\alpha$ are band-limited functions which even satisfy $\hat{\psi}_{x,\eta,t}\in C_c^\infty(\mathbb{R}^2)$. This implies $\psi_{x,\eta,t}\in\mathcal{S}(\mathbb{R}^2)$ for all $(x,\eta,t)\in\mathbb{X}$ and allows to define the extended transform $V_{\mathfrak{C}_\alpha}$ for $f\in\mathcal{S}'(\mathbb{R}^2)$ as

$$V_{\mathfrak{C}_\alpha} f(\mathbf{x}) := \langle f,\psi_{\mathbf{x}}\rangle_{\mathcal{S}'\times\mathcal{S}}, \quad \mathbf{x}\in\mathbb{X}. \tag{4.7}$$

This transform $V_{\mathfrak{C}_\alpha}$ "extends" the $L_2$-version from (4.1) in the sense depicted in the following commutative diagram:

$$\begin{array}{ccc} L_2(\mathbb{R}^2) & \xrightarrow{V_{\mathfrak{C}_\alpha}} & L_2(\mathbb{X}) \\ R\big\downarrow & & \big\downarrow\overline{(\cdot)} \\ L_2(\mathbb{R}^2)' & \xrightarrow{V_{\mathfrak{C}_\alpha}} & L_2(\mathbb{X}) \end{array} \quad . \tag{4.8}$$

However, even with this extended definition of $V_{\mathfrak{C}_\alpha}$ at hand, it is not yet clear how useful this transform actually is for the analysis of signals in $\mathcal{S}'(\mathbb{R}^2)$. In particular, it is not a-priori self-evident that $V_{\mathfrak{C}_\alpha}$ is still injective, a prerequisite if we want to be able to uniquely reconstruct signals from their transforms.

Fortunately, it turns out that even in its extended form the injectivity of the transform $V_{\mathfrak{C}_\alpha}$ is guaranteed. This is a consequence of the fact that $\mathfrak{C}_\alpha$ is total in $\mathcal{S}(\mathbb{R}^2)$, i.e., that the linear span of $\mathfrak{C}_\alpha$ is dense in $\mathcal{S}(\mathbb{R}^2)$. This will be the statement of Lemma 4.1.3, but before we can give a proof of this we need some preparation.

For every $N \in \mathbb{N}_0$, let us introduce the auxiliary spaces

$$\mathcal{B}_N(\mathbb{R}^2) := \Big\{ f \in C^N(\mathbb{R}^2) \ : \ \|f\|_N < \infty \Big\}$$

where $\|\cdot\|_N$ is given as in (4.5). These spaces are Banach spaces as the following lemma shows. They are useful since they "approximate" $\mathcal{S}(\mathbb{R}^2)$ in the following sense:

$$\mathcal{B}_{N+1}(\mathbb{R}^2) \subset \mathcal{B}_N(\mathbb{R}^2) \quad \text{for } N \in \mathbb{N}_0 \qquad \text{and} \quad \mathcal{S}(\mathbb{R}^2) = \bigcap_{N \in \mathbb{N}_0} \mathcal{B}_N(\mathbb{R}^2).$$

Moreover, the family of nested spaces $\{\mathcal{B}_N(\mathbb{R}^2)\}_{N \in \mathbb{N}_0}$ captures the topology of $\mathcal{S}(\mathbb{R}^2)$.

**Lemma 4.1.1.** *For each $N \in \mathbb{N}_0$ the space $\mathcal{B}_N(\mathbb{R}^2)$ equipped with the norm $\|\cdot\|_N$ is a Banach space. Moreover, if $N \geq 2$ it is continuously and densely embedded into $L_2(\mathbb{R}^2)$.*

*Proof.* The vector space properties of $\mathcal{B}_N$ are obvious. Further, $\|\cdot\|_N$ clearly defines a norm on $\mathcal{B}_N$. To prove the completeness of $\mathcal{B}_N$, note that $B_N \hookrightarrow C^N$. Hence, every Cauchy sequence $(f_n)_{n \in \mathbb{N}} \subset \mathcal{B}_N$ has at least a $C^N$-limit $f \in C^N$ with $\|f_n - f\|_{C^N} \to 0$ for $n \to \infty$. It remains to show $f \in B_N$ and $f_n \to f$ in $\mathcal{B}_N$.

For this, let $\varepsilon > 0$ be arbitrary. Then there exists $M \in \mathbb{N}$ such that $\|f_n - f_m\|_N \leq \varepsilon$ for all $n, m \geq M$. From $\|f_n - f\|_{C^N} \to 0$ we deduce $\partial^\gamma f_n(x) \to \partial^\gamma f(x)$ pointwise for all $x \in \mathbb{R}^2$ and $|\gamma| \leq N$. By continuity, we deduce for $n \geq M$

$$(1 + |x|)^N \sum_{|\gamma| \leq N} |\partial^\gamma (f - f_n)(x)| \leq \varepsilon \quad \text{for all } x \in \mathbb{R}^2.$$

It follows $\|f - f_n\|_N \leq \varepsilon$ for all $n \geq M$, which implies $f \in B_N$ since in particular $\|f\|_N \leq \|f_M\|_N + \|f - f_M\|_N < \infty$. In addition, we can conclude $f_n \to f$ in $\mathcal{B}_N$ since $\varepsilon$ was arbitrary.

For $f \in \mathcal{B}_N$ we have $f(x) \lesssim (1 + |x|)^{-N}$. Hence, if $N \geq 2$, we clearly have $f \in L_2$. Further, $\mathcal{B}_N$ is dense in $L_2$ due to the density of the subspace $C_c^\infty$. Finally, to see that the embedding $\mathcal{B}_N \hookrightarrow L_2$ is continuous, we estimate

$$\|f|L_2\| \leq \|f(1 + |\cdot|)^N|L_\infty\| \|(1 + |\cdot|)^{-N}|L_2\| \lesssim \|f\|_N. \qquad \square$$

Next, we establish a strong form of the reconstruction formula (4.2) for Schwartz functions $\varphi \in \mathcal{S}(\mathbb{R}^2)$.

**Lemma 4.1.2.** *Let $\varphi \in \mathcal{S}(\mathbb{R}^2)$. Then, for every $N \in \mathbb{N}_0$ the reconstruction formula*

$$\varphi = \int_{\mathbb{X}} V_{\mathfrak{C}_\alpha}\varphi(\mathbf{x}) \psi_{\mathbf{x}} \, d\mu(\mathbf{x})$$

*holds in strong Bochner sense in $\mathcal{B}_N(\mathbb{R}^2)$.*

*Proof.* Let $N \in \mathbb{N}_0$. We first prove that the integral converges in Bochner sense in $\mathcal{B}_N$. For this we verify the Bochner criterion

$$\int_{\mathbb{X}} |V_{\mathfrak{C}_\alpha}\varphi(\mathbf{x})| \|\psi_{\mathbf{x}}\|_N \, d\mu(\mathbf{x}) < \infty.$$

The integrand of this integral is clearly measurable. It thus only remains to prove the boundedness. To this end, we first show that there is a constant $C_N > 0$ such that

$$\|\psi_{\mathbf{x}}\|_N \le C_N t^{-(1+\alpha)/2} t^{-N} (1+|x|_2)^N \quad \text{uniformly for all } \mathbf{x} = (x,\eta,t) \in \mathbb{X}. \tag{4.9}$$

Recall that with $g_t$ given as in (3.9)

$$\psi_{x,\eta,t} = t^{-(1+\alpha)/2} g_t\big(A_{\alpha,t}^{-1} R_\eta(\cdot - x)\big),$$

which implies

$$\hat{\psi}_{x,\eta,t} = t^{(1+\alpha)/2} \hat{g}_t(A_{\alpha,t} R_\eta \,\cdot\,) \exp(-2\pi i \langle x, \cdot\rangle).$$

Let us now estimate the Schwartz semi-norms $|\cdot|_{\kappa,\nu}$ from (4.4) for the $\alpha$-curvelets $\psi_{x,\eta,t} \in \mathfrak{C}_\alpha$. For $\kappa = (\kappa_1,\kappa_2) \in \mathbb{N}_0^2$ and $\nu = (\nu_1,\nu_2) \in \mathbb{N}_0^2$, with $|\kappa|_1, |\nu|_1 \le N$, we have

$$|\psi_{x,\eta,t}|_{\kappa,\nu} = \sup_{\xi \in \mathbb{R}^2} |\xi^\kappa \partial^\nu \psi_{x,\eta,t}(\xi)| \lesssim \|\partial^\kappa(\xi^\nu \hat{\psi}_{x,\eta,t}) | L_1\|.$$

Further

$$\begin{aligned}
\Big\|\partial^\kappa\{\xi \mapsto \xi^\nu \hat{\psi}_{x,\eta,t}(\xi)\} | L_1\Big\| &\lesssim t^{(1+\alpha)/2} \Big\|\partial^\kappa\{\xi \mapsto \xi^\nu \hat{g}_t(A_{\alpha,t}R_\eta \xi)\exp(-2\pi i\langle x,\xi\rangle)\} | L_1\Big\| \\
&\lesssim t^{(1+\alpha)/2} \sup_{l \le \kappa} \Big\|\partial^l\{\xi \mapsto \xi^\nu \hat{g}_t(A_{\alpha,t}R_\eta\xi)\} | L_1\Big\| |x_1|^{\kappa_1 - l_1} |x_2|^{\kappa_2 - l_2} \\
&\le t^{(1+\alpha)/2} (1+|x|_2)^N \sup_{|l|_1 \le N} \Big\|\partial^l\{\xi \mapsto \xi^\nu \hat{g}_t(A_{\alpha,t}R_\eta\xi)\} | L_1\Big\|,
\end{aligned}$$

where we used $|x_1| \le |x|_2$ and $|x_2| \le |x|_2$ to obtain

$$|x_1|^{\kappa_1-l_1}|x_2|^{\kappa_2-l_2} \le (1+|x_1|)^{\kappa_1-l_1}(1+|x_2|)^{\kappa_2-l_2} \le (1+|x|_2)^{\kappa_1+\kappa_2} \le (1+|x|_2)^N.$$

Finally, for $l = (l_1,l_2) \in \mathbb{N}_0^2$ with $|l|_1 \le N$ we deduce

$$\Big\|\partial^l\{\xi^\nu \hat{g}_t(A_{\alpha,t}R_\eta \cdot)\} | L_1\Big\| \lesssim \sup_{m \le \min\{l,\nu\}} \Big\|\xi^{\nu-m}\partial^{l-m}\hat{g}_t(A_{\alpha,t}R_\eta\cdot) | L_1\Big\|.$$

Taking into account $t \in (0,1]$, we can estimate

$$\Big|\partial^{l-m}\hat{g}_t(A_{\alpha,t}R_\eta\cdot)\Big| \lesssim \sum_{|n|_1 \le N} \Big|(\partial^n \hat{g}_t)(A_{\alpha,t}R_\eta\cdot)\Big|.$$

Using $|\xi^{\nu-m}| \le (1+|\xi|_2)^N$, we then obtain for each $m = (m_1,m_2) \in \mathbb{N}_0^2$ with $m \le \min\{l,\nu\}$

$$\begin{aligned}
\Big\|\xi^{\nu-m}\partial^{l-m}\hat{g}_t(A_{\alpha,t}R_\eta\cdot) | L_1\Big\| &\lesssim \sum_{|n|_1 \le N} \Big\|(1+|\xi|_2)^N(\partial^n\hat{g}_t)(A_{\alpha,t}R_\eta\cdot) | L_1\Big\| \\
&= t^{-(1+\alpha)} \sum_{|n|_1 \le N} \Big\|(1+|R_\eta^{-1}A_{\alpha,t}^{-1}\xi|_2)^N(\partial^n\hat{g}_t)(\cdot) | L_1\Big\| \\
&\lesssim t^{-(1+\alpha)} t^{-N} \sup_{|n|_1 \le N} \Big\|(1+|\xi|_2)^N(\partial^n\hat{g}_t)(\cdot) | L_1\Big\|.
\end{aligned}$$

Putting everything together, this yields

$$|\psi_{x,\eta,t}|_{\kappa,\nu} \lesssim \|\partial^\kappa(\xi^\nu \hat{\psi}_{x,\eta,t})|L_1\| \lesssim t^{-(1+\alpha)/2} t^{-N}(1+|x|_2)^N \sup_{|n|_1\le N} \Big\|(1+|\xi|_2)^N(\partial^n \hat{g}_t)(\cdot)|L_1\Big\|.$$

Now we use the $\alpha$-molecule estimate (2.5) with $N_1 > N+2$ and obtain the desired result,

$$|\psi_{x,\eta,t}|_{\kappa,\nu} \lesssim t^{-(1+\alpha)/2}(1+|x|_2)^N t^{-N}\|\{\xi \mapsto (1+|\xi|_2)^N(1+|\xi|_2^2)^{-N_1/2}\}|L_1\|.$$

Finally, we note that uniformly in $(x,\eta,t) \in \mathbb{X}$

$$\|\psi_{x,\eta,t}\|_N \lesssim \sup_{|\kappa|_1,|\nu|_1\le N} |\psi_{x,\eta,t}|_{\kappa,\nu}.$$

This proves (4.9).

Next, we interpret $\varphi \in \mathcal{S}(\mathbb{R}^2)$ as a system of $\alpha$-molecules of order $(\infty,\infty,\infty,\infty)$ consisting of just one element with the phase space coordinates $(0,0,1) \in \mathbb{P}$. Then we obtain from Theorem 2.2.2 for arbitrary but fixed $\tilde{N} > 0$ the estimate

$$|V_{\mathfrak{C}_\alpha}\varphi(x,\eta,t)| = |\langle \varphi, \psi_{x,\eta,t}\rangle| \le C_{\tilde{N},\varphi} \cdot t^{\tilde{N}}(1+|x|^2)^{-\tilde{N}} \tag{4.10}$$

with a constant $C_{\tilde{N},\varphi} > 0$ independent of $(x,\eta,t)$.

Altogether, (4.9) and (4.10) prove the Bochner criterion. Hence the integral

$$\varphi_N := \int_{\mathbb{X}} V_{\mathfrak{C}_\alpha}\varphi(\mathbf{x})\psi_{\mathbf{x}}\, d\mu(\mathbf{x}) \tag{4.11}$$

converges in Bochner sense to a function $\varphi_N$ in $\mathcal{B}_N$. It remains to prove $\varphi = \varphi_N$.

Let us first assume $N \ge 2$. Then $\mathcal{B}_N \hookrightarrow L_2$ and the function $\varphi_N$ is also the (strong and weak) $L_2$-limit of this integral. This implies $\varphi_N = \varphi$ almost everywhere since the reconstruction formula (4.2) holds weakly in $L_2$. Moreover, since both, $\varphi_N \in \mathcal{B}_N$ and $\varphi \in \mathcal{S}$, are continuous, we even have pointwise equality. For the case $N \in \{0,1\}$, let us note that $\mathcal{B}_{N+1} \hookrightarrow \mathcal{B}_N$ for all $N \in \mathbb{N}_0$. Altogether, this establishes the reconstruction formula in strong sense in $\mathcal{B}_N$ for all $N \in \mathbb{N}_0$. □

With the previous result, we are now ready to give a proof of Lemma 4.1.3.

**Lemma 4.1.3.** *The continuous $\alpha$-curvelet frame $\mathfrak{C}_\alpha$ is total in $\mathcal{S}(\mathbb{R}^2)$.*

*Proof.* We have seen that the reconstruction formula (4.11) holds in a strong Bochner sense in $\mathcal{B}_N$ for functions in $\mathcal{S}$. Let $U \subset \mathcal{S}$ be an open neighborhood in $\mathcal{S}$ of some $\varphi \in \mathcal{S}$. Then $U$ is open in $\mathcal{B}_N$ for sufficiently large $N$. Since the formula (4.11) holds strongly in $\mathcal{B}_N$, we can deduce that $U \cap \operatorname{span}\mathfrak{C}_\alpha \neq \emptyset$. □

As a consequence of Lemma 4.1.3, the extension of $V_{\mathfrak{C}_\alpha}$ defined in (4.7) is injective on $\mathcal{S}'(\mathbb{R}^2)$. Hence, it is an invertible transform and the reconstruction of signals is possible. In fact, we have the following reconstruction formula for signals in $\mathcal{S}'(\mathbb{R}^2)$.

**Proposition 4.1.4.** *For signals $f \in \mathcal{S}'(\mathbb{R}^2)$ the reconstruction formula* (4.2) *holds $*$-weakly in $\mathcal{S}'(\mathbb{R}^2)$, i.e., for all $\varphi \in \mathcal{S}(\mathbb{R}^2)$*

$$\langle f, \varphi\rangle_{\mathcal{S}'\times\mathcal{S}} = \int_{\mathbb{X}} V_{\mathfrak{C}_\alpha} f(\mathbf{x})\langle \varphi, \psi_{\mathbf{x}}\rangle\, d\mu(\mathbf{x}). \tag{4.12}$$

*Proof.* We have already observed that $\{\|\cdot\|_N\}_{N\in\mathbb{N}_0}$ is a filtrating family of semi-norms, which generates the topology of $\mathcal{S}$. Hence, according to a fundamental result on continuous functionals on locally convex spaces (see e.g. [105, page 96]), for fixed $f \in \mathcal{S}'(\mathbb{R}^2)$ there exist corresponding $N_f \in \mathbb{N}_0$ and $C_f > 0$ such that

$$|\langle f, \varphi\rangle_{\mathcal{S}'\times\mathcal{S}}| \leq C_f \|\varphi\|_{N_f} = C_f \sup_{x\in\mathbb{R}^2} (1+|x|)^{N_f} \sum_{|\gamma|\leq N_f} |\partial^\gamma \varphi(x)| \quad \text{for all } \varphi \in \mathcal{S}.$$

Since $\mathcal{S}(\mathbb{R}^2) \hookrightarrow B_{N_f}(\mathbb{R}^2)$, we can thus extend $f$ to a functional $\tilde{f} \in B'_{N_f}(\mathbb{R}^2)$ by the Hahn-Banach extension theorem (see e.g. [112, Satz VIII.2.8]). Moreover, if $N \geq 2$, we have the embeddings

$$\mathcal{S}(\mathbb{R}^2) \hookrightarrow B_N(\mathbb{R}^2) \hookrightarrow L_2(\mathbb{R}^2) \quad \text{and} \quad L_2(\mathbb{R}^2)' \hookrightarrow B'_N(\mathbb{R}^2) \hookrightarrow \mathcal{S}'(\mathbb{R}^2).$$

Hence, using Lemma 4.1.2, we can argue as follows with $N = N_f$,

$$\begin{aligned}\langle f, \varphi\rangle_{\mathcal{S}'\times\mathcal{S}} &= \langle \tilde{f}, \varphi\rangle_{\mathcal{B}'_N\times\mathcal{B}_N} = \langle \tilde{f}, \int_{\mathbb{X}} V_{\mathfrak{C}_\alpha}\varphi(\mathbf{x})\psi_{\mathbf{x}}\, d\mu(\mathbf{x})\rangle_{\mathcal{B}'_N\times\mathcal{B}_N} \\ &= \int_{\mathbb{X}} V_{\mathfrak{C}_\alpha}\varphi(\mathbf{x})\langle \tilde{f}, \psi_{\mathbf{x}}\rangle_{\mathcal{B}'_N\times\mathcal{B}_N}\, d\mu(\mathbf{x}) = \int_{\mathbb{X}} V_{\mathfrak{C}_\alpha}\varphi(\mathbf{x})\langle f, \psi_{\mathbf{x}}\rangle_{\mathcal{S}'\times\mathcal{S}}\, d\mu(\mathbf{x})\,.\end{aligned}$$

This establishes the reconstruction formula (4.12). □

Finally, we also extend the reproducing formula (4.3) to all tempered distributions.

**Proposition 4.1.5.** *The reproducing formula* (4.3) *holds for tempered distributions, i.e., for all* $f \in \mathcal{S}'(\mathbb{R}^2)$

$$V_{\mathfrak{C}_\alpha} f(\mathbf{x}) = \int_{\mathbb{X}} \overline{\mathcal{G}[\mathfrak{C}_\alpha]}(\mathbf{x},\mathbf{y}) V_{\mathfrak{C}_\alpha} f(\mathbf{y})\, d\mu(\mathbf{y})\,, \quad \mathbf{x} \in \mathbb{X}.$$

*Proof.* By plugging in $\psi_{\mathbf{x}}$ for $\varphi$ in (4.12) we directly obtain

$$\begin{aligned}V_{\mathfrak{C}_\alpha} f(\mathbf{x}) = \langle f, \psi_{\mathbf{x}}\rangle_{\mathcal{S}'\times\mathcal{S}} &= \int_{\mathbb{X}} V_{\mathfrak{C}_\alpha} f(\mathbf{y}) \langle \psi_{\mathbf{x}}, \psi_{\mathbf{y}}\rangle\, d\mu(\mathbf{y}) \\ &= \int_{\mathbb{X}} \overline{\mathcal{G}[\mathfrak{C}_\alpha]}(\mathbf{x},\mathbf{y}) V_{\mathfrak{C}_\alpha} f(\mathbf{y})\, d\mu(\mathbf{y}) = \overline{\mathcal{G}[\mathfrak{C}_\alpha]} V_{\mathfrak{C}_\alpha} f(\mathbf{x}),\end{aligned}$$

pointwise for all $\mathbf{x} \in \mathbb{X}$. □

Note that in contrast to (4.3) we need to use the conjugate reproducing kernel $\overline{\mathcal{G}[\mathfrak{C}_\alpha]}$. This is a consequence of the relation between the extended version of the $\alpha$-curvelet transform (4.7) and the $L_2$-version (4.1), as depicted in (4.8).

## 4.2 QBF-Spaces on the Curvelet Domain

The continuous $\alpha$-curvelet transform $V_{\mathfrak{C}_\alpha}$ defined in Section 4.1 is a powerful tool for signal analysis. In the sequel, we will use its extended version (4.7) for the characterization of signals $f \in \mathcal{S}'(\mathbb{R}^2)$ on the transform domain $\mathbb{X}$.

Concretely, our objects of interest will be so-called *$\alpha$-curvelet coorbit spaces*

$$\mathsf{Co}(\mathfrak{C}_\alpha, Y) := \Big\{ f \in \mathcal{S}'(\mathbb{R}^2) \ : \ V_{\mathfrak{C}_\alpha} f \in Y \Big\}, \tag{4.13}$$

where $Y$ is some suitable function space on $\mathbb{X}$. Our investigation of such spaces will be based on the theory presented in [74]. One required assumption there is that $Y$ constitutes a rich solid quasi-Banach function space, a notion recalled below.

A *quasi-Banach function space*, for which we subsequently use the abbreviation *QBF-space*, with associated domain $\mathbb{X}$ is by definition a subset of the $\mu$-measurable functions from $\mathbb{X}$ to $\mathbb{C}$, which is linearly closed and complete with respect to some given quasi-norm. Functions which coincide apart from a null-set are thereby identified. In the Banach case, we speak of a *Banach function space*, or *BF-space* for short. Note, that QBF-spaces in our sense need not be continuously embedded into $L_1^{loc}(\mathbb{X})$ as sometimes required in the literature, for example in [43].

As a reminder, a *quasi-norm* on a linear space $Y$ is defined in the same way as a norm, with the only difference that the triangle inequality need not hold in a strict sense. It suffices if it is satisfied up to a multiplicative constant $C_Y \geq 1$ called the *quasi-norm constant*, i.e., if

$$\|f+g\| \leq C_Y \big(\|f\| + \|g\|\big) \quad \text{for all } f,\, g \in Y.$$

Another concept, closely related to a quasi-norm, is the notion of an *$r$-norm*, where $0 < r \leq 1$ and the usual triangle inequality is replaced by the $r$-triangle inequality

$$\|f+g\|^r \leq \|f\|^r + \|g\|^r \quad \text{for all } f,\, g \in Y.$$

It is straightforward to show that every $r$-norm on $Y$ constitutes a quasi-norm with associated quasi-norm constant $C_Y = 2^{1/r-1}$. Vice versa, while a quasi-norm need not be an $r$-norm itself, there at least always exists an equivalent $r$-norm generating the same topology with $r$ satisfying $C_Y = 2^{1/r-1}$. This is the statement of the Aoki-Rolewicz theorem [3, 101]. As a consequence, every quasi-normed space $Y$ with quasi-norm constant $C_Y$ can be regarded as an $r$-normed space, with $r = (\log_2(C_Y) + 1)^{-1}$ being the so-called *exponent* of $Y$.

A QBF-space $Y$ on $\mathbb{X}$ is called *solid* if for every $\mu$-measurable function $f : \mathbb{X} \to \mathbb{C}$ we have

$$|f| \leq g \text{ for some } g \in Y \Rightarrow f \in Y \text{ and } \|f|Y\| \leq \|g|Y\|.$$

It is called *rich* if it contains all characteristic functions $\mathcal{X}_U$ corresponding to compact subsets $U \subset \mathbb{X}$.

The following lemma draws a connection between the convergence of a sequence of functions in a solid quasi-normed function space $Y$ to the pointwise convergence of subsequences.

**Lemma 4.2.1** ([74, Lem. 2.2]). *Let $Y$ be a solid quasi-normed function space on $\mathbb{X}$, and assume that $f_n \to f$ in $Y$ for a sequence of functions $(f_n)_{n\in\mathbb{N}}$. Then for almost all $\mathbf{x} \in \mathbb{X}$ there is a subsequence $(f_{n_k})_{k\in\mathbb{N}}$, whose choice may depend on the particular $\mathbf{x} \in \mathbb{X}$, such that $f_{n_k}(\mathbf{x}) \to f(\mathbf{x})$.*

*Proof.* Let us assume that $f_n \to f$ in $Y$. Then $g_n := f_n - f \to 0$ in $Y$ or equivalently $\|g_n|Y\| \to 0$. Since $\inf_{m\geq n} |g_m|$ is a measurable function with $\inf_{m\geq n} |g_m| \leq |g_k|$ for all

$k \geq n$ the solidity of $Y$ yields $\inf_{m\geq n} |g_m| \in Y$ and $\| \inf_{m\geq n} |g_m||Y\| \leq \|g_k|Y\|$ for all $k \geq n$. We deduce

$$0 \leq \| \inf_{m\geq n} |g_m||Y\| \leq \inf_{m\geq n} \|g_m|Y\| = 0,$$

and as a consequence $\inf_{m\geq n} |g_m|(\mathbf{x}) = 0$ for almost every $\mathbf{x} \in \mathbb{X}$. For each of these $\mathbf{x}$, we can hence find a subsequence $(g_{n_k})_{k\in\mathbb{N}}$ such that $g_{n_k}(\mathbf{x}) \to 0$. This implies $f_{n_k}(\mathbf{x}) \to f(\mathbf{x})$ and yields the result. □

We will subsequently restrict our attention to a special scale of function spaces on $\mathbb{X}$ corresponding to Besov-type characterizations of the signals $f \in \mathcal{S}'(\mathbb{R}^2)$. These spaces can be viewed as straight-forward generalizations of the mixed-norm Lebesgue spaces used in [111] for the coorbit description of the classic inhomogeneous Besov spaces.

We remark, that also other scales of function spaces on $\mathbb{X}$ could be considered, for example spaces leading to Triebel-Lizorkin type characterizations. The bulk of the subsequent exposition remains the same, significant adaptions are only needed in Subsection 4.5.5.

### 4.2.1 The Mixed-norm Lebesgue Spaces $L^s_{p,q}(\mathbb{X})$

We now define a scale $L^s_{p,q}(\mathbb{X})$ of function spaces on $\mathbb{X}$ which is inspired by the mixed-norm Lebesgue spaces on the inhomogeneous wavelet domain $\mathbb{R}^2 \times [(0,1) \cup \{\infty\}]$, considered in [111].

**Definition 4.2.2.** Let $0 < p, q < \infty$ and $s \in \mathbb{R}$. We then define the function space

$$L^s_{p,q}(\mathbb{X}) := \Big\{ F : \mathbb{X} \to \mathbb{C} \quad \mu\text{-measurable} \;:\; \|F|L^s_{p,q}\| < \infty \Big\}$$

with respective quasi-norm

$$\|F|L^s_{p,q}\| := \Big( \int_0^{2\pi} \Big\| F(\cdot,\eta,1)|L_p \Big\|^q \,d\eta \Big)^{1/q} + \Big( \int_0^1 \int_0^{2\pi} t^{-sq} \Big\| F(\cdot,\eta,t)|L_p \Big\|^q \frac{d\eta\,dt}{t} \Big)^{1/q}.$$

As we will see in Proposition 4.2.4, this space constitutes a rich solid QBF-space on $\mathbb{X}$ with associated exponent $r := \min\{1, p, q\}$. For the proof of the completeness, we will use the fact that Fatou's lemma is valid in $L^s_{p,q}(\mathbb{X})$.

**Lemma 4.2.3.** *Let $(F_n)_{n\in\mathbb{N}}$ be a sequence of functions $F_n : \mathbb{X} \to [0,\infty)$ in $L^s_{p,q}(\mathbb{X})$ such that $\liminf_{n\to\infty} \|F_n|L^s_{p,q}\| < \infty$. Then $F := \liminf_{n\to\infty} F_n \in L^s_{p,q}(\mathbb{X})$ with $\|F|L^s_{p,q}\| \leq \liminf_{n\to\infty} \|F_n|L^s_{p,q}\|$.*

*Proof.* The assertion is a direct consequence of the classic lemma of Fatou for Lebesgue spaces [2, Lem. A1.20]. For the inhomogeneous part, we obtain

$$\begin{aligned} \Big( \int_0^{2\pi} \Big\| \liminf_{n\to\infty} F_n(\cdot,\eta,1)|L_p \Big\|^q \,d\eta \Big)^{1/q} &\leq \Big( \int_0^{2\pi} \liminf_{n\to\infty} \Big\| F_n(\cdot,\eta,1)|L_p \Big\|^q \,d\eta \Big)^{1/q} \\ &\leq \liminf_{n\to\infty} \Big( \int_0^{2\pi} \Big\| F_n(\cdot,\eta,1)|L_p \Big\|^q \,d\eta \Big)^{1/q}. \end{aligned}$$

The calculation for the homogeneous part is similar. □

Now we are ready to prove the following result.

**Proposition 4.2.4.** *The spaces $L^s_{p,q}(\mathbb{X})$ are rich solid QBF-spaces on $\mathbb{X}$ with associated exponent $r := \min\{1, p, q\}$.*

*Proof.* It is straightforward to verify that $L^s_{p,q}$ is a quasi-normed space on $\mathbb{X}$. A direct calculation with $\tilde{p} := p/r$, $\tilde{q} := q/r$, $\tilde{s} := sr$, namely

$$\|f + g|L^s_{p,q}\|^r = \||f + g|^r|L^{\tilde{s}}_{\tilde{p},\tilde{q}}\| \leq \||f|^r|L^{\tilde{s}}_{\tilde{p},\tilde{q}}\| + \||g|^r|L^{\tilde{s}}_{\tilde{p},\tilde{q}}\| = \|f|L^s_{p,q}\|^r + \|g|L^s_{p,q}\|^r,$$

further shows for $r := \min\{1, p, q\}$ and arbitrary $f, g \in L^s_{p,q}$

$$\|f + g|L^s_{p,q}\|^r \leq \|f|L^s_{p,q}\|^r + \|g|L^s_{p,q}\|^r.$$

As a consequence, $\| \cdot |L^s_{p,q}\|$ is also an $r$-norm. The solidity of $L^s_{p,q}$ is obvious. For the proof of the richness, we refer to Lemma 4.2.11.

At last, we prove the completeness of $L^s_{p,q}$ and consider a Cauchy sequence $(F_n)_{n\in\mathbb{N}}$. Without loss of generality, we may assume the property

$$\|F_{n+1} - F_n|L^s_{p,q}\| \leq 2^{-n/r} \quad \text{for all } n \in \mathbb{N}.$$

We then define the $\mu$-measurable functions

$$G_m := \sum_{n=1}^{m} |F_{n+1} - F_n|\,, \quad m \in \mathbb{N},$$

which are elements of $L^s_{p,q}$ due to the estimate

$$\|G_m|L^s_{p,q}\|^r \leq \sum_{n=1}^{m} \|F_{n+1} - F_n|L^s_{p,q}\|^r \leq \sum_{n=1}^{\infty} \|F_{n+1} - F_n|L^s_{p,q}\|^r \leq \sum_{n=1}^{\infty} 2^{-n} = 1.$$

Further, since the sequence $(G_m)_{m\in\mathbb{N}}$ is monotonically increasing, we obtain with Lemma 4.2.3

$$G := \lim_{m\to\infty} G_m = \liminf_{m\to\infty} G_m \in L^s_{p,q},$$

and further

$$\|G|L^s_{p,q}\| \leq \liminf_{m\to\infty} \|G_m|L^s_{p,q}\| \leq 1.$$

In particular, the function $G$ is finite almost everywhere and the sum $\sum_{n=1}^{\infty} |F_{n+1}(\mathbf{x}) - F_n(\mathbf{x})|$ converges for almost every $\mathbf{x} \in \mathbb{X}$. At those points, the sequence $(F_n(\mathbf{x}))_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{C}$ and the limit

$$F(\mathbf{x}) := \lim_{n\to\infty} F_n(\mathbf{x})$$

is well-defined, giving rise to a $\mu$-measurable function $F$ on $\mathbb{X}$.

Using Lemma 4.2.3 and the fact that $(F_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^s_{p,q}$, we deduce

$$\|F - F_n|L^s_{p,q}\| = \| \lim_{m\to\infty} |F_m - F_n||L^s_{p,q}\| \leq \liminf_{m\to\infty} \|F_m - F_n|L^s_{p,q}\| \to 0 \quad (n \to \infty).$$

This implies $F \in L^s_{p,q}$ and the convergence $F_n \to F$ in $L^s_{p,q}$. □

In the next subsection, we introduce another scale of function spaces on $\mathbb{X}$ closely related to $L^s_{p,q}(\mathbb{X})$. Those spaces, denoted by $\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$, feature more regularity. For instance, in contrast to $L^s_{p,q}(\mathbb{X})$, they are continuously embedded into $L^{loc}_1(\mathbb{X})$, even in the quasi-Banach case.

### 4.2.2 The Associated Wiener Spaces $\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$

The Wiener spaces $\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$ are obtained by a Wiener-type amalgamization (see [43, Def. 3.1]) of $L^s_{p,q}(\mathbb{X})$ with the local component $L_\infty(\mathbb{X})$. This means that we utilize a suitable family of window functions $\{W_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ to take the $L_\infty$-norm locally around points $\mathbf{x}\in\mathbb{X}$. Afterwards, for a measurable function $F:\mathbb{X}\to\mathbb{C}$ the so-called *control function*

$$\mathbf{x}\mapsto \|\mathcal{X}_{W_{\mathbf{x}}}F|L_\infty\|, \tag{4.14}$$

is measured globally in the $L^s_{p,q}$-quasi-norm. The outcome of this procedure clearly depends on the utilized windows and we need to carefully choose those.

On the wavelet domain, which naturally carries a group structure, a canonical way to generate suitable windows is by the action of the group on some fixed prototype window. Since we do not have a group structure on $\mathbb{X}$, we need to take a different route and resort to the quasi-metric structure of $\mathbb{X}$ instead. Hence, we use the $\alpha$-balls $B^\alpha_\tau(\mathbf{x})$, or – to the same effect – their more practical relatives $V^\alpha_\tau(\mathbf{x})$, defined in (2.25) and (2.26), to localize the functions around points $\mathbf{x}\in\mathbb{X}$.

By appropriately restricting the sets in the collection $\mathcal{V}^\alpha_\tau[\mathbb{X}] := \{V^\alpha_\tau(\mathbf{x}) : \mathbf{x}\in\mathbb{X}\}$, we obtain for each $\alpha\in[0,1]$ and $\tau\ge 0$ the family

$$\mathcal{U}^\alpha_\tau[\mathbb{X}] := \Big\{U^\alpha_\tau(\mathbf{x}) : \mathbf{x}\in\mathbb{X}\Big\} \quad\text{with}\quad U^\alpha_\tau(\mathbf{x}) := \begin{cases} V^\alpha_\tau(\mathbf{x})\cap\mathbb{X}_0 & ,\ \mathbf{x}\in\mathbb{X}_0,\\ V^\alpha_\tau(\mathbf{x})\cap\mathbb{X}_1 & ,\ \mathbf{x}\in\mathbb{X}_1.\end{cases}$$

For convenience, let us state the explicit form of these sets,

$$U^\alpha_\tau(\mathbf{x}) = \begin{cases} (x+R_\eta^{-1}A_{\alpha,t}Q^\tau)\times(\eta+t^{1-\alpha}I^\tau)_{2\pi}\times(tJ^\tau\cap(0,1)) & ,\ \mathbf{x}\in\mathbb{X}_0\,,\\ (x+R_\eta^{-1}Q^\tau)\times(\eta+I^\tau)_{2\pi}\times(\{1\}) & ,\ \mathbf{x}\in\mathbb{X}_1\,.\end{cases}$$

where $Q^\tau := [-\tau,\tau]^2$, $I^\tau := [-\tau,\tau]$, and $J^\tau := [2^{-\tau},2^\tau]$. Utilizing

$$Q^{\alpha,\tau}_{x,\eta,t} := x+R_\eta^{-1}A_{\alpha,t}Q^\tau\,,\quad I^{\alpha,\tau}_{\eta,t} := (\eta+t^{1-\alpha}I^\tau)_{2\pi}\,,\quad J^\tau_t := \begin{cases} tJ^\tau\cap(0,1) & ,\ 0<t<1,\\ \{1\} & ,\ t=1,\end{cases}$$

we can also write

$$U^\alpha_\tau(\mathbf{x}) = Q^{\alpha,\tau}_{x,\eta,t}\times I^{\alpha,\tau}_{\eta,t}\times J^\tau_t\,. \tag{4.15}$$

The corresponding dual sets $U'^{,\alpha}_\tau(\mathbf{x})$ are given by

$$U'^{,\alpha}_\tau(\mathbf{x}) := \Big\{\mathbf{y}\in\mathbb{X}\ :\ \mathbf{x}\in U^\alpha_\tau(\mathbf{y})\Big\} = \begin{cases} V'^{,\alpha}_\tau(\mathbf{x})\cap\mathbb{X}_0 & ,\ \mathbf{x}\in\mathbb{X}_0,\\ V'^{,\alpha}_\tau(\mathbf{x})\cap\mathbb{X}_1 & ,\ \mathbf{x}\in\mathbb{X}_1,\end{cases}$$

with

$$U'^{,\alpha}_\tau(\mathbf{x}) = \begin{cases} \{\mathbf{y}\in\mathbb{X}_0\ :\ y\in x+R_\theta^{-1}A_{\alpha,u}Q^\tau,\ \theta\in(\eta+u^{1-\alpha}I^\tau)_{2\pi},\ u\in tJ^\tau\} & ,\ \mathbf{x}\in\mathbb{X}_0\,,\\ \{\mathbf{y}\in\mathbb{X}_1\ :\ y\in x+R_\theta^{-1}Q^\tau,\ \theta\in(\eta+I^\tau)_{2\pi}\} & ,\ \mathbf{x}\in\mathbb{X}_1\,,\end{cases}$$

or alternatively

$$U'^{,\alpha}_\tau(\mathbf{x}) = \Big\{\mathbf{y}\in\mathbb{X}\ :\ y\in Q^{\alpha,\tau}_{x,\theta,u},\ \theta\in I^{\alpha,\tau}_{\eta,u},\ u\in J^\tau_t\Big\}. \tag{4.16}$$

Some elementary but essential properties of the sets $U^\alpha_\tau(\mathbf{x})$ carry over from Lemma 2.2.17. It is worth noting that the results are now uniform for all $\mathbf{x}\in\mathbb{X}$. Moreover, they are independent of $\alpha\in[0,1]$.

**Lemma 4.2.5.** *Let $\alpha \in [0,1]$ and $\mathbf{x} = (x,\eta,t) \in \mathbb{X}$ be fixed. For $\tau \geq 0$ define the function $m(\tau) := 2^\tau(1+\tau)$. For $\tau,\sigma \geq 0$ the following holds true:*

*i) If $\tau < \sigma$ then $U^\alpha_\tau(\mathbf{x}) \subset U^\alpha_\sigma(\mathbf{x})$, and $U^\alpha_0(\mathbf{x}) = \bigcap_{\tau>0} U^\alpha_\tau(\mathbf{x}) = \{\mathbf{x}\}$.*

*ii) $\mathbf{y} \in U^\alpha_\tau(\mathbf{x}) \Rightarrow \mathbf{x} \in U^\alpha_{\tau m(\tau)}(\mathbf{y})$.*

*iii) $\mathbf{y} \in U^\alpha_\tau(\mathbf{x})$ and $\mathbf{z} \in U^\alpha_\sigma(\mathbf{y}) \Rightarrow \mathbf{z} \in U^\alpha_{f(\tau,\sigma)}(\mathbf{x})$ with $f(\tau,\sigma) := \tau + \sigma m(\tau)$.*

*iv) $\mathbf{y} \in U^\alpha_\tau(\mathbf{x})$ and $\mathbf{z} \in U^\alpha_\sigma(\mathbf{x}) \Rightarrow \mathbf{z} \in U^\alpha_{g(\tau,\sigma)}(\mathbf{y})$ with $g(\tau,\sigma) := (\tau+\sigma)m(\tau)$.*

*v) $\mathbf{x} \in U^\alpha_\tau(\mathbf{y}) \cap U^\alpha_\sigma(\mathbf{z}) \Rightarrow \mathbf{z} \in U^\alpha_{h(\tau,\sigma)}(\mathbf{y})$ with $h(\tau,\sigma) = \tau + \sigma m(\tau)m(\sigma)$.*

*Proof.* We distinguish two cases, either $\mathbf{x} \in \mathbb{X}_0$ or $\mathbf{x} \in \mathbb{X}_1$. If $\mathbf{x} \in \mathbb{X}_0$ then $U^\alpha_\tau(\mathbf{x}) = V^\alpha_\tau(\mathbf{x}) \cap \mathbb{X}_0$ and consequently $\mathbf{y},\mathbf{z} \in \mathbb{X}_0$. Analogously, $\mathbf{x} \in \mathbb{X}_1$ implies $U^\alpha_\tau(\mathbf{x}) = V^\alpha_\tau(\mathbf{x}) \cap \mathbb{X}_1$ and thus $\mathbf{y},\mathbf{z} \in \mathbb{X}_1$. Since $m_t(\tau) = m(\tau)$ for $t \leq 1$, the assertions follow then from Lemma 2.2.17. $\square$

Now we are ready to define the *$\alpha$-anisotropic Wiener maximal operator* $\mathbf{W}^\alpha_\tau$ depending on $\alpha \in [0,1]$ and $\tau > 0$. For a function $F : \mathbb{X} \to \mathbb{C}$ we put

$$\mathbf{W}^\alpha_\tau F(\mathbf{x}) := \|F\mathcal{X}_{U^\alpha_\tau(\mathbf{x})}|L_\infty\| = \operatorname*{ess\,sup}_{\mathbf{y}\in U^\alpha_\tau(\mathbf{x})} |F(\mathbf{y})|\,, \quad \mathbf{x} \in \mathbb{X}. \tag{4.17}$$

Further, if $\tau = 1$ we use the simplified notation $\mathbf{W}^\alpha := \mathbf{W}^\alpha_1$.

The term *maximal operator* is justified, since $\mathbf{W}^\alpha_\tau$ has the following majorizing property,

$$|F(\mathbf{x})| \leq \mathbf{W}^\alpha_\tau F(\mathbf{x}) \quad \text{for a.e. } \mathbf{x} \in \mathbb{X}. \tag{4.18}$$

But an even stronger result holds true, stated in the following lemma.

**Lemma 4.2.6.** *Let $\alpha \in [0,1]$ and let $\tau > \sigma > 0$. Then for any function $F : \mathbb{X} \to \mathbb{C}$ we have for almost every $\mathbf{x} \in \mathbb{X}$*

$$|F(\mathbf{x})|\mathcal{X}_{U^\alpha_\sigma(\mathbf{y})}(\mathbf{x}) = |F(\mathbf{x})|\mathcal{X}_{U'^{,\alpha}_\sigma(\mathbf{x})}(\mathbf{y}) \leq \mathbf{W}^\alpha_\tau F(\mathbf{y}) \quad \textit{for all } \mathbf{y} \in \mathbb{X}.$$

*Proof.* In a first step we prove (4.18). For this choose $\rho := \min\{1, \tau/8\} > 0$ such that $g(\rho,\rho) \leq \tau$, where $g$ is the function from Lemma 4.2.5(iv). According to Lemma 4.2.5(iv), we then have

$$\bigcap_{\mathbf{y}\in U^\alpha_\rho(\mathbf{x})} U^\alpha_\tau(\mathbf{y}) \supseteq U^\alpha_\rho(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathbb{X}.$$

It follows that for each $\mathbf{x} \in \mathbb{X}$ the following relation holds true, for almost all $\mathbf{y} \in U^\alpha_\rho(\mathbf{x})$,

$$|F(\mathbf{y})| \leq \operatorname*{ess\,sup}_{\mathbf{z}\in U^\alpha_\rho(\mathbf{x})} |F(\mathbf{z})| \leq \operatorname*{ess\,sup}_{\mathbf{z}\in U^\alpha_\tau(\mathbf{y})} |F(\mathbf{z})| = \mathbf{W}^\alpha_\tau F(\mathbf{y}).$$

The relation (4.18) is a direct consequence, since we can find a sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ such that $\{U^\alpha_\rho(\mathbf{x}_n)\}_{n\in\mathbb{N}}$ is a countable covering of $\mathbb{X}$.

Secondly, we now turn to the proof of the more general assertion of the lemma. According to (4.18), for each $n \in \mathbb{N}$ there exists a null-set $\mathbf{N}_n \subset \mathbb{X}$ such that $|F(\mathbf{x})| \leq \mathbf{W}^\alpha_{1/n}F(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{X}\backslash\mathbf{N}_n$. Putting $\mathbf{N} := \bigcup_n \mathbf{N}_n$, which is again a null-set, we then have $|F(\mathbf{x})| \leq \mathbf{W}^\alpha_\tau F(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{X}\backslash\mathbf{N}$ and all $\tau > 0$. Since by assumption $\sigma < \tau$, we can choose

$\rho := (\tau - \sigma)/m(\sigma) > 0$ such that $f(\sigma, \rho) = \tau$ for the function $f$ from Lemma 4.2.5(iii). Then for every $\mathbf{x} \in \mathbb{X} \backslash \mathbf{N}$

$$|F(\mathbf{x})| \leq \mathbf{W}^{\alpha}_{\rho} F(\mathbf{x}) \leq \mathbf{W}^{\alpha}_{\tau} F(\mathbf{y}) \quad \text{for all } \mathbf{y} \in U'^{,\alpha}_{\sigma}(\mathbf{x}),$$

where $U'^{,\alpha}_{\sigma}(\mathbf{x})$ denotes the dual ball of $U^{\alpha}_{\sigma}(\mathbf{x})$. Since $\mathcal{X}_{U^{\alpha}_{\sigma}(\mathbf{y})}(\mathbf{x}) = \mathcal{X}_{U'^{,\alpha}_{\sigma}(\mathbf{x})}(\mathbf{y})$, the proof is finished. □

Utilizing the maximal operator $\mathbf{W}^{\alpha} = \mathbf{W}^{\alpha}_{1}$, we can now define the spaces $\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$. Note that in contrast to the original scale $L^{s}_{p,q}(\mathbb{X})$ the new scale $\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$ also depends on the parameter $\alpha \in [0, 1]$.

**Definition 4.2.7.** Let $\alpha \in [0, 1]$, $0 < p, q < \infty$, and $s \in \mathbb{R}$. We define the *$\alpha$-anisotropic Wiener space* associated to $L^{s}_{p,q}(\mathbb{X})$ as

$$\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X}) := \Big\{ F : \mathbb{X} \to \mathbb{C} \quad \mu\text{-measurable} \ : \ \|\mathbf{W}^{\alpha} F | L^{s}_{p,q}\| < \infty \Big\}$$

with quasi-norm $\| \cdot | \mathbb{L}^{\alpha,s}_{p,q}\| := \|\mathbf{W}^{\alpha}(\cdot) | L^{s}_{p,q}\|$.

The spaces $\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$ inherit many properties from the scale $L^{s}_{p,q}(\mathbb{X})$. In particular, like those, they are rich solid QBF-spaces on $\mathbb{X}$ with the same exponent $r = \min\{1, p, q\}$. For the proof of this fact we need the following commutation property of $\mathbf{W}^{\alpha}_{\tau}$.

**Lemma 4.2.8.** *Let $\alpha \in [0, 1]$ and $\tau > 0$. Assume that $F_n \to F$ converges pointwise almost everywhere for $n \to \infty$, in a uniform way on compacta. Then*

$$\mathbf{W}^{\alpha}_{\tau} F(\mathbf{x}) = \lim_{n \to \infty} \mathbf{W}^{\alpha}_{\tau} F_n(\mathbf{x}) \quad \textit{for every } \mathbf{x} \in \mathbb{X}.$$

*Proof.* First, we show that for any two functions $F, G : \mathbb{X} \to \mathbb{C}$ and all $\mathbf{x} \in \mathbb{X}$

$$|\mathbf{W}^{\alpha}_{\tau} G(\mathbf{x}) - \mathbf{W}^{\alpha}_{\tau} F(\mathbf{x})| \leq \mathbf{W}^{\alpha}_{\tau}(G - F)(\mathbf{x}). \tag{4.19}$$

Without loss of generality, we can assume that $\mathbf{W}^{\alpha}_{\tau} G(\mathbf{x}) \geq \mathbf{W}^{\alpha}_{\tau} F(\mathbf{x})$. Further, there exists a null-set $\mathbf{N} \subset \mathbb{X}$ such that $|F(\mathbf{y})| \leq \operatorname{ess\,sup}_{\mathbf{z} \in U^{\alpha}_{\tau}(\mathbf{x})} |F(\mathbf{z})|$ for all $\mathbf{y} \in U^{\alpha}_{\tau}(\mathbf{x}) \backslash \mathbf{N}$. Then (4.19) follows from the estimate

$$\begin{aligned}
\operatorname*{ess\,sup}_{\mathbf{y} \in U^{\alpha}_{\tau}(\mathbf{x})} |G(\mathbf{y})| - \operatorname*{ess\,sup}_{\mathbf{z} \in U^{\alpha}_{\tau}(\mathbf{x})} |F(\mathbf{z})| &= \operatorname*{ess\,sup}_{\mathbf{y} \in U^{\alpha}_{\tau}(\mathbf{x}) \backslash \mathbf{N}} \Big( |G(\mathbf{y})| - \operatorname*{ess\,sup}_{\mathbf{z} \in U^{\alpha}_{\tau}(\mathbf{x})} |F(\mathbf{z})| \Big) \\
&\leq \operatorname*{ess\,sup}_{\mathbf{y} \in U^{\alpha}_{\tau}(\mathbf{x}) \backslash \mathbf{N}} \Big( |G(\mathbf{y})| - |F(\mathbf{y})| \Big) \leq \operatorname*{ess\,sup}_{\mathbf{y} \in U^{\alpha}_{\tau}(\mathbf{x})} |G(\mathbf{y}) - F(\mathbf{y})|.
\end{aligned}$$

The assertion of the lemma is now a direct consequence of the validity of

$$|\mathbf{W}^{\alpha}_{\tau} F_n(\mathbf{x}) - \mathbf{W}^{\alpha}_{\tau} F(\mathbf{x})| \leq \mathbf{W}^{\alpha}_{\tau}(F_n - F)(\mathbf{x}) = \operatorname*{ess\,sup}_{\mathbf{y} \in U^{\alpha}_{\tau}(\mathbf{x})} |F_n(\mathbf{y}) - F(\mathbf{y})|$$

for every $\mathbf{x} \in \mathbb{X}$ and the fact that

$$\operatorname*{ess\,sup}_{\mathbf{y} \in U^{\alpha}_{\tau}(\mathbf{x})} |F_n(\mathbf{y}) - F(\mathbf{y})| \to 0 \quad (n \to \infty).$$

□

Now we formulate the companion result to Proposition 4.2.4.

**Proposition 4.2.9.** *The spaces $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$ are rich solid QBF-spaces on $\mathbb{X}$ with associated exponent $r := \min\{1, p, q\}$.*

*Proof.* It is clear that $\mathbb{L}_{p,q}^{\alpha,s}$ is a quasi-normed space on $\mathbb{X}$ with the same exponent $r = \min\{1, p, q\}$ as $L_{p,q}^s$. Further, the solidity of $\mathbb{L}_{p,q}^{\alpha,s}$ directly follows from the solidity of $L_{p,q}^s$ and the monotonicity of the Wiener maximal operator. To verify the richness, we use Lemma 4.2.5(v) to estimate

$$\mathbf{W}^\alpha \mathcal{X}_{U_\sigma^\alpha(\mathbf{x})} \le \mathcal{X}_{U_{h(\sigma,1)}^\alpha(\mathbf{x})} \quad \text{for every } \mathbf{x} \in \mathbb{X} \text{ and all } \sigma > 0. \tag{4.20}$$

Taking into account the richness and the solidity of $L_{p,q}^s$, this implies $\mathbf{W}^\alpha \mathcal{X}_{U_\sigma^\alpha(\mathbf{x})} \in L_{p,q}^s$ and thus $\mathcal{X}_{U_\sigma^\alpha(\mathbf{x})} \in \mathbb{L}_{p,q}^{\alpha,s}$.

It remains to show the completeness of $\mathbb{L}_{p,q}^{\alpha,s}$. For this, we use an embedding which will be established in the next subsection. According to (4.26), we have $\mathbb{L}_{p,q}^{\alpha,s} \hookrightarrow L_\infty^{loc}$. Due to this result, every Cauchy sequence $(F_n)_{n\in\mathbb{N}}$ in $\mathbb{L}_{p,q}^{\alpha,s}$ yields Cauchy sequences $(F_n(\mathbf{x}))_{n\in\mathbb{N}}$ in $\mathbb{C}$ for almost every $\mathbf{x} \in \mathbb{X}$.

Hence, $(F_n)_{n\in\mathbb{N}}$ converges pointwise almost everywhere to a $\mu$-measurable function $F$ on $\mathbb{X}$. In addition, the convergence is uniform on compacta. Hence, we can apply Lemma 4.2.8 which yields for fixed $m \in \mathbb{N}$ the pointwise convergence

$$\mathbf{W}_\tau^\alpha(F_m - F) = \lim_{n\to\infty} \mathbf{W}_\tau^\alpha(F_m - F_n).$$

Since $(F_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{L}_{p,q}^{\alpha,s}$, we can further verify for every $m \in \mathbb{N}$

$$\liminf_{n\to\infty} \|\mathbf{W}_\tau^\alpha(F_m - F_n)|L_{p,q}^s\| = \liminf_{n\to\infty} \|F_m - F_n|\mathbb{L}_{p,q}^{\alpha,s}\| < \infty.$$

Hence, using the Fatou property of $L_{p,q}^s$ proved in Lemma 4.2.3, we get $\mathbf{W}_\tau^\alpha(F_m - F) \in L_{p,q}^s$ and thus $F_m - F \in \mathbb{L}_{p,q}^{\alpha,s}$ for all $m \in \mathbb{N}$. In particular, since $F_m \in \mathbb{L}_{p,q}^{\alpha,s}$, this implies $F \in \mathbb{L}_{p,q}^{\alpha,s}$. Further, we have $F_m \to F$ in $\mathbb{L}_{p,q}^{\alpha,s}$ since

$$\|F_m - F|\mathbb{L}_{p,q}^{\alpha,s}\| \le \liminf_{n\to\infty} \|\mathbf{W}_\tau^\alpha(F_m - F_n)|L_{p,q}^s\| = \liminf_{n\to\infty} \|F_m - F_n|\mathbb{L}_{p,q}^{\alpha,s}\| \to 0 \quad (m \to \infty).$$

□

We end this subsection with an important embedding, relating the spaces $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$ and $L_{p,q}^s(\mathbb{X})$. It is a direct consequence of (4.18).

**Proposition 4.2.10.** *We have the continuous embedding*

$$\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X}) \hookrightarrow L_{p,q}^s(\mathbb{X}).$$

*Proof.* As a consequence of (4.18) and the solidity of $L_{p,q}^s$, we obtain

$$\|F|L_{p,q}^s\| \le \|\mathbf{W}^\alpha F|L_{p,q}^s\| = \|F|\mathbb{L}_{p,q}^{\alpha,s}\|$$

for every measurable function $F : \mathbb{X} \to \mathbb{C}$. This yields the result. □

Another important embedding of $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$ is derived in the next subsection. Proposition 4.2.13 will show that these Wiener spaces are naturally embedded into certain weighted $L_\infty$-spaces.

### 4.2.3 The Associated Canonical Weights $\mathbf{v}_{p,q}^{\alpha,s}$

We have already mentioned at the end of Subsection 4.2.1 that the spaces $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$ are continuously embedded into $L_1^{loc}(\mathbb{X})$. We will see that this is a consequence of the local boundedness of the functions contained in $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$, a property which we will study in more detail in this subsection. As it turns out, to each space $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$ there belongs an *associated canonical weight* $\mathbf{v}_{p,q}^{\alpha,s}: \mathbb{X} \to \mathbb{R}_+$ such that

$$\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X}) \hookrightarrow L_\infty^{1/\mathbf{v}_{p,q}^{\alpha,s}}(\mathbb{X}).$$

To determine the weight $\mathbf{v}_{p,q}^{\alpha,s}$, let us introduce the characteristic functions

$$\mathcal{X}_{\mathbf{x}}^{\alpha,\tau}(\mathbf{y}) := \mathcal{X}_{U_\tau^\alpha(\mathbf{x})}(\mathbf{y}) \quad \text{and} \quad \tilde{\mathcal{X}}_{\mathbf{x}}^{\alpha,\tau}(\mathbf{y}) := \mathcal{X}_{U_\tau^{\prime,\alpha}(\mathbf{x})}(\mathbf{y}), \tag{4.21}$$

where $U_\tau^\alpha(\mathbf{x})$ is as in (4.15) and $U_\tau^{\prime,\alpha}(\mathbf{x})$ denotes the dual ball from (4.16). Note that

$$\mathcal{X}_{\mathbf{x}}^{\alpha,\tau}(\mathbf{y}) = \sup_{\mathbf{z}\in U_\tau^{\prime,\alpha}(\mathbf{y})} \mathcal{X}_{\{\mathbf{x}\}}(\mathbf{z}) \quad \text{and} \quad \tilde{\mathcal{X}}_{\mathbf{x}}^{\alpha,\tau}(\mathbf{y}) = \sup_{\mathbf{z}\in U_\tau^{\alpha}(\mathbf{y})} \mathcal{X}_{\{\mathbf{x}\}}(\mathbf{z}),$$

where $\mathcal{X}_{\{\mathbf{x}\}}$ is the unit-height spike at $\mathbf{x} \in \mathbb{X}$.

According to (4.15), the characteristic function $\mathcal{X}_{\mathbf{x}}^{\alpha,\tau} = \mathcal{X}_{x,\eta,t}^{\alpha,\tau}$ of the set $U_\tau^\alpha(\mathbf{x})$ can be decomposed in the form

$$\mathcal{X}_{x,\eta,t}^{\alpha,\tau}(y,\theta,u) = \chi_{x,\eta,t}^{\alpha,\tau}(y) \cdot \chi_{\eta,t}^{\alpha,\tau}(\theta) \cdot \chi_t^\tau(u),$$

where $(y,\theta,u) \in \mathbb{X}$ and

$$\chi_{x,\eta,t}^{\alpha,\tau}(y) := \mathcal{X}_{Q_{x,\eta,t}^{\alpha,\tau}}(y)\,, \quad \chi_{\eta,t}^{\alpha,\tau}(\theta) := \mathcal{X}_{I_{\eta,t}^{\alpha,\tau}}(\theta)\,, \quad \chi_t^\tau(u) := \mathcal{X}_{J_t^\tau}(u). \tag{4.22}$$

Due to the tent-like structure (4.16) of $U_\tau^{\prime,\alpha}(\mathbf{x})$, the decomposition of the characteristic functions $\tilde{\mathcal{X}}_{\mathbf{x}}^{\alpha,\tau} = \tilde{\mathcal{X}}_{x,\eta,t}^{\alpha,\tau}$ has a slightly different form, namely

$$\tilde{\mathcal{X}}_{x,\eta,t}^{\alpha,\tau}(y,\theta,u) = \chi_{x,\theta,u}^{\alpha,\tau}(y) \cdot \chi_{\eta,u}^{\alpha,\tau}(\theta) \cdot \chi_t^\tau(u) \quad \text{for } (y,\theta,u) \in \mathbb{X}.$$

We now calculate the $L_{p,q}^s$-quasi-norms of these characteristic functions, since they play an important role, as we will see below.

**Lemma 4.2.11.** *Let $\alpha \in [0,1]$ and let $\tau > 0$ be fixed. We then have*

$$\|\mathcal{X}_{x,\eta,t}^{\alpha,\tau}|L_{p,q}^s\| \asymp \|\tilde{\mathcal{X}}_{x,\eta,t}^{\alpha,\tau}|L_{p,q}^s\| \asymp t^{-s+(1+\alpha)/p+(1-\alpha)/q} \quad \textit{uniformly in } (x,\eta,t) \in \mathbb{X}.$$

*Proof.* For $(x,\eta,t) \in \mathbb{X}_0$, we first calculate

$$\int_{\mathbb{R}^2} \chi_{x,\eta,t}^{\alpha,\tau}(y)\,dy \asymp t^{1+\alpha}\,, \quad \int_{\mathbb{T}} \chi_{\eta,t}^{\alpha,\tau}(\theta)\,d\theta \asymp t^{1-\alpha}\,, \quad \int_0^1 u^{-sq}\chi_t^\tau(u)\,\frac{du}{u} \asymp t^{-sq}. \tag{4.23}$$

Using (4.23) and $\|\chi_{x,\eta,t}^{\alpha,\tau}|L_p\| = \|\chi_{x,\eta,t}^{\alpha,\tau}|L_1\|^{1/p}$, we obtain

$$\begin{aligned}
\|\mathcal{X}_{x,\eta,t}^{\alpha,\tau}|L_{p,q}^s\| &= \left( \int_0^1 \int_0^{2\pi} u^{-sq}\chi_t^\tau(u)\chi_{\eta,t}^{\alpha,\tau}(\theta)\|\chi_{x,\eta,t}^{\alpha,\tau}|L_p\|^q \frac{d\theta\,du}{u} \right)^{1/q} \\
&\asymp \left( \int_0^1 u^{-sq}\chi_t^\tau(u) t^{1-\alpha} t^{(1+\alpha)q/p} \frac{du}{u} \right)^{1/q} \asymp t^{-s} t^{(1+\alpha)/p} t^{(1-\alpha)/q}.
\end{aligned}$$

Further, we also have

$$\begin{aligned}\|\tilde{\mathcal{X}}^{\alpha,\tau}_{x,\eta,t}|L^s_{p,q}\| &= \left(\int_0^1\int_0^{2\pi} u^{-sq}\chi_t^\tau(u)\chi^{\alpha,\tau}_{\eta,u}(\theta)\|\chi^{\alpha,\tau}_{x,\theta,u}|L_p\|^q\,\frac{d\theta\,du}{u}\right)^{1/q}\\ &\asymp \left(\int_0^1 u^{-sq}\chi_t^\tau(u)u^{1-\alpha}u^{(1+\alpha)q/p}\frac{du}{u}\right)^{1/q} \asymp t^{-s}t^{(1+\alpha)/p}t^{(1-\alpha)/q}.\end{aligned}$$

An analogous calculation yields

$$\|\mathcal{X}^{\alpha,\tau}_{x,\eta,1}|L^s_{p,q}\| \asymp \|\tilde{\mathcal{X}}^{\alpha,\tau}_{x,\eta,1}|L^s_{p,q}\| \asymp 1$$

for $(x,\eta,1)\in\mathbb{X}_1$. □

As a byproduct of the previous lemma, we obtain the quasi-invariance of the $\mu$-measure of the sets $U^\alpha_\tau(\mathbf{x})$ and their duals $U'^{,\alpha}_\tau(\mathbf{x})$ from their positions $\mathbf{x}\in\mathbb{X}$.

**Corollary 4.2.12.** *Let $\alpha\in[0,1]$, and $\tau>0$ be fixed. Then we have*

$$\mu(U^\alpha_\tau(\mathbf{x})) \asymp \mu(U'^{,\alpha}_\tau(\mathbf{x})) \asymp 1 \quad \text{uniformly in } \mathbf{x}\in\mathbb{X}.$$

*Proof.* Using Lemma 4.2.11, we obtain

$$\mu(U^\alpha_\tau(\mathbf{x})) = \int_{\mathbb{X}} \mathcal{X}^{\alpha,\tau}_{\mathbf{x}}(\mathbf{y})\,d\mu(\mathbf{y}) = \|\mathcal{X}^{\alpha,\tau}_{\mathbf{x}}|L^2_{1,1}\| \asymp 1\,,$$

and similarly $\mu(U'^{,\alpha}_\tau(\mathbf{x}))\asymp 1$. □

Lemma 4.2.11 paves the way for the definition of the weight functions $\mathbf{v}^{\alpha,s}_{p,q}:\mathbb{X}\to\mathbb{R}_+$. Recall that our aim are embeddings of the form (4.25). Following intuition, we need to calculate $\|\mathcal{X}_{U^\alpha_\tau(\mathbf{x})}|\mathbb{L}^{\alpha,s}_{p,q}\| = \|\mathbf{W}^\alpha\mathcal{X}_{U^\alpha_\tau(\mathbf{x})}|L^s_{p,q}\|$ for small $\tau>0$ to approximate $\mathbf{v}^{\alpha,s}_{p,q}(\mathbf{x})$. This motivates to use $\|\mathcal{X}_{U'^{,\alpha}_1(\mathbf{x})}|L^s_{p,q}\| = \|\sup_{\mathbf{z}\in U^\alpha_1(\cdot)}\mathcal{X}_{\{\mathbf{x}\}}(\mathbf{z})|L^s_{p,q}\|$ for the definition. In view of Lemma 4.2.11, it thus makes sense to define

$$\mathbf{v}^{\alpha,s}_{p,q}(\mathbf{x}) := t^{-s+(1+\alpha)/p+(1-\alpha)/q}\,, \quad \mathbf{x}=(x,\eta,t)\in\mathbb{X}\,. \tag{4.24}$$

Indeed, we then have the following result.

**Proposition 4.2.13.** *We have the continuous embedding*

$$\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X}) \hookrightarrow L^{1/\mathbf{v}^{\alpha,s}_{p,q}}_\infty(\mathbb{X}). \tag{4.25}$$

*Proof.* Take $F\in\mathbb{L}^{\alpha,s}_{p,q}$. Then $F\in L^s_{p,q}$ and, due to Lemma 4.2.6 and the solidity of $L^s_{p,q}$, it holds for almost every $\mathbf{x}\in\mathbb{X}$

$$\|F|\mathbb{L}^{\alpha,s}_{p,q}\| = \|\mathbf{W}^\alpha F|L^s_{p,q}\| \ge |F(\mathbf{x})|\|\mathcal{X}_{U'^{,\alpha}_{\frac12}(\mathbf{x})}|L^s_{p,q}\| = |F(\mathbf{x})|\|\tilde{\mathcal{X}}^{\alpha,\frac12}_{\mathbf{x}}|L^s_{p,q}\|.$$

Using Lemma 4.2.11, this implies $F\in L^{1/\mathbf{v}^{\alpha,s}_{p,q}}_\infty$ and the embedding $\mathbb{L}^{\alpha,s}_{p,q}\hookrightarrow L^{1/\mathbf{v}^{\alpha,s}_{p,q}}_\infty$ is continuous. □

Since the weights $\mathbf{v}_{p,q}^{\alpha,s}$ are *locally bounded*, i.e., they constitute bounded functions on every compact subset of $\mathbb{X}$, we finally arrive at the following chain of continuous embeddings

$$\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X}) \hookrightarrow L_\infty^{1/\mathbf{v}_{p,q}^{\alpha,s}}(\mathbb{X}) \hookrightarrow L_\infty^{loc}(\mathbb{X}) \hookrightarrow L_1^{loc}(\mathbb{X}). \tag{4.26}$$

In particular, these embeddings show that the convergence in $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$ is locally uniform almost everywhere, a fact which was used in the proof of Proposition 4.2.9.

We are now well-prepared for the definition and analysis of the coorbit spaces $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})$ associated to the $\alpha$-curvelet transform $V_{\mathfrak{C}_\alpha}$ and the spaces $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$. This will be the topic of the remaining two sections.

## 4.3 $\alpha$-Molecule Coorbit Spaces

In the previous section, we have introduced two scales of function spaces on the curvelet domain $\mathbb{X}$, namely $L_{p,q}^{s}(\mathbb{X})$ (Definition 4.2.2) and $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$ (Definition 4.2.7). Plugging the spaces $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$ into the general definition (4.13) of an $\alpha$-curvelet coorbit space, we obtain the special scale of coorbits $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})$ which we will subsequently analyze in more detail. In the end, our analysis will reveal, via Theorem 4.3.13, that these spaces can be identified with other Besov-type function spaces, for example those considered in [85].

An important tool in our investigation will be the *continuous $\alpha$-molecule transform* which is a natural generalization of the continuous $\alpha$-curvelet transform $V_{\mathfrak{C}_\alpha}$ from Section 4.1. It is introduced in Subsection 4.3.2 and enables a much broader approach to the analysis of the coorbits $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})$. In particular, it leads to the more general notion of an *$\alpha$-molecule coorbit space* introduced and analyzed in Subsection 4.3.3.

### 4.3.1 $\alpha$-Curvelet Coorbit Spaces associated to $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$

Let us begin with the definition of the $\alpha$-curvelet coorbit spaces associated to the scale $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$. They are obtained for the special choice $Y := \mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$ in (4.13).

**Definition 4.3.1.** The *$\alpha$-curvelet coorbit space* with respect to $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$ is defined as

$$\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s}) := \Big\{ f \in \mathcal{S}'(\mathbb{R}^2) \ : \ V_{\mathfrak{C}_\alpha} f \in \mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X}) \Big\}$$

and equipped with the quasi-norm $\| \cdot | \mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})\| := \|V_{\mathfrak{C}_\alpha}(\cdot)|\mathbb{L}_{p,q}^{\alpha,s}\|$.

To analyze these spaces, we will utilize the abstract machinery provided by the theory of coorbit spaces – here especially the exposition from [74] – combined with the theory of $\alpha$-molecules from Chapter 2. In particular, in Theorem 4.3.8, we will see that they constitute quasi-Banach spaces.

For the subsequent investigation, it is advantageous to view $\alpha$-curvelets as special instances of $\alpha$-molecules. It turns out that without much effort we can then take a broader approach and base our entire investigation on the more general concept of *$\alpha$-molecule coorbit spaces*.

### 4.3.2 The Continuous $\alpha$-Molecule Transform

In this subsection, $\mathfrak{M}_\alpha = \{m_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ shall always denote a system of $\alpha$-molecules with respect to the canonical parametrization. If such a system constitutes a tight frame for $L_2(\mathbb{R}^2)$ it gives rise to an associated transform $V_{\mathfrak{M}_\alpha} : L_2(\mathbb{R}^2) \to L_2(\mathbb{X})$ defined by

$$V_{\mathfrak{M}_\alpha} f(\mathbf{x}) := \langle f, m_{\mathbf{x}} \rangle , \quad \mathbf{x} \in \mathbb{X}. \tag{4.27}$$

Let us subsequently call it the *continuous $\alpha$-molecule transform* associated to $\mathfrak{M}_\alpha$.

In analogy to the definition of the $\alpha$-curvelet coorbit spaces in Definition 4.3.1, we now aim to define coorbits of $\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$ with respect to this more general transform $V_{\mathfrak{M}_\alpha}$. Before we can do this, however, we need to suitably enlarge the reservoir of $V_{\mathfrak{M}_\alpha}$ similar to the extension of the $\alpha$-curvelet transform $V_{\mathfrak{C}_\alpha}$ to the space of tempered distributions $\mathcal{S}'(\mathbb{R}^2)$ in Subsection 4.1.2.

Here some more care is required, though. Whereas the inclusion $\mathfrak{C}_\alpha \subset \mathcal{S}(\mathbb{R}^2)$ allowed to use the whole space $\mathcal{S}'(\mathbb{R}^2)$ as a reservoir for $V_{\mathfrak{C}_\alpha}$, we may not be so lucky with the transform $V_{\mathfrak{M}_\alpha}$ since in general $\mathfrak{M}_\alpha \not\subset \mathcal{S}(\mathbb{R}^2)$. To obtain a suitable reservoir for $V_{\mathfrak{M}_\alpha}$, we thus need to slightly modify the procedure used for the extension of $V_{\mathfrak{C}_\alpha}$.

The basic idea for the subsequent exposition stems from abstract coorbit theory and goes back to [42, 43]. Instead of $\mathcal{S}(\mathbb{R}^2)$, we consider a subspace $\mathcal{H}^\nu_1$ of $L_2(\mathbb{R}^2)$ of the form

$$\mathcal{H}^\nu_1 := \Big\{ f \in L_2(\mathbb{R}^2) \ : \ V_{\mathfrak{M}_\alpha} f \in L^\nu_1(\mathbb{X}) \Big\}, \tag{4.28}$$

where $\nu : \mathbb{X} \to [1,\infty)$ is some suitable weight function. With $\| \cdot | \mathcal{H}^\nu_1 \| := \| V_{\mathfrak{M}_\alpha}(\cdot) | L^\nu_1 \|$ as a norm, this space is topologized differently than by the usual subspace topology induced by $L_2(\mathbb{R}^2)$. Provided that the canonical injection $\mathcal{H}^\nu_1 \hookrightarrow L_2(\mathbb{R}^2)$ is continuous and dense, we then obtain a Gelfand triple resembling (4.6), namely

$$\mathcal{H}^\nu_1 \hookrightarrow L_2(\mathbb{R}^2) \hookrightarrow (\mathcal{H}^\nu_1)^\urcorner. \tag{4.29}$$

Here, one usually uses the anti-dual $(\mathcal{H}^\nu_1)^\urcorner$, i.e., the space of bounded conjugate-linear functionals on $\mathcal{H}^\nu_1$. This has the advantage that the duality product $\langle \cdot, \cdot \rangle_{(\mathcal{H}^\nu_1)^\urcorner \times \mathcal{H}^\nu_1}$ canonically extends the scalar product $\langle \cdot, \cdot \rangle$ on $L_2(\mathbb{R}^2)$, which by convention is conjugate-linear in the second component.

In view of (4.29), if the weight $\nu$ is chosen such that $\mathfrak{M}_\alpha \subset \mathcal{H}^\nu_1$, the transform $V_{\mathfrak{M}_\alpha}$ can be extended via the duality product

$$V_{\mathfrak{M}_\alpha} f(\mathbf{x}) := \langle f, m_{\mathbf{x}} \rangle_{(\mathcal{H}^\nu_1)^\urcorner \times \mathcal{H}^\nu_1} , \quad \mathbf{x} \in \mathbb{X}. \tag{4.30}$$

The difficulty of this approach is to find a suitable weight $\nu : \mathbb{X} \to [1,\infty)$ such that the above conditions are fulfilled, i.e., that $\mathcal{H}^\nu_1 \hookrightarrow L_2(\mathbb{R}^2)$ is a continuous and dense embedding and that $\mathfrak{M}_\alpha \subset \mathcal{H}^\nu_1$.

A helpful criterion can be found in [74, Lem. 2.14]. Let $m_\nu$ denote the bivariate weight associated to $\nu : \mathbb{X} \to \mathbb{R}_+$ given by

$$m_\nu(\mathbf{x}, \mathbf{y}) := \max \Big\{ \frac{\nu(\mathbf{x})}{\nu(\mathbf{y})}, \frac{\nu(\mathbf{y})}{\nu(\mathbf{x})} \Big\} , \quad \mathbf{x}, \mathbf{y} \in \mathbb{X}.$$

Then, according to [74, Lem. 2.14], the desired conditions essentially hold true if the frame elements of $\mathfrak{M}_\alpha$ satisfy $\|m_{\mathbf{x}}|L_2\| \lesssim \nu(\mathbf{x})$ for every $\mathbf{x} \in \mathbb{X}$ and if the associated Gramian kernel

$$\mathcal{G}[\mathfrak{M}_\alpha](\mathbf{x},\mathbf{y}) := \langle m_{\mathbf{y}}, m_{\mathbf{x}} \rangle\,, \quad \mathbf{x},\mathbf{y} \in \mathbb{X},$$

is contained in the kernel algebra $\mathcal{A}_{m_\nu}$ defined below (see also [74, eq. (2.8)]). Under these assumptions, we then have a Gelfand triple as in (4.29) and $m_{\mathbf{x}} \in \mathcal{H}_1^\nu$ holds true at least for almost all elements $m_{\mathbf{x}} \in \mathfrak{M}_\alpha$.

In our setting, the algebra $\mathcal{A}_{m_\nu}$ takes the form

$$\mathcal{A}_{m_\nu} := \Big\{ K : \mathbb{X} \times \mathbb{X} \to \mathbb{C} \text{ measurable } : \ \|K|\mathcal{A}_{m_\nu}\| < \infty \Big\},$$

with $\|K|\mathcal{A}_{m_\nu}\|$ being the expression

$$\max\Big\{ \operatorname*{ess\,sup}_{\mathbf{x}\in\mathbb{X}} \int_{\mathbf{y}\in\mathbb{X}} |K(\mathbf{x},\mathbf{y}) m_\nu(\mathbf{x},\mathbf{y})| \, d\mu(\mathbf{y})\,, \operatorname*{ess\,sup}_{\mathbf{y}\in\mathbb{X}} \int_{\mathbf{x}\in\mathbb{X}} |K(\mathbf{x},\mathbf{y}) m_\nu(\mathbf{x},\mathbf{y})| \, d\mu(\mathbf{x}) \Big\}.$$

A closer inspection of $\mathcal{A}_{m_\nu}$ reveals that $\mathcal{A}_{m_\nu}$ is a Banach space with $\|\cdot|\mathcal{A}_{m_\nu}\|$ as a norm. Further, with the multiplication given by (4.51), $\mathcal{A}_{m_\nu}$ even is a Banach algebra. However, since these details are not essential for the main exposition, they are outsourced to the appendix, Section 4.5.

It remains to find a suitable weight for the analyzing frame $\mathfrak{M}_\alpha$. A plausible ansatz are weights of the form

$$\nu_\gamma(x,\eta,t) := t^{-\gamma}\,, \quad \gamma \ge 0, \tag{4.31}$$

which promote decay of the transform $|V_{\mathfrak{M}_\alpha} f(x,\eta,t)|$ along the scale variable $t$ in the direction $t \searrow 0$. In view of Paley-Wiener, such weights are associated to the smoothness of the functions $f \in \mathcal{H}_1^\nu$.

Next, we utilize Theorem 4.5.5 from the appendix. It gives a condition when a Gramian kernel $\mathcal{G}[\mathfrak{M}_\alpha]$ belongs to $\mathcal{A}_{m_\nu}$ for weights of type $\nu = \nu_\gamma$. Even more, the condition in Theorem 4.5.5 ensures that the so-called *cross-Gramian maximal kernels*, associated to two possibly different systems of $\alpha$-molecules $\mathfrak{M}_\alpha$ and $\widetilde{\mathfrak{M}}_\alpha$, belong to $\mathcal{A}_{m_\nu}$.

For $\tau \ge 0$, those are given as follows,

$$\mathcal{M}^\alpha_\tau[\widetilde{\mathfrak{M}}_\alpha, \mathfrak{M}_\alpha](\mathbf{x},\mathbf{y}) := \sup_{\mathbf{z}\in U^\alpha_\tau(\mathbf{y})} |\langle m_{\mathbf{x}}, \tilde{m}_{\mathbf{z}} \rangle|\,, \quad \mathbf{x},\mathbf{y} \in \mathbb{X}, \tag{4.32}$$

where $U^\alpha_\tau(\mathbf{y}) \subset \mathbb{X}$ are the subsets from (4.15). This definition should be compared with the so-called *cross-Gramian kernels* (see also (2.7)) defined by

$$\mathcal{G}[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha](\mathbf{x},\mathbf{y}) := \langle m_{\mathbf{x}}, \tilde{m}_{\mathbf{y}} \rangle\,, \quad \mathbf{x},\mathbf{y} \in \mathbb{X}.$$

Note, in particular, that for $\tau = 0$ we have

$$|\mathcal{G}[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha]| = \mathcal{M}^\alpha_0[\widetilde{\mathfrak{M}}_\alpha, \mathfrak{M}_\alpha]. \tag{4.33}$$

To simplify the notation in case $\mathfrak{M}_\alpha = \widetilde{\mathfrak{M}}_\alpha$, we put $\mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha] := \mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \mathfrak{M}_\alpha]$ in line with $\mathcal{G}[\mathfrak{M}_\alpha] := \mathcal{G}[\mathfrak{M}_\alpha, \mathfrak{M}_\alpha]$.

Applying Theorem 4.5.5 in combination with [74, Lem. 2.14] yields the following result.

**Lemma 4.3.2.** *Let $\alpha \in [0,1]$, and let $\mathfrak{M}_\alpha = \{m_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ be a tight frame of $\alpha$-molecules of order $(L, M, N_1, N_2)$ with respect to the canonical parametrization. Further assume that $\nu = \nu_\gamma$ is a weight of the form* (4.31) *with $\gamma \geq 0$ such that*

$$L > 2(\gamma+2)\,, \quad M > 3(\gamma+2) - \frac{3-\alpha}{2}\,, \quad N_1 > \gamma + 2 + \frac{1+\alpha}{2}\,, \quad N_2 > 2(\gamma+2)\,. \quad (4.34)$$

*Then the space $\mathcal{H}_1^\nu$ defined as in* (4.28) *is a Banach space with associated norm $\|\cdot|\mathcal{H}_1^\nu\| := \|V_{\mathfrak{M}_\alpha}(\cdot)|L_1^\nu\|$. Moreover, it is continuously and densely embedded in $L_2(\mathbb{R}^2)$ and $\mathfrak{M}_\alpha$ is a total subset of $\mathcal{H}_1^\nu$.*

*Proof.* If condition (4.34) is fulfilled, we have $N_1 > 1$ and thus, by Lemma 2.1.4, $\|m_{\mathbf{x}}|L_2\| \lesssim 1 \leq \nu(\mathbf{x})$ uniformly in $\mathbf{x} \in \mathbb{X}$. Further, due to Theorem 4.5.5, condition (4.34) also ensures that $\mathcal{M}_\tau^\alpha[\mathfrak{M}_\alpha] \in \mathcal{A}_{m_\nu}$ for arbitrary $\tau \geq 0$. As a consequence of (4.33) and the solidity of $\mathcal{A}_{m_\nu}$, then also $\mathcal{G}[\mathfrak{M}_\alpha] \in \mathcal{A}_{m_\nu}$. Hence, the prerequisites of [74, Lem. 2.14] are fulfilled, and thus $\mathcal{H}_1^\nu$ is a Banach space, continuously and densely embedded in $L_2(\mathbb{R}^2)$. Furthermore, almost all frame elements $m_{\mathbf{x}} \in \mathfrak{M}_\alpha$ are contained in $\mathcal{H}_1^\nu$.

Using $\mathcal{M}_\tau^\alpha[\mathfrak{M}_\alpha] \in \mathcal{A}_{m_\nu}$ for $\tau > 0$, we can even show the full inclusion $\mathfrak{M}_\alpha \subset \mathcal{H}_1^\nu$. To this end, let $\mathbf{x} \in \mathbb{X}$ be arbitrary but fixed. Then there exists $\tilde{\mathbf{x}} \in \mathbb{X}$ such that $\mathbf{x} \in U_\tau^\alpha(\tilde{\mathbf{x}})$ and

$$\int_{\mathbb{X}} \mathcal{M}_\tau^\alpha[\mathfrak{M}_\alpha](\tilde{\mathbf{x}},\mathbf{y}) m_\nu(\tilde{\mathbf{x}},\mathbf{y})\, d\mu(\mathbf{y}) < \infty.$$

Now we can deduce $m_{\mathbf{x}} \in \mathcal{H}_1^\nu$ since $V_{\mathfrak{M}_\alpha}(m_{\mathbf{x}}) \in L_1^\nu(\mathbb{X})$ due to the estimate

$$\|V_{\mathfrak{M}_\alpha}(m_{\mathbf{x}})|L_1^\nu\| = \|\mathcal{G}[\mathfrak{M}_\alpha](\mathbf{x},\cdot)\nu(\cdot)|L_1\| \leq \nu(\tilde{\mathbf{x}})\|\mathcal{M}_\tau^\alpha[\mathfrak{M}_\alpha](\tilde{\mathbf{x}},\cdot)m_\nu(\tilde{\mathbf{x}},\cdot)|L_1\| < \infty.$$

This proves $\mathfrak{M}_\alpha \subset \mathcal{H}_1^\nu$, and in view of [74, Cor. 2.20] the system $\mathfrak{M}_\alpha$ is even total in $\mathcal{H}_1^\nu$. □

As a consequence of Lemma 4.3.2, if the order $(L, M, N_1, N_2)$ of the system $\mathfrak{M}_\alpha$ satisfies (4.34), the transform $V_{\mathfrak{M}_\alpha}$ extends naturally to $(\mathcal{H}_1^\nu)^\urcorner$ with $\nu = \nu_\gamma$. In the following proposition, we collect some properties of the extended transform $V_{\mathfrak{M}_\alpha}$.

**Proposition 4.3.3.** *Under the assumptions of Lemma 4.3.2 the transform $V_{\mathfrak{M}_\alpha}$ defined in* (4.27) *possesses a natural extension to $(\mathcal{H}_1^\nu)^\urcorner$ given by* (4.30). *The extended transform is an injective bounded linear operator*

$$V_{\mathfrak{M}_\alpha} : (\mathcal{H}_1^\nu)^\urcorner \to L_\infty^{1/\nu}(\mathbb{X}),$$

*mapping signals $f \in (\mathcal{H}_1^\nu)^\urcorner$ to functions $V_{\mathfrak{M}_\alpha} f \in L_\infty^{1/\nu}(\mathbb{X})$ given by*

$$V_{\mathfrak{M}_\alpha} f(\mathbf{x}) = \langle f, m_{\mathbf{x}}\rangle_{(\mathcal{H}_1^\nu)^\urcorner \times \mathcal{H}_1^\nu}\,, \quad \mathbf{x} \in \mathbb{X}.$$

*Proof.* This follows from [74, Lem. 2.15] and [74, Cor. 2.19] since the prerequisites of [74, Lem. 2.14] are fulfilled, as shown in the proof of Lemma 4.3.2. □

Finally, we clarify the relation of the auxiliary space $\mathcal{H}_1^\nu$ with $\nu = \nu_\gamma$ to the Schwartz space $\mathcal{S}(\mathbb{R}^2)$.

**Lemma 4.3.4.** *Assume that the assumptions of Lemma 4.3.2 are fulfilled. Then we have the continuous embedding $\mathcal{S}(\mathbb{R}^2) \hookrightarrow \mathcal{H}_1^\nu$ for $\nu = \nu_\gamma$.*

*Proof.* Let $f \in \mathcal{S}(\mathbb{R}^2)$ and $F(\mathbf{x}) := V_{\mathfrak{M}_\alpha} f(\mathbf{x}) = \langle f, m_{\mathbf{x}} \rangle$ with $m_{\mathbf{x}} \in \mathfrak{M}_\alpha$. We interpret $f$ as a single $\alpha$-molecule with phase space coordinates $(0,0,1)$. According to the assumptions, there is $N > 2+\gamma$ such that Theorem 2.2.2 yields for all $(x,\eta,t) \in \mathbb{X}$

$$|V_{\mathfrak{M}_\alpha} f(x,\eta,t)| = |\langle f, m_{\mathbf{x}} \rangle| \leq C_{N,f} \cdot t^N (1+|x|)^{-N}$$

with a constant $C_{N,f} > 0$ depending on $N$ and $f$. Since $N > 2$ and $N - (2+\gamma) > 0$, we obtain, with $d\mu_0 = \frac{dx\, d\eta\, dt}{t^3}$ and $d\mu_1 = dx\, d\eta$,

$$\int_{\mathbb{X}_0} |F(\mathbf{x})| \nu_\gamma(\mathbf{x})\, d\mu_0(\mathbf{x}) \leq C_{N,f} \int_0^1 \int_{\mathbb{T}} \int_{\mathbb{R}^2} t^{N-(2+\gamma)} (1+|x|)^{-N} \frac{dx\, d\eta\, dt}{t} < \infty,$$
$$\int_{\mathbb{X}_1} |F(\mathbf{x})| \nu_\gamma(\mathbf{x})\, d\mu_1(\mathbf{x}) \leq C_{N,f} \int_{\mathbb{T}} \int_{\mathbb{R}^2} (1+|x|)^{-N}\, dx\, d\eta < \infty.$$

This shows $F \in L_1^\nu(\mathbb{X}, \mu)$ for $\nu = \nu_\gamma$ and thus $f \in \mathcal{H}_1^\nu$.

To show the continuity of the embedding $\mathcal{S} \hookrightarrow \mathcal{H}_1^\nu$, let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{S}$ such that $f_n \to 0$ in $\mathcal{S}$. Then there exists a sequence of constants $C_n > 0$ with $C_n \to 0$ as $n \to \infty$ such that for all $\rho \in \mathbb{N}_0^2$ with $|\rho|_1 \leq L$

$$|\partial^\rho \hat{f}_n(\xi)| \leq C_n \cdot \langle |\xi| \rangle^{-N_1} \langle \xi_2 \rangle^{-N_2} \quad \text{for all } \xi = (\xi_1, \xi_2) \in \mathbb{R}^2.$$

In particular, the system $\{C_n^{-1} f_n : n \in \mathbb{N}\}$ is a system of $\alpha$-molecules of order $(M, L, N_1, N_2)$ with respect to the parametrization $n \mapsto (x_n, \eta_n, t_n) := (0,0,1)$. Invoking Theorem 2.2.2 thus yields a constant $C_N \geq 0$ such that

$$|F_n(\mathbf{x})| := |V_{\mathfrak{M}_\alpha} f_n(\mathbf{x})| = |\langle f_n, m_{\mathbf{x}} \rangle| \leq C_n C_N \cdot t^N (1+|x|)^{-N} \quad \text{for all } \mathbf{x} \in \mathbb{X} \text{ and } n \in \mathbb{N}.$$

As a consequence, $\|F_n | L_1^\nu\| \lesssim C_n$ for $n \in \mathbb{N}$. This proves $\|F_n | L_1^\nu\| \to 0$ and verifies $\mathcal{S} \hookrightarrow \mathcal{H}_1^\nu$ since $f_n \to 0$ in $\mathcal{H}_1^\nu$. $\square$

### 4.3.3 $\alpha$-Molecule Coorbit Spaces associated to $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$

With a suitable reservoir for the continuous $\alpha$-molecule transform $V_{\mathfrak{M}_\alpha}$ at hand, we can now proceed to define coorbits of $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$ with respect to $V_{\mathfrak{M}_\alpha}$.

**Definition 4.3.5.** Let $\mathfrak{M}_\alpha$ be a tight frame of $\alpha$-molecules of order $(L, M, N_1, N_2)$ with respect to the canonical parametrization. Assume that $\gamma \geq 0$ satisfies condition (4.34) and let $\nu_\gamma$ denote the weight from (4.31). Then the *$\alpha$-molecule coorbit space* with respect to $\nu_\gamma$ and $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$ is defined as

$$\mathsf{Co}(\nu_\gamma, \mathfrak{M}_\alpha, \mathbb{L}_{p,q}^{\alpha,s}) := \left\{ f \in (\mathcal{H}_1^{\nu_\gamma})^\urcorner \ : \ V_{\mathfrak{M}_\alpha} f \in \mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X}) \right\}$$

and equipped with the quasi-norm $\| \cdot | \mathsf{Co}(\nu_\gamma, \mathfrak{M}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})\| := \|V_{\mathfrak{M}_\alpha}(\cdot) | \mathbb{L}_{p,q}^{\alpha,s}\|$.

Due to Proposition 4.3.3, these spaces are well-defined. In contrast to the $\alpha$-curvelet coorbits $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})$ from Definition 4.3.1, they also depend on the utilized weight $\nu_\gamma$.

Let us investigate some properties of these spaces, where we again resort to the abstract theory from [74]. We know from [74, Def. 2.25] and [74, Thm. 2.31] that they constitute quasi-Banach spaces if the analyzing frame $\mathfrak{M}_\alpha$ has PROPERTY $F(\nu, Y)$ with respect to the weight $\nu = \nu_\gamma$ and the space $Y = \mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$.

Before we recall this notion from [74, Def. 2.24], it is convenient to introduce a subalgebra of $\mathcal{A}_{m_\nu}$ as follows. Let $Y$ be a rich solid QBF-space on $\mathbb{X}$, and let $\nu : \mathbb{X} \to [1,\infty)$ be a measurable weight. Then we define

$$\mathcal{B}_{m_\nu,Y} := \Big\{ K : \mathbb{X} \times \mathbb{X} \to \mathbb{C} \ : \ K \in \mathcal{A}_{m_\nu} \text{ and } K : Y \to Y \text{ is bounded} \Big\}.$$

More details on these spaces are provided in the appendix, Section 4.5. In particular, it is shown that $\mathcal{B}_{m_\nu,Y}$ is a solid quasi-Banach algebra endowed with the quasi-norm

$$\|K|\mathcal{B}_{m_\nu,Y}\| := \max\{\|K|\mathcal{A}_{m_\nu}\|, \|K|Y \to Y\|\}.$$

Now we are ready to give the definition of PROPERTY $F(\nu, Y)$ in the concrete situation of an $\alpha$-molecule frame $\mathfrak{M}_\alpha$.

**Definition 4.3.6** (compare [74, Def. 2.24]). Let $\nu \geq 1$ be a weight on $\mathbb{X}$ and let $Y$ be a rich solid QBF-space on $\mathbb{X}$. A tight frame of $\alpha$-molecules $\mathfrak{M}_\alpha = \{m_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ is said to have PROPERTY $F(\nu, Y)$ if

(i) $\|m_{\mathbf{x}}|L_2\| \leq C_B \nu(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{X}$ and some fixed constant $C_B > 0$,

(ii) $\mathcal{G}[\mathfrak{M}_\alpha] \in \mathcal{B}_{m_\nu,Y}$,

(iii) $\mathcal{G}[\mathfrak{M}_\alpha] : Y \to L^{1/\nu}_\infty(\mathbb{X})$ is a continuous operator from $Y$ to $L^{1/\nu}_\infty(\mathbb{X})$.

We now derive a sufficient condition on the order of an $\alpha$-molecule frame $\mathfrak{M}_\alpha$ such that conditions (i)-(iii) are fulfillable for $Y = \mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$. Recall $\mathbf{v}^{\alpha,s}_{p,q}$, the weight from (4.24) associated to $\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$. Subsequently, we use the modified weight $\nu^{\alpha,s}_{p,q} : \mathbb{X} \to [1,\infty)$ given by

$$\nu^{\alpha,s}_{p,q}(\mathbf{x}) := \max\{1, \mathbf{v}^{\alpha,s}_{p,q}(\mathbf{x})\} = t^{-\max\{0,\tilde{s}\}}, \quad \mathbf{x} = (x,\eta,t) \in \mathbb{X}, \tag{4.35}$$

where $\tilde{s} := s - (1+\alpha)/p - (1-\alpha)/q$. Note that this weight is a weight of the form $\nu_\gamma$ as in (4.31) with $\gamma = \max\{0, \tilde{s}\} \geq 0$.

Finally, we are ready to prove the first result concerning $\alpha$-molecule coorbits of $\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$.

**Theorem 4.3.7.** *Let $\alpha \in [0,1]$, and let $\mathfrak{M}_\alpha = \{m_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ be a tight frame of $\alpha$-molecules of order $(L, M, N_1, N_2)$ with respect to the canonical parametrization. Further, let $\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$ be the space from Definition 4.2.7 with fixed parameters $0 < p, q < \infty$ and $s \in \mathbb{R}$. If*

$$L > 2(\rho+2)\,, \quad M > 3(\rho+2) - \frac{3-\alpha}{2}\,, \quad N_1 > \rho + 2 + \frac{1+\alpha}{2}\,, \quad N_2 > 2(\rho+2) \tag{4.36}$$

*holds true for $\rho := \max\{|s| + 2(1/r - 1), |\tilde{s}|\}$, where $r := \min\{1, p, q\}$ and $\tilde{s} := s - (1+\alpha)/p - (1-\alpha)/q$, then the $\alpha$-molecule coorbit space*

$$\mathsf{Co}(\nu^{\alpha,s}_{p,q}, \mathfrak{M}_\alpha, \mathbb{L}^{\alpha,s}_{p,q}) = \Big\{ f \in (\mathcal{H}_1^{\nu^{\alpha,s}_{p,q}})^\urcorner \ : \ V_{\mathfrak{M}_\alpha} f \in \mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X}) \Big\}$$

*is a quasi-Banach space with quasi-norm $\|\cdot|\mathsf{Co}(\nu^{\alpha,s}_{p,q}, \mathfrak{M}_\alpha, \mathbb{L}^{\alpha,s}_{p,q})\| := \|V_{\mathfrak{M}_\alpha}(\cdot)|\mathbb{L}^{\alpha,s}_{p,q}\|$. The quasi-norm constant of $\mathsf{Co}(\nu^{\alpha,s}_{p,q}, \mathfrak{M}_\alpha, \mathbb{L}^{\alpha,s}_{p,q})$ is thereby inherited from $\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$.*

*Proof.* Let us assume that the order of the $\alpha$-molecule frame $\mathfrak{M}_\alpha$ fulfills condition (4.36). We will show that under this assumption $\mathfrak{M}_\alpha$ has Property $F(\nu, Y)$ as introduced in Definition 4.3.6 for the choice $\nu := \nu_{p,q}^{\alpha,s}$ and $Y := \mathbb{L}_{p,q}^{\alpha,s}$. An application of [74, Thm. 2.31] thus finishes the proof.

It remains to verify conditions (i)-(iii) in Definition 4.3.6. Since $N_1 > 1$ we have $\|m_{\mathbf{x}}|L_2\| \lesssim \nu(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{X}$ by Lemma 2.1.4, proving (i). Further, as a consequence of Theorem 4.5.5, condition (ii) holds true, namely $\mathcal{G}[\mathfrak{M}_\alpha] \in \mathcal{B}_{m_\nu,Y}$ which entails that $\mathcal{G}[\mathfrak{M}_\alpha] : Y \to Y$ is a bounded linear operator. As a consequence of the embedding $Y \hookrightarrow L_\infty^{1/\nu}$, this also implies that the operation $\mathcal{G}[\mathfrak{M}_\alpha] : Y \to L_\infty^{1/\nu}$ is well-defined and continuous. Hence, also condition (iii) is true. $\square$

An important feature of the theory of $\alpha$-molecules is the transfer principle, already encountered in Theorem 2.3.6 of Section 2.3 in a discrete setup. It allows to transfer certain properties between $\alpha$-molecule systems via the concept of sparsity equivalence.

The next result, Theorem 4.3.8, can be interpreted as another occurrence of the transfer principle, this time in a continuous setting. It relates the $\alpha$-molecule coorbit spaces $\mathsf{Co}(\nu_{p,q}^{\alpha,s}, \mathfrak{M}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})$ to the $\alpha$-curvelet coorbits from Definition 4.3.1. Its proof is based on Theorem 4.5.5(ii) and the abstract result [74, Lem. 2.28].

**Theorem 4.3.8.** *For every tight frame of $\alpha$-molecules $\mathfrak{M}_\alpha$ subject to the assumptions of Theorem 4.3.7 we have the identification*

$$\mathsf{Co}(\nu_{p,q}^{\alpha,s}, \mathfrak{M}_\alpha, \mathbb{L}_{p,q}^{\alpha,s}) \asymp \mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})$$

*in the sense of equivalent quasi-norms.*

*Proof.* First of all note that the $\alpha$-curvelet frame $\mathfrak{C}_\alpha$ is a concrete instance of an $\alpha$-molecule frame satisfying the assumptions of Theorem 4.3.7. Hence, the coorbit $\mathsf{Co}(\nu_{p,q}^{\alpha,s}, \mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})$ is a well-defined quasi-Banach space. Further, we have shown in Lemma 4.3.4 that $\mathcal{S} \hookrightarrow \mathcal{H}_1^\nu$ for $\nu := \nu_{p,q}^{\alpha,s}$ and, according to Proposition 4.1.5, the reproducing formula of $V_{\mathfrak{C}_\alpha}$ extends to $\mathcal{S}'$. Hence, all prerequisites to apply [74, Lem. 2.28] in our concrete situation are fulfilled. We obtain

$$\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s}) \asymp \mathsf{Co}(\nu_{p,q}^{\alpha,s}, \mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s}).$$

Let us now turn to a general $\alpha$-molecule frame $\mathfrak{M}_\alpha$ subject to the assumptions of Theorem 4.3.7. It follows from Theorem 4.5.5 that the cross-Gramian maximal kernels $\mathcal{M}_\tau^\alpha[\mathfrak{C}_\alpha, \mathfrak{M}_\alpha]$ and $\mathcal{M}_\tau^\alpha[\mathfrak{M}_\alpha, \mathfrak{C}_\alpha]$ (see (4.32) for definition) both belong to $\mathcal{B}_{m_\nu,Y}$ with $Y := \mathbb{L}_{p,q}^{\alpha,s}$ and $\nu := \nu_{p,q}^{\alpha,s}$. As a consequence of the solidity of $\mathcal{B}_{m_\nu,Y}$ (see Proposition 4.5.4), in particular $\mathcal{G}[\mathfrak{C}_\alpha, \mathfrak{M}_\alpha] \in \mathcal{B}_{m_\nu,Y}$ and $\mathcal{G}[\mathfrak{M}_\alpha, \mathfrak{C}_\alpha] \in \mathcal{B}_{m_\nu,Y}$ hold true. With [74, Lem. 2.29] we can therefore deduce

$$\mathsf{Co}(\nu_{p,q}^{\alpha,s}, \mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s}) \asymp \mathsf{Co}(\nu_{p,q}^{\alpha,s}, \mathfrak{M}_\alpha, \mathbb{L}_{p,q}^{\alpha,s}). \qquad \square$$

Theorem 4.3.8 is a powerful tool for the analysis of $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})$. As an immediate consequence, for example, we obtain in conjunction with Theorem 4.3.7 that $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})$ is a quasi-Banach space with the same quasi-norm constant as $\mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X})$. In the next subsection, we will further use this theorem to deduce a discrete characterization of $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})$. The obtained discrete description in Theorem 4.3.13 will then allow us to identify $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})$ with the shearlet smoothness spaces from [85].

### 4.3.4 Characterization via Discrete $\alpha$-Curvelets

In this subsection, we will apply Theorem 4.3.8 for a special choice of $\mathfrak{M}_\alpha$ leading to a discrete characterization of $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q})$. This will in particular relate these coorbit spaces to familiar scales of smoothness spaces considered for example in [85].

Let us first recall the discrete Parseval frame of $\alpha$-curvelets $\mathfrak{C}^\bullet_\alpha$ from Definition 3.2.6 indexed by the discrete $\alpha$-curvelet index set

$$M = \mathbb{J} \times \mathbb{Z}^2. \tag{4.37}$$

Hereby $\mathbb{J} = \{J = (j,\ell) : j \in \mathbb{N}_0, \ell \in \mathbb{L}_j\}$ is a collection of scale-angle pairs with $\mathbb{L}_j := \{0,1,\dots,L_j-1\}$ and $L_j := 2^{\lfloor j(1-\alpha)\rfloor}$ for $j \in \mathbb{N}_0$. The system $\mathfrak{C}^\bullet_\alpha$ has the form

$$\mathfrak{C}^\bullet_\alpha = \{\psi_\mu\}_{\mu\in M} = \left\{\psi_{j,\ell,k} : j \in \mathbb{N}_0, \ell \in \mathbb{L}_j, k \in \mathbb{Z}^2\right\}$$

and consists of band-limited $\alpha$-curvelets $\psi_\mu = \psi_{j,\ell,k} \in L_2(\mathbb{R}^2)$. It is useful to distinguish the coarse-scale elements from the high-scale functions, corresponding to respective index sets

$$M_1 := \mathbb{J}_1 \times \mathbb{Z}^2 \quad \text{with} \quad \mathbb{J}_1 := \{(0,0)\},$$

and $$M_0 := \mathbb{J}_0 \times \mathbb{Z}^2 \quad \text{with} \quad \mathbb{J}_0 := \{J = (j,\ell) : j \in \mathbb{N}, \ell \in \mathbb{L}_j\}.$$

It was shown in Lemma 3.2.7 that $\mathfrak{C}^\bullet_\alpha$ constitutes a Parseval frame for $L_2(\mathbb{R}^2)$. Moreover, according to Proposition 3.2.8, it is a system of $\alpha$-molecules of order $(\infty,\infty,\infty,\infty)$ with respect to the specific $\alpha$-curvelet parametrization (3.28), namely $(M, \Phi_M)$ with

$$\Phi_M : \ M \to \mathbb{P} \ : \ (j,\ell,k) \mapsto (x_{j,\ell,k}, \ell\omega_j, 2^{-j}) = \left(R^{-1}_{\ell\omega_j} A^{-1}_{\alpha,2^j} k, \ell\omega_j, 2^{-j}\right),$$

where $\omega_j := 2^{-\lfloor j(1-\alpha)\rfloor}\pi$. For convenience, we will subsequently use the abbreviations $R_{j,\ell} := R_{\ell\omega_j}$ and $A_j := A_{\alpha,2^j}$ as in (3.24).

Recall the subsets $U^\alpha_\tau(\mathbf{x})$ of $\mathbb{X}$ defined in (4.15) in Section 4.2. For $\mu = (j,\ell,k) \in M$ and $\tau > 0$, let us now define the sets

$$U^{\alpha,\tau}_\mu := U^{\alpha,\tau}_{j,\ell,k} := U^\alpha_\tau(\mathbf{x}_{j,\ell,k}) \quad \text{with } \mathbf{x}_{j,\ell,k} := \mathbf{x}_\mu := \Phi_M(\mu).$$

More concretely, those sets can be written in the form

$$U^{\alpha,\tau}_{j,\ell,k} = Q^{\alpha,\tau}_{j,\ell,k} \times I^{\alpha,\tau}_{j,\ell} \times J^\tau_j, \tag{4.38}$$

where, with $Q^\tau := [-\tau,\tau]^2$, $I^\tau := [-\tau,\tau]$, and $J^\tau := [2^{-\tau}, 2^\tau]$,

$$Q^{\alpha,\tau}_{j,\ell,k} := x_{j,\ell,k} + R^{-1}_{j,\ell} A^{-1}_j Q^\tau,\ I^{\alpha,\tau}_{j,\ell} := \left(\ell\omega_j + 2^{-j(1-\alpha)} I^\tau\right)_{2\pi},\ J^\tau_j := \begin{cases} 2^{-j} J^\tau \cap (0,1) & ,\ j \in \mathbb{N}, \\ \{1\} & ,\ j = 0. \end{cases}$$

For the choice $\tau = \frac{1}{3}$ the collection $\{U^{\alpha,\tau}_\mu\}_{\mu\in M}$ consists of pairwise disjoint subsets of $\mathbb{X}$. We now put $U_\mu := U^{\alpha,\frac{1}{3}}_\mu$ and introduce the positive weights

$$w_\mu := \int\limits_{\mathbb{X}} \mathcal{X}_{U_\mu}(\mathbf{x})\, d\mu(\mathbf{x}) > 0. \tag{4.39}$$

Then we define

$$\overline{\mathfrak{C}^{\bullet}_{\alpha}} := \{\tilde{\psi}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}} \quad \text{where} \quad \begin{cases} \tilde{\psi}_{x,\eta,t} := \psi_\mu \omega_\mu^{-\frac{1}{2}}\,, & (x,\eta,t)\in U_\mu, \\ \tilde{\psi}_{x,\eta,t} := 0\,, & \text{else.} \end{cases}$$

We call $\overline{\mathfrak{C}^{\bullet}_{\alpha}}$ the *continuization* of the discrete $\alpha$-curvelet frame $\mathfrak{C}^{\bullet}_{\alpha}$.

**Lemma 4.3.9.** *The system $\overline{\mathfrak{C}^{\bullet}_{\alpha}} = \{\tilde{\psi}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ is a continuous Parseval frame of $\alpha$-molecules of order $(\infty,\infty,\infty,\infty)$ with respect to the canonical parametrization.*

*Proof.* The system $\overline{\mathfrak{C}^{\bullet}_{\alpha}} = \{\tilde{\psi}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ inherits the Parseval property from $\mathfrak{C}^{\bullet}_{\alpha} = \{\psi_\mu\}_{\mu\in M}$. Indeed, for every $f\in L_2(\mathbb{R}^2)$ we have

$$\int_{\mathbb{X}} |\langle f,\tilde{\psi}_{\mathbf{x}}\rangle|^2\,d\mu(\mathbf{x}) = \sum_{\mu\in M}\int_{U_\mu} |\langle f,\psi_\mu\rangle|^2 w_\mu^{-1}\,d\mu(\mathbf{x}) = \sum_{\mu\in M} |\langle f,\psi_\mu\rangle|^2 = \|f|L_2\|^2.$$

For the verification of the $\alpha$-molecule property of $\overline{\mathfrak{C}^{\bullet}_{\alpha}}$, we define for $(x,\eta,t)\in\mathbb{X}$

$$h_{x,\eta,t} := t^{(1+\alpha)/2}\tilde{\psi}_{x,\eta,t}\big(R_\eta^{-1}A_{\alpha,t}\cdot+x\big)$$

such that

$$\tilde{\psi}_{x,\eta,t} = t^{-(1+\alpha)/2}h_{x,\eta,t}\big(A_{\alpha,t}^{-1}R_\eta(\cdot-x)\big).$$

To finish the proof, it remains to show condition (2.5) for the functions $h_{x,\eta,t}$.

We first recall that, according to (3.26), every $\alpha$-curvelet $\psi_{j,\ell,k}\in\mathfrak{C}^{\bullet}_{\alpha}$ has the form

$$\psi_{j,\ell,k} = 2^{j(1+\alpha)/2}a_j\big(A_jR_{j,\ell}\cdot-k\big)$$

with a generator $a_j\in L_2(\mathbb{R}^2)$ given by (3.25).

Hence, in case $(x,\eta,t)\in U_\mu$ for some $\mu=(j,\ell,k)\in M$, we can deduce

$$\begin{aligned} h_{x,\eta,t} &= t^{(1+\alpha)/2}\tilde{\psi}_{x,\eta,t}\big(R_\eta^{-1}A_{\alpha,t}\cdot+x\big) \\ &= \omega_\mu^{-\frac{1}{2}}(t2^j)^{(1+\alpha)/2}a_j\big(A_jR_{j,\ell}R_\eta^{-1}A_{\alpha,t}\cdot+(A_jR_{j,\ell}x-k)\big). \end{aligned}$$

Altogether, since $\tilde{\psi}_{x,\eta,t}=0$ if $(x,\eta,t)\notin U_\mu$ for every $\mu\in M$, we obtain the spatial representation

$$h_{x,\eta,t} = \begin{cases} \omega_\mu^{-\frac{1}{2}}(\Delta t)^{(1+\alpha)/2}a_j\big(T_{\eta,t}^{-T}\cdot+\Delta k\big)\,, & \text{if } (x,\eta,t)\in U_\mu, \\ 0\,, & \text{else,} \end{cases}$$

where we abbreviate $\Delta t := t2^j$, $\Delta k := A_jR_{j,\ell}x-k$, and

$$T_{\eta,t} := A_j^{-1}R_{j,\ell}R_\eta^{-1}A_{\alpha,t}^{-1} =: \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix}.$$

On the Fourier side, we get

$$\hat{h}_{x,\eta,t} = \begin{cases} \omega_\mu^{-\frac{1}{2}}(\Delta t)^{-(1+\alpha)/2}\hat{a}_j\big(T_{\eta,t}\cdot\big)\exp\Big(2\pi i\langle\Delta k, T_{\eta,t}\cdot\rangle\Big)\,, & \text{if } (x,\eta,t)\in U_\mu, \\ 0\,, & \text{else.} \end{cases}$$

Our first goal is now to show that for every fixed $\rho = (\rho_1, \rho_2) \in \mathbb{N}_0^2$

$$\sup_{\xi \in \mathbb{R}^2} |\partial^\rho \hat{h}_{x,\eta,t}(\xi)| \lesssim 1 \quad \text{uniformly in } (x, \eta, t) \in \mathbb{X}. \tag{4.40}$$

Note that since $\hat{h}_{x,\eta,t} \in C^\infty(\mathbb{R}^2)$ the derivatives $\partial^\rho \hat{h}_{x,\eta,t}$ are well-defined. With the chain rule, we calculate for $\xi \in \mathbb{R}^2$

$$\begin{aligned}
\partial_1(\hat{a}_j(T_{\eta,t}\cdot))(\xi) &= t_{11}(\partial_1 \hat{a}_j)(T_{\eta,t}\xi) + t_{21}(\partial_2 \hat{a}_j)(T_{\eta,t}\xi), \\
\partial_2[\hat{a}_j(T_{\eta,t}\cdot)](\xi) &= t_{12}(\partial_1 \hat{a}_j)(T_{\eta,t}\xi) + t_{22}(\partial_2 \hat{a}_j)(T_{\eta,t}\xi).
\end{aligned}$$

By iteration, we obtain for $\rho = (\rho_1, \rho_2) \in \mathbb{N}_0^2$ the expansion

$$\partial_1^{\rho_1} \partial_2^{\rho_2} [\hat{a}_j(T_{\eta,t}\cdot)](\xi) = \sum_{a=0}^{\rho_1} \sum_{b=0}^{\rho_2} n_{a,b} S_{a,b}(\xi; \eta, t)$$

with combinatorial coefficients $n_{a,b} \in \mathbb{N}$ and terms $S_{a,b}(\xi; \eta, t)$ of the form

$$S_{a,b}(\xi; \eta, t) := (t_{11})^a (t_{21})^{\rho_1 - a} (t_{12})^b (t_{22})^{\rho_2 - b} (\partial_1^{a+b} \partial_2^{\rho_1 + \rho_2 - a - b} \hat{a}_j)(T_{\eta,t}\xi).$$

To estimate these terms, note that $T_{\eta,t}$ is a matrix of the type investigated in Lemma 2.2.18 and Corollary 2.2.19. With the notation there, we have

$$T_{\eta,t}^{-T} = A_j R_{j,\ell} R_\eta^{-1} A_{\alpha,t} = T(\ell\omega_j, 2^{-j}, \eta, t).$$

In the relevant case $(x, \eta, t) \in U_\mu$, it holds $\max\{\Delta t, (\Delta t)^{-1}\} \leq 2^\tau$ and $d_{\mathbb{S}}(\ell\omega_j, \eta) \leq \tau 2^{-j(1-\alpha)}$ for $\tau = \frac{1}{3}$. Hence, we can apply Lemma 2.2.18 and obtain that the entries of $T_{\eta,t}^{-T}$ and $T_{\eta,t}^T$ are uniformly bounded. As a consequence, also the entries $t_{11}$, $t_{12}$, $t_{21}$, and $t_{22}$ of $T_{\eta,t}$ are uniformly bounded. This yields $|S_{a,b}(\xi; \eta, t)| \lesssim |(\partial_1^{a+b} \partial_2^{\rho_1+\rho_2-a-b} \hat{a}_j)(T_{\eta,t}\xi)|$.

Further, we recall from the proof of Proposition 3.2.8 that, for any given $\rho = (\rho_1, \rho_2) \in \mathbb{N}_0^2$, we have uniformly in $j \in \mathbb{N}_0$

$$\sup_{\xi \in \mathbb{R}^2} |\partial_1^{\rho_1} \partial_2^{\rho_2} \hat{a}_j(\xi)| \lesssim 1.$$

Hence, we obtain

$$\sup_{\xi \in \mathbb{R}^2} |(\partial_1^{a+b} \partial_2^{\rho_1+\rho_2-a-b} \hat{a}_j)(T_{\eta,t}\xi)| \lesssim 1 \quad \text{uniformly in } (x, \eta, t) \in \mathbb{X}.$$

All in all, this proves $|S_{a,b}(\xi; \eta, t)| \lesssim 1$ and we can deduce

$$\sup_{\xi \in \mathbb{R}^2} |\partial_1^{\rho_1} \partial_2^{\rho_2} [\hat{a}_j(T_{\eta,t}\cdot)](\xi)| \lesssim 1 \quad \text{uniformly in } (x, \eta, t) \in \mathbb{X}.$$

For the exponential term, we note

$$\exp(2\pi i \langle \Delta k, T_{\eta,t}\cdot \rangle) = \exp(2\pi i \langle T_{\eta,t}^T(\Delta k), \cdot \rangle).$$

Whenever $(x, \eta, t) \in U_\mu = U_{j,\ell,k}$ we have $x \in Q_{j,\ell,k}^{\alpha,\frac{1}{3}}$ and thus

$$\Delta k = A_j R_{j,\ell} x - k = A_j R_{j,\ell}(x - x_{j,\ell,k}) \in Q^{\frac{1}{3}}.$$

In other words $\|\Delta k\|_\infty \leq \frac{1}{3}$. We now invoke Corollary 2.2.19 which yields $\|T_{\eta,t}^T\|_{\infty\to\infty} \leq 2^{\frac{1}{3}}(1+\frac{1}{3}) \leq 2$. Hence, we get $T_{\eta,t}^T(\Delta k) \in Q^2 = [-2,2]^2$.

For every $\rho = (\rho_1,\rho_2) \in \mathbb{N}_0^2$ with $|\rho|_1 \leq L$ and every $(x,\eta,t) \in \mathbb{X}$ we thus obtain

$$\|\partial_1^{\rho_1}\partial_2^{\rho_2} \exp\left(2\pi i\langle \Delta k, T_{\eta,t}\cdot\rangle\right)|L_\infty\| \leq (4\pi)^L \lesssim 1.$$

Finally, we take care of the prefactor $\omega_\mu^{-\frac{1}{2}}(\Delta t)^{-(1+\alpha)/2}$. Clearly, $2^{-\frac{1}{3}} \leq \Delta t \leq 2^{\frac{1}{3}}$ due to $(x,\eta,t) \in U_\mu$. We further have $w_\mu \asymp 1$ by Corollary 4.2.12.

Taken all together, our estimates prove (4.40).

Next, we analyze the support properties of the functions $\partial^\rho \hat{h}_{x,\eta,t}$. From the support of the generators $\hat{a}_j$ (see (3.27) and the following discussion) one can directly deduce that there exist constants $C \geq c > 0$ such that $\operatorname{supp}\hat{a}_0 \subset [-C,C]^2$ and for $j \geq 1$

$$\operatorname{supp}\hat{a}_j \subset [-C,C]^2 \backslash \left([-c,c] \times [-2^{j(1-\alpha)}c, 2^{j(1-\alpha)}c]\right).$$

Furthermore, we clearly have

$$\operatorname{supp}\partial^\rho \hat{h}_{x,\eta,t} \subseteq \operatorname{supp}\hat{h}_{x,\eta,t} = \operatorname{supp}\hat{a}_j(T_{\eta,t}\cdot\,).$$

An application of Corollary 2.2.19 yields $\|T_{\eta,t}\|_{\infty\to\infty} \leq 2^{\frac{1}{3}}(1+\frac{1}{3}) \leq 2$ and $\|T_{\eta,t}^{-1}\|_{\infty\to\infty} \leq 2^{\frac{1}{3}}(1+\frac{1}{3}) \leq 2$. Hence, in case $(x,\eta,t) \in U_\mu$ for $\mu = (j,\ell,k) \in M_0$ with $j \geq 1$, we obtain

$$\operatorname{supp}\partial^\rho \hat{h}_{x,\eta,t} \subseteq [-2C,2C]^2 \backslash \left([-\tfrac{c}{2},\tfrac{c}{2}] \times [-2^{j(1-\alpha)}\tfrac{c}{2}, 2^{j(1-\alpha)}\tfrac{c}{2}]\right). \tag{4.41}$$

If $(x,\eta,t) \in U_\mu$ for $\mu = (0,\ell,k) \in M_1$ with $j = 0$, we necessarily have $t = 1$ and

$$\operatorname{supp}\partial^\rho \hat{h}_{x,\eta,1} \subseteq [-2C,2C]^2. \tag{4.42}$$

The uniform boundedness of the functions $\partial^\rho \hat{h}_{x,\eta,t}$ shown in (4.40) together with the support properties (4.41) and (4.42) imply condition (2.5) for $\partial^\rho \hat{h}_{x,\eta,t}$ for arbitrary orders $(L,M,N_1,N_2)$. This is the same argument already encountered in Propositions 3.1.3 and 3.2.8. The proof is finished. $\square$

Let us remark that the utilized continuization procedure does not work in a generic $\alpha$-molecule setting. In the proof of Lemma 4.3.9 we have based our arguments on the fact that the $\alpha$-curvelets are bandlimited, a specific property stronger than the generic $\alpha$-molecule decay conditions. In a general setting, the proof would thus not go through as above.

Since $\overline{\mathfrak{C}_\alpha^\bullet}$ is a continuous Parseval frame of $\alpha$-molecules with respect to the canonical parametrization, we can now apply Theorem 4.3.8 to obtain the following result.

**Lemma 4.3.10.** *We have the equivalence*

$$\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s}) \asymp \mathsf{Co}(\overline{\mathfrak{C}_\alpha^\bullet}, \mathbb{L}_{p,q}^{\alpha,s}) := \left\{ f \in \mathcal{S}'(\mathbb{R}^2) \;:\; V_{\overline{\mathfrak{C}_\alpha^\bullet}} f \in \mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X}) \right\},$$

*where* $\mathsf{Co}(\overline{\mathfrak{C}_\alpha^\bullet}, \mathbb{L}_{p,q}^{\alpha,s})$ *is equipped with the quasi-norm* $\|\cdot|\mathsf{Co}(\overline{\mathfrak{C}_\alpha^\bullet}, \mathbb{L}_{p,q}^{\alpha,s})\| := \|V_{\overline{\mathfrak{C}_\alpha^\bullet}}(\cdot)|\mathbb{L}_{p,q}^{\alpha,s}\|$.

*Proof.* For the proof, we show

$$\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q}) \asymp \mathsf{Co}(\nu^{\alpha,s}_{p,q}, \overline{\mathfrak{C}^\bullet_\alpha}, \mathbb{L}^{\alpha,s}_{p,q}) \asymp \mathsf{Co}(\overline{\mathfrak{C}^\bullet_\alpha}, \mathbb{L}^{\alpha,s}_{p,q}).$$

Since $\mathfrak{C}_\alpha$ and $\overline{\mathfrak{C}^\bullet_\alpha}$ are both Parseval frames of $\alpha$-molecules with respect to the canonical parametrization, and since for both the respective order is arbitrarily large (see Propositions 3.1.1 and 3.1.3 and Lemma 4.3.9), an application of Theorem 4.3.8 yields the first equivalence

$$\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q}) \asymp \mathsf{Co}(\nu^{\alpha,s}_{p,q}, \overline{\mathfrak{C}^\bullet_\alpha}, \mathbb{L}^{\alpha,s}_{p,q}).$$

It remains to prove the second equivalence

$$\mathsf{Co}(\nu^{\alpha,s}_{p,q}, \overline{\mathfrak{C}^\bullet_\alpha}, \mathbb{L}^{\alpha,s}_{p,q}) \asymp \mathsf{Co}(\overline{\mathfrak{C}^\bullet_\alpha}, \mathbb{L}^{\alpha,s}_{p,q}).$$

Subsequently, we argue similarly as in the proof of Theorem 4.3.8, with $\mathfrak{C}_\alpha$ replaced by $\overline{\mathfrak{C}^\bullet_\alpha}$. First of all notice that $\overline{\mathfrak{C}^\bullet_\alpha} = \{\tilde{\psi}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}} \subset \mathcal{S}$, and hence the associated transform $V_{\overline{\mathfrak{C}^\bullet_\alpha}}$ extends naturally to $\mathcal{S}'$. The identification of the spaces is then proved by applying the abstract result [74, Lem. 2.28]. For this, we need to show that the reproducing formula of $V_{\overline{\mathfrak{C}^\bullet_\alpha}}$ extends to $\mathcal{S}'$, which works analogously as for $\mathfrak{C}_\alpha$. In essence, we need to imitate the arguments of Subsection 4.1.2 and in particular transfer Lemma 4.1.2.

The proof of this key lemma is based on (4.9) and (4.10). Hence we need to adapt these estimates to $\overline{\mathfrak{C}^\bullet_\alpha}$. Concerning (4.10), we obtain analogously for $\varphi \in \mathcal{S}$ and arbitrary but fixed $\tilde{N} > 0$

$$|V_{\overline{\mathfrak{C}^\bullet_\alpha}}\varphi(x,\eta,t)| = |\langle \varphi, \tilde{\psi}_{x,\eta,t}\rangle| \le C_{\tilde{N},\varphi} \cdot t^{\tilde{N}}(1+|x|^2)^{-\tilde{N}}$$

with a constant $C_{\tilde{N},\varphi} > 0$ independent of $(x,\eta,t) \in \mathbb{X}$. In view of Lemma 4.3.9, this is a direct consequence of Theorem 2.2.2.

Concerning (4.9), we first observe that an analogous estimate holds true for the discrete $\alpha$-curvelets $\psi_{j,\ell,k} \in \mathfrak{C}^\bullet_\alpha$. There exists a constant $C_N > 0$ such that

$$\|\psi_{j,\ell,k}\|_N \le C_N 2^{j(1+\alpha)/2} 2^{jN} (1+|x_{j,\ell,k}|_2)^N \quad \text{uniformly for all } (j,\ell,k) \in M. \tag{4.43}$$

This follows analogously as in the proof of Lemma 4.1.2. The only difference is that, instead of the generators $g_t$ of the continuous $\alpha$-curvelet frame given by (3.9), we need to use the functions $a_t$ as in (3.25).

Now we can also show that, with a constant $\tilde{C}_N > 0$,

$$\|\tilde{\psi}_{\mathbf{x}}\|_N \le \tilde{C}_N t^{-(1+\alpha)/2} t^{-N} (1+|x|_2)^N \quad \text{for all } \mathbf{x} = (x,\eta,t) \in \mathbb{X}.$$

Indeed, let $\mathbf{x} = (x,\eta,t) \in U^{\alpha,\frac{1}{3}}_{j,\ell,k}$. Then there exist $\Delta t \in J^{\frac{1}{3}} = [2^{-\frac{1}{3}}, 2^{\frac{1}{3}}]$ and $\Delta k \in Q^{\frac{1}{3}} = [-\frac{1}{3}, \frac{1}{3}]^2$ such that $x = R^{-1}_{j,\ell} A^{-1}_j (k + \Delta k)$ and $2^j t = \Delta t$. Now we use (4.43) and

$$1 + |x_{j,\ell,k}|_2 = 1 + |A^{-1}_j k|_2 \le (1 + |A^{-1}_j (k+\Delta k)|_2)(1 + |A^{-1}_j \Delta k|_2) \le (1+|x|_2)(1+|\Delta k|_2)$$

to get with $\tilde{C}_N := C_N 2^{(1+\alpha)/6+N/3}(1+\sqrt{2}/3)^N$ as desired

$$\|\tilde{\psi}_{\mathbf{x}}\|_N = \|\psi_{j,\ell,k}\|_N \le C_N 2^{j(1+\alpha)/2} 2^{jN} (1 + |A^{-1}_j k|_2)^N \le \tilde{C}_N t^{-(1+\alpha)/2} t^{-N} (1+|x|_2)^N.$$

The rest of the proof is analogous to Subsection 4.1.2. □

Next, we introduce a scale of sequence spaces on the discrete $\alpha$-curvelet index set $M$.

**Definition 4.3.11.** Let $\alpha \in [0,1]$, $s \in \mathbb{R}$, $0 < p,q < \infty$. Further, $M = \mathbb{J} \times \mathbb{Z}^2$ shall be the $\alpha$-curvelet index set as defined in (4.37). We then define the sequence space

$$\ell^s_{p,q}(M) := \Big\{ (c_\mu)_{\mu \in M} \subset \mathbb{C} \;:\; \|(c_\mu)_\mu | \ell^s_{p,q}\| < \infty \Big\},$$

with associated quasi-norm defined for $(c_\mu)_\mu = (c_{j,\ell,k})_{j,\ell,k}$ by

$$\|(c_\mu)_\mu | \ell^s_{p,q}\| := \Big( \sum_{(j,\ell)\in\mathbb{J}} 2^{jsq} \Big( \sum_{k\in\mathbb{Z}^2} |c_{j,\ell,k}|^p \Big)^{q/p} \Big)^{1/q}.$$

The following lemma is the final discretization step on our way to Theorem 4.3.13.

**Lemma 4.3.12.** *Define $\tilde{s} := s - (1+\alpha)/p - (1-\alpha)/q$. Then*

$$\mathsf{Co}(\overline{\mathfrak{C}^\bullet_\alpha}, \mathbb{L}^{\alpha,s}_{p,q}) \asymp \mathsf{Co}(\mathfrak{C}^\bullet_\alpha, \ell^{\tilde{s}}_{p,q}) := \Big\{ f \in \mathcal{S}'(\mathbb{R}^2) \;:\; V_{\mathfrak{C}^\bullet_\alpha} f \in \ell^{\tilde{s}}_{p,q} \Big\},$$

*where $\mathsf{Co}(\mathfrak{C}^\bullet_\alpha, \ell^{\tilde{s}}_{p,q})$ is equipped with the quasi-norm $\|\cdot|\mathsf{Co}(\mathfrak{C}^\bullet_\alpha, \ell^{\tilde{s}}_{p,q})\| := \|V_{\mathfrak{C}^\bullet_\alpha}(\cdot)|\ell^{\tilde{s}}_{p,q}\|$.*

*Proof.* Let $f \in \mathcal{S}'$, and let $\mathfrak{C}^\bullet_\alpha = \{\psi_\mu\}_{\mu\in M}$ be the discrete $\alpha$-curvelet frame defined in Definition 3.2.6. We have, with $c_\mu := \langle f, \psi_\mu \rangle_{\mathcal{S}'\times\mathcal{S}}$ being the curvelet coefficients,

$$V_{\overline{\mathfrak{C}^\bullet_\alpha}} f(\mathbf{x}) = \sum_\mu c_\mu \omega_\mu^{-1/2} \mathcal{X}^{\alpha,\frac{1}{3}}_\mu(\mathbf{x}), \quad \mathbf{x} \in \mathbb{X},$$

where $\omega_\mu = \omega_{j,\ell,k}$ are the weights defined in (4.39). Further, the short-hand notation $\mathcal{X}^{\alpha,\tau}_\mu := \mathcal{X}^{\alpha,\tau}_{j,\ell,k} := \mathcal{X}_{U^{\alpha,\tau}_{j,\ell,k}}$ for $\tau > 0$ is used. Recalling (4.38), namely $U^{\alpha,\tau}_{j,\ell,k} = Q^{\alpha,\tau}_{j,\ell,k} \times I^{\alpha,\tau}_{j,\ell} \times J^\tau_j$, it is useful to decompose

$$\mathcal{X}^{\alpha,\tau}_{j,\ell,k}(x,\eta,t) = \chi^\tau_j(t) \cdot \chi^{\alpha,\tau}_{j,\ell}(\eta) \cdot \chi^{\alpha,\tau}_{j,\ell,k}(x), \quad (x,\eta,t) \in \mathbb{X},$$

with $\chi^\tau_j = \mathcal{X}_{J^\tau_j}$, $\chi^{\alpha,\tau}_{j,\ell} = \mathcal{X}_{I^{\alpha,\tau}_{j,\ell}}$, and $\chi^{\alpha,\tau}_{j,\ell,k} = \mathcal{X}_{Q^{\alpha,\tau}_{j,\ell,k}}$.

Since, as a consequence of Corollary 4.2.12, we have $\omega_\mu \asymp 1$ uniformly for all $\mu \in M$, we can deduce for $\mathbf{x} \in \mathbb{X}$

$$|V_{\overline{\mathfrak{C}^\bullet_\alpha}} f(\mathbf{x})| = \sum_\mu |c_\mu| \omega_\mu^{-1/2} \mathcal{X}^{\alpha,\frac{1}{3}}_\mu(\mathbf{x}) \asymp \sum_\mu |c_\mu| \mathcal{X}^{\alpha,\frac{1}{3}}_\mu(\mathbf{x}).$$

For arbitrary $\tau > 0$, we then calculate for the homogeneous component,

$$
\begin{aligned}
&\Big( \int_0^1 \int_{\mathbb{T}} t^{-sq} \Big\| \sum_{j,\ell,k} c_{j,\ell,k} \mathcal{X}^{\alpha,\tau}_{j,\ell,k}(\cdot,\eta,t) \Big| L_p \Big\|^q \frac{d\eta\, dt}{t} \Big)^{1/q} \\
&= \Big( \int_0^1 \int_{\mathbb{T}} t^{-sq} \Big( \sum_{j,\ell} \chi^{\tau}_j(t) \chi^{\alpha,\tau}_{j,\ell}(\eta) \sum_k |c_{j,\ell,k}|^p \|\chi^{\alpha,\tau}_{j,\ell,k}|L_1\| \Big)^{q/p} \frac{d\eta\, dt}{t} \Big)^{1/q} \\
&\asymp \Big( \int_0^1 \int_{\mathbb{T}} t^{-sq} \sum_{j,\ell} \chi^{\tau}_j(t) \chi^{\alpha,\tau}_{j,\ell}(\eta) \Big( \sum_k |c_{j,\ell,k}|^p \|\chi^{\alpha,\tau}_{j,\ell,k}|L_1\| \Big)^{q/p} \frac{d\eta\, dt}{t} \Big)^{1/q} \\
&= \Big( \sum_{j,\ell} \int_0^1 \int_{\mathbb{T}} t^{-sq} \chi^{\tau}_j(t) \chi^{\alpha,\tau}_{j,\ell}(\eta) \frac{d\eta\, dt}{t} \Big( \sum_k |c_{j,\ell,k}|^p \|\chi^{\alpha,\tau}_{j,\ell,k}|L_1\| \Big)^{q/p} \Big)^{1/q} \\
&\asymp \Big( \sum_{(j,\ell)\in\mathbb{J}_0} 2^{j\tilde{s}q} \Big( \sum_k |c_{j,\ell,k}|^p \Big)^{q/p} \Big)^{1/q} = \|(c_\mu)_\mu | \ell^{\tilde{s}}_{p,q}(M_0)\|,
\end{aligned}
$$

with $\tilde{s} = s - (1+\alpha)/p - (1-\alpha)/q$ and implicit constants depending on $\tau$.

Hence we deduce, with $\tau > 0$ sufficiently large,

$$
\begin{aligned}
\|(c_\mu)_\mu | \ell^{\tilde{s}}_{p,q}(M_0)\| &\asymp \Big( \int_0^1 \int_{\mathbb{T}} t^{-sq} \Big\| \sum_\mu c_\mu \mathcal{X}^{\alpha,\frac{1}{3}}_\mu \Big| L_p \Big\|^q \frac{d\eta\, dt}{t} \Big)^{1/q} \asymp \|V_{\overline{\mathfrak{C}^\bullet_\alpha}} f | L^s_{p,q}(\mathbb{X}_0)\| \\
&\le \|V_{\overline{\mathfrak{C}^\bullet_\alpha}} f | \mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X}_0)\| \lesssim \Big( \int_0^1 \int_{\mathbb{T}} t^{-sq} \Big\| \sum_\mu c_\mu \mathcal{X}^{\alpha,\tau}_\mu \Big| L_p \Big\|^q \frac{d\eta\, dt}{t} \Big)^{1/q} \asymp \|(c_\mu)_\mu | \ell^{\tilde{s}}_{p,q}(M_0)\|.
\end{aligned}
$$

Here we used $\|V_{\overline{\mathfrak{C}^\bullet_\alpha}} f | \mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X}_0)\| = \|\mathbf{W}^\alpha V_{\overline{\mathfrak{C}^\bullet_\alpha}} f | L^s_{p,q}(\mathbb{X}_0)\|$, where $\mathbf{W}^\alpha = \mathbf{W}^\alpha_1$ is the $\alpha$-anisotropic Wiener maximal operator from (4.17), and that we can estimate for sufficiently large $\tau > 0$,

$$
\mathbf{W}^\alpha V_{\overline{\mathfrak{C}^\bullet_\alpha}} f(\mathbf{x}) \le \sum_\mu |c_\mu| \omega_\mu^{-1/2} \mathbf{W}^\alpha \mathcal{X}^{\alpha,\frac{1}{3}}_\mu(\mathbf{x}) \lesssim \sum_\mu |c_\mu| \mathcal{X}^{\alpha,\tau}_\mu(\mathbf{x}) , \quad \mathbf{x} \in \mathbb{X}.
$$

Similarly, we argue for the inhomogeneous part. Here, for $\tau > 0$ sufficiently large,

$$
\begin{aligned}
\|(c_\mu)_\mu | \ell^{\tilde{s}}_{p,q}(M_1)\| &\asymp \Big( \int_{\mathbb{T}} \Big\| \sum_\mu c_\mu \mathcal{X}^{\alpha,\frac{1}{3}}_\mu(\cdot,\eta,1) \Big| L_p \Big\|^q d\eta \Big)^{1/q} \asymp \|V_{\overline{\mathfrak{C}^\bullet_\alpha}} f | L^s_{p,q}(\mathbb{X}_1)\| \\
&\le \|V_{\overline{\mathfrak{C}^\bullet_\alpha}} f | \mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X}_1)\| \lesssim \Big( \int_{\mathbb{T}} \Big\| \sum_\mu c_\mu \mathcal{X}^{\alpha,\tau}_\mu(\cdot,\eta,1) \Big| L_p \Big\|^q d\eta \Big)^{1/q} \asymp \|(c_\mu)_\mu | \ell^{\tilde{s}}_{p,q}(M_1)\|,
\end{aligned}
$$

since for arbitrary $\tau > 0$

$$\Big( \int_{\mathbb{T}} \Big\| \sum_{j,\ell,k} c_{j,\ell,k} \mathcal{X}^{\alpha,\tau}_{j,\ell,k}(\cdot,\eta,1) \Big| L_p \Big\|^q \, d\eta \Big)^{1/q}$$
$$= \Big( \int_{\mathbb{T}} \Big( \sum_{j,\ell} \chi^\tau_j(1) \chi^{\alpha,\tau}_{j,\ell}(\eta) \sum_k |c_{j,\ell,k}|^p \|\chi^{\alpha,\tau}_{j,\ell,k}|L_1\| \Big)^{q/p} d\eta \Big)^{1/q}$$
$$\asymp \Big( \sum_{\ell\in\mathbb{L}_0} \int_{\mathbb{T}} \chi^{\alpha,\tau}_{0,\ell}(\eta)\, d\eta \Big( \sum_k |c_{0,\ell,k}|^p \|\chi^{\alpha,\tau}_{0,\ell,k}|L_1\| \Big)^{q/p} \Big)^{1/q}$$
$$\asymp \Big( \sum_{\ell\in\mathbb{L}_0} \Big( \sum_k |c_{0,\ell,k}|^p \Big)^{q/p} \Big)^{1/q} = \|(c_\mu)_\mu | \ell^{\tilde{s}}_{p,q}(M_1)\|.$$

Altogether, this proves $\mathsf{Co}(\overline{\mathfrak{C}^\bullet_\alpha}, \mathbb{L}^{\alpha,s}_{p,q}) \asymp \mathsf{Co}(\mathfrak{C}^\bullet_\alpha, \ell^{\tilde{s}}_{p,q})$. □

Finally, we can formulate the main result of this subsection, Theorem 4.3.13, giving a discrete characterization of $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q})$.

**Theorem 4.3.13.** *It holds with $\tilde{s} := s - (1+\alpha)/p - (1-\alpha)/q$ and equivalent quasi-norms*

$$\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q}) \asymp \mathsf{Co}(\mathfrak{C}^\bullet_\alpha, \ell^{\tilde{s}}_{p,q}).$$

*Proof.* For the proof, we just need to combine Lemma 4.3.10 and Lemma 4.3.12,

$$\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q}) \asymp \mathsf{Co}(\overline{\mathfrak{C}^\bullet_\alpha}, \mathbb{L}^{\alpha,s}_{p,q}) \asymp \mathsf{Co}(\mathfrak{C}^\bullet_\alpha, \ell^{\tilde{s}}_{p,q}).$$ □

We are now able to draw a connection between the coorbits $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q})$ and other known scales of function spaces. The discrete characterization of Theorem 4.3.13 allows to identify them with shearlet and curvelet smoothness spaces or more general decomposition spaces (see e.g. [6]), for which equivalent discretizations have been derived. For example, they coincide, up to equivalence of quasi-norms, with the shearlet smoothness spaces considered in [85].

The discretization procedure presented in this subsection, based on the continuization of a discrete frame, is more or less a hands-on technique to obtain discrete descriptions. A more systematic approach to discretizations is developed in the next section.

## 4.4 Discretization Theory

As we have seen in the previous section, the coorbit spaces $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q})$ coincide with the smoothness spaces defined and analyzed in [85]. In particular, they can also be characterized as decomposition spaces (see [6]). Both, the coorbit approach as well as the decomposition approach, offer their own advantages. One feature of coorbit theory is the rich and powerful discretization machinery it comes along with. In the following, it allows us to derive very general discrete descriptions of $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q})$ (see our main Theorems 4.4.19 and 4.4.21).

Let us first recall the collection $\mathcal{U}^\alpha_\tau = \mathcal{U}^\alpha_\tau[\mathbb{X}] = \{U^\alpha_\tau(\mathbf{x}) : \mathbf{x} \in \mathbb{X}\}$ consisting of the sets $U^\alpha_\tau(\mathbf{x})$ defined in (4.15) as

$$U^\alpha_\tau(\mathbf{x}) = \begin{cases} (x + R^{-1}_\eta A_{\alpha,t} Q^\tau) \times (\eta + t^{1-\alpha} I^\tau)_{2\pi} \times (tJ^\tau \cap (0,1)) & , \mathbf{x} \in \mathbb{X}_0\,, \\ (x + R^{-1}_\eta Q^\tau) \times (\eta + I^\tau)_{2\pi} \times (\{1\}) & , \mathbf{x} \in \mathbb{X}_1\,, \end{cases} \tag{4.44}$$

where $Q^\tau = [-\tau,\tau]^2$, $I^\tau = [-\tau,\tau]$, and $J^\tau = [2^{-\tau}, 2^\tau]$. This collection $\mathcal{U}^\alpha_\tau$ constitutes a continuous covering of $\mathbb{X}$ depending on $\alpha \in [0,1]$ and a density parameter $\tau > 0$. Later, by suitably sampling the transform domain $\mathbb{X}$, we will extract from $\mathcal{U}^\alpha_\tau$ discrete coverings of $\mathbb{X}$. Thereby even certain types of irregular samplings will be allowed.

By a sampling $\mathcal{P}_\Lambda$ of the curvelet domain $\mathbb{X}$ we thereby mean a pair $\mathcal{P}_\Lambda = (\Lambda, \mathcal{P})$ of some countable index set $\Lambda$ and a map $\mathcal{P} : \Lambda \to \mathbb{X}$. For convenience, this notation will also be used for the point family $\mathcal{P}_\Lambda = \{\mathbf{x}_\lambda\}_{\lambda\in\Lambda}$ where $\mathbf{x}_\lambda := \mathcal{P}(\lambda)$. A sampling of $\mathbb{X}$ at the points $\mathcal{P}_\Lambda$ naturally leads to the family

$$\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda] := \Big\{ U^\alpha_\tau(\mathbf{x}_\lambda) \ : \ \lambda \in \Lambda \Big\}$$

of subsets of $\mathbb{X}$, whereby $\alpha \in [0,1]$ and $\tau > 0$ are fixed parameters.

An important concept for the discretization theory is the notion of an admissible covering of the transform domain. Let us recall this notion [74, Def. 2.4] and extend it to admissible collections of subsets of $\mathbb{X}$.

**Definition 4.4.1.** Let $\Lambda$ be a countable index set. A collection $\mathcal{U} = \{U_\lambda\}_{\lambda\in\Lambda}$ of subsets of $\mathbb{X}$ is called *admissible* if it is non-empty, locally finite and if it further has the following properties:

i) The sets $U_\lambda$ are measurable, relatively compact, and have non-void interior;

ii) The *intersection number* $\mathfrak{s}(\mathcal{U})$ of the collection $\mathcal{U}$ is finite, i.e.,

$$\mathfrak{s}(\mathcal{U}) := \sup_{\lambda\in\Lambda} \sharp\{\mu \in \Lambda : U_\lambda \cap U_\mu \neq \emptyset\} < \infty \,. \tag{4.45}$$

Note that for a partition $\mathcal{U}$ the intersection number $\mathfrak{s}(\mathcal{U})$ is 1.

If additionally $\mathbb{X} = \bigcup_{\lambda\in\Lambda} U_\lambda$ is fulfilled we call $\mathcal{U}$ an *admissible covering.*

We call a collection $\mathcal{U}$ of subsets of $\mathbb{X}$ *relatively separated* if we can partition $\mathcal{U}$ into a finite number of subcollections, each of which only contains pairwise disjoint sets. The *order* of the separation is the smallest possible number of such subcollections.

Note, that admissible collections of subsets of $\mathbb{X}$, due to their local finiteness, are necessarily countable since $\mathbb{X}$ is $\sigma$-compact. Further, we will see in the next lemma that a collection of sets with finite intersection number $N$ is also relatively separated of order at most $N$. The other direction of this statement is not always true. However, for collections of the type $\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]$ the reverse implication also holds, which will be proved below in Corollary 4.4.7.

**Lemma 4.4.2.** *Let $\mathcal{U} = \{U_\lambda\}_{\lambda\in\Lambda}$ be a collection of subsets of $\mathbb{X}$. If the intersection number $\mathfrak{s}(\mathcal{U})$ defined in* (4.45) *is finite, then $\mathcal{U}$ is relatively separated.*

*Proof.* By Zorn's lemma we can find a maximal subcollection $\mathcal{U}_1$ consisting of pairwise disjoint sets. Taking out this subfamily decreases the intersection number of the remaining family by at least 1. Iterating this process, we end up with a partition $\mathcal{U} = \bigcup_{i=1}^r \mathcal{U}_i$ into at most $\mathfrak{s}(\mathcal{U})$ subcollections of pairwise disjoint sets. □

Next, we are going to characterize those samplings $\mathcal{P}_\Lambda = (\Lambda, \mathcal{P})$ of $\mathbb{X}$ which lead to admissible coverings of the form $\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]$.

### 4.4.1 Well-spread Families of Points in $\mathbb{X}$

We make the following definition in analogy to [43, Def. 3.2], where similar notions were introduced on locally compact groups.

**Definition 4.4.3.** Let $\alpha \in [0,1]$ and let $\tau > 0$. A family of points $\mathcal{P}_\Lambda = \{\mathbf{x}_\lambda\}_{\lambda\in\Lambda}$ in $\mathbb{X}$ is called $(\alpha,\tau)$-*dense* if $\bigcup_{\lambda\in\Lambda} U_\tau^\alpha(\mathbf{x}_\lambda) = \mathbb{X}$. It is called $(\alpha,\tau)$-*separated* if $U_\tau^\alpha(\mathbf{x}_\lambda) \cap U_\tau^\alpha(\mathbf{x}_\mu) = \emptyset$ for $\lambda \neq \mu \in \Lambda$. It is called *relatively $\alpha$-separated* if there exists $\tau > 0$ such that $\mathcal{P}_\Lambda$ is a finite union of $(\alpha,\tau)$-separated subfamilies. It is called *$\alpha$-well-spread* if it is relatively $\alpha$-separated and if it is $(\alpha,\tau)$-dense for some $\tau > 0$.

The finite union of relatively $\alpha$-separated families of points is again relatively $\alpha$-separated.

**Remark 4.4.4.** Note that every subset of $\mathbb{X}$ can be interpreted as a family by indexing each element with itself. This perspective allows to apply the notions defined for point families in Definition 4.4.3 to subsets of $\mathbb{X}$ in a canonical way.

We will later need the following intuitive fact for which we nevertheless provide a short proof for reasons of mathematical rigor.

**Lemma 4.4.5.** *Let $\mathcal{P}_\Lambda$ be an $(\alpha,\tau)$-separated family of points in $\mathbb{X}$, $\alpha \in [0,1]$ and $\tau > 0$ be fixed. Then there exists an extension of $\mathcal{P}_\Lambda$ to an $\alpha$-well-spread family of points in $\mathbb{X}$, which is still $(\alpha,\tau)$-separated.*

*Proof.* The proof is based upon Zorn's lemma. Let $\mathcal{P}(\Lambda)$ denote the sampling points of $\mathcal{P}_\Lambda$ interpreted as a subset of $\mathbb{X}$. Then consider the class $\mathfrak{C}$ of all $(\alpha,\tau)$-separated extensions of $\mathcal{P}(\Lambda)$, partially ordered by inclusion, whereby $(\alpha,\tau)$-separation shall be understood in the light of Remark 4.4.4. The union of the points in each chain in $\mathfrak{C}$ is again an element of the class $\mathfrak{C}$, i.e., an $(\alpha,\tau)$-separated point set. Moreover, it majorizes the chain and hence, by Zorn's lemma, there exists a maximal element in $\mathfrak{C}$, which we denote by $\mathcal{P}(\Lambda)^{\mathrm{ext}}$.

This subset of $\mathbb{X}$ is $(\alpha,\tau)$-separated and extends $\mathcal{P}(\Lambda)$. Moreover, $\mathcal{P}(\Lambda)^{\mathrm{ext}}$ is $(\alpha,\sigma)$-dense for $\sigma \geq h(\tau,\tau)$, since otherwise there would be $\mathbf{x} \notin \bigcup_{\mathbf{y}\in\mathcal{P}(\Lambda)^{\mathrm{ext}}} U_\sigma^\alpha(\mathbf{y})$, and thus $U_\tau^\alpha(\mathbf{x}) \cap \bigcup_{\mathbf{y}\in\mathcal{P}(\Lambda)^{\mathrm{ext}}} U_\tau^\alpha(\mathbf{y}) = \emptyset$ by Lemma 4.2.5(v). This would be a contradiction to the maximality of $\mathcal{P}(\Lambda)^{\mathrm{ext}}$. Finally note that the set $\mathcal{P}(\Lambda)^{\mathrm{ext}}$ is countable since, due to the $\sigma$-compactness of $\mathbb{X}$, all elements of $\mathfrak{C}$ are countable. In particular, the index set $\Lambda$ of $\mathcal{P}_\Lambda$ can be extended to some countable index set $\Lambda^{\mathrm{ext}}$, and $\mathcal{P}(\Lambda)^{\mathrm{ext}}$, suitably indexed by $\Lambda^{\mathrm{ext}}$, becomes an extension of the family $\mathcal{P}_\Lambda$ with the desired properties. $\square$

The following result establishes a connection between the relative $\alpha$-separation of points and the intersection number of associated $\alpha$-patches. Note that the finite union of set collections with finite intersection numbers need not have a finite intersection number any more. Even the union of two collections of pairwise disjoint sets can have an infinite intersection number in general.

**Lemma 4.4.6.** *Assume that $\mathcal{P}_\Lambda$ is a relatively $\alpha$-separated family of points in $\mathbb{X}$. Then for each $\tau > 0$ the family $\mathcal{U}_\tau^\alpha[\mathcal{P}_\Lambda]$ has a finite intersection number, dependent on $\tau$.*

*Proof.* Let $\tau > 0$ be fixed, and take an arbitrary $\mathbf{x} \in \mathcal{P}(\Lambda)$, where $\mathcal{P}(\Lambda)$ – as in the proof of the previous lemma – denotes the subset of $\mathbb{X}$ consisting of the points contained in the family $\mathcal{P}_\Lambda$. For the proof we will subsequently count the number of non-trivial intersections the set $U_\tau^\alpha(\mathbf{x})$ has with other sets in $\mathcal{U}_\tau^\alpha[\mathcal{P}_\Lambda]$.

Since $\mathcal{P}_\Lambda$ is relatively $\alpha$-separated there exist $\sigma > 0$, $N \in \mathbb{N}$, and a partition $\Lambda = \bigcup_{k=1}^N \Lambda^k$ such that the subfamilies $\mathcal{P}_{\Lambda^k}$ are $(\alpha, \sigma)$-separated. For all $\mathbf{y} \in \mathcal{P}(\Lambda)$ with $\mathbf{y} \in \mathbb{X} \backslash U^\alpha_{h(\tau,\tau)}(\mathbf{x})$ we have $U^\alpha_\tau(\mathbf{x}) \cap U^\alpha_\tau(\mathbf{y}) = \emptyset$ due to Lemma 4.2.5(v). Hence, it only remains to count the number of points in $\mathcal{P}(\Lambda^k) \cap U^\alpha_{h(\tau,\tau)}(\mathbf{x})$ for each $k \in \{1, \dots, N\}$. For this, we note that all the sets in $\mathcal{U}^\alpha_\sigma[\mathcal{P}_{\Lambda^k}]$ with centers in $U^\alpha_{h(\tau,\tau)}(\mathbf{x})$ are contained in the larger set $U^\alpha_{f(h(\tau,\tau),\sigma)}(\mathbf{x})$, which is true due to Lemma 4.2.5(iii). The volume of this set is upper-bounded by some constant $C > 0$, independent of $\mathbf{x} \in \mathbb{X}$, according to Corollary 4.2.12. Further, also by Corollary 4.2.12, the volumes of the sets $U^\alpha_\sigma(\mathbf{y})$, $\mathbf{y} \in \mathcal{P}(\Lambda)$, are uniformly lower-bounded by some $D > 0$. Since $\mathcal{P}_{\Lambda^k}$ is $(\alpha, \sigma)$-separated, we conclude that the number of points in $\mathcal{P}(\Lambda^k) \cap U^\alpha_{h(\tau,\tau)}(\mathbf{x})$ is bounded by $C/D$ for each $k \in \{1, \dots, N\}$. $\square$

As a consequence of Lemma 4.4.2 and Lemma 4.4.6, we can formulate the following result establishing relations between the notions of relatively $\alpha$-separated points, corresponding set collections, and finite intersection numbers.

**Corollary 4.4.7.** *Let $\mathcal{P}_\Lambda$ be a family of points in $\mathbb{X}$. For arbitrary $\tau > 0$ we have the equivalence:*

*i) $\mathcal{P}_\Lambda$ is relatively $\alpha$-separated,*

*ii) $\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]$ is relatively separated,*

*iii) $\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]$ has a finite intersection number.*

*Proof.* We use a circle argument. Assume $\mathcal{P}_\Lambda$ is relatively $\alpha$-separated. Then by Lemma 4.4.6 the collection $\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]$ has a finite intersection number. By Lemma 4.4.2 this in turn implies that $\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]$ is relatively separated. Finally, if $\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]$ is relatively separated, then there exist $N \in \mathbb{N}$ and a partition $\Lambda = \bigcup_{k=1}^N \Lambda^k$ such that $\mathcal{U}^\alpha_\tau[\mathcal{P}_{\Lambda^k}]$ consists of pairwise disjoint sets for every $k \in \{1, \dots, N\}$. As a consequence, the families $\mathcal{P}_{\Lambda^k}$ are $(\alpha, \tau)$-separated, and in turn $\mathcal{P}_\Lambda$ is relatively $\alpha$-separated. $\square$

As a consequence, with $\alpha \in [0, 1]$ and $\tau > 0$ fixed and $\mathcal{P}_\Lambda$ being a family of points in $\mathbb{X}$, we have the following equivalence:

$$\mathcal{P}_\Lambda \text{ is relatively } \alpha\text{-separated.} \qquad \Longleftrightarrow \qquad \mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda] \text{ is an admissible collection.}$$

Now we are ready to prove the close relationship between admissible coverings of the form $\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]$ and $\alpha$-well-spread point families. One should compare this result to [43, Lem. 3.3], for example.

**Proposition 4.4.8.** *Let $\mathcal{P}_\Lambda$ be a family of points in $\mathbb{X}$. The collection $\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]$ is an admissible covering of $\mathbb{X}$ if and only if $\mathcal{P}_\Lambda$ is $\alpha$-well-spread and $(\alpha, \tau)$-dense in $\mathbb{X}$. Moreover, if we have an admissible covering of the form $\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]$ for some family $\mathcal{P}_\Lambda$ in $\mathbb{X}$ and $\tau > 0$, then by choosing one point in each set one obtains an $\alpha$-well-spread family of points.*

*Proof.* Assume that $\mathcal{P}_\Lambda$ is $\alpha$-well-spread. Then $\mathcal{P}_\Lambda$ is relatively $\alpha$-separated and, according to Lemma 4.4.6, the intersection number of $\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]$ remains finite for all $\tau > 0$. Assuming, in addition, $(\alpha, \tau)$-density, the properties of an admissible covering are easy to establish.

Conversely, if $\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]$ is an admissible covering, then $\mathcal{P}_\Lambda$ is in particular $(\alpha, \tau)$-dense. Moreover, also $\mathcal{U}^\alpha_{\tau+1}[\mathcal{P}_\Lambda]$ is an admissible covering, whose intersection number $N$ is finite.

By Lemma 4.4.2 we can decompose $\mathcal{U}^{\alpha}_{\tau+1}[\mathcal{P}_\Lambda]$ into (at most) $N$ subfamilies $\mathcal{U}^{\alpha}_{\tau+1}[\mathcal{P}_{\Lambda^k}]$, $k \in \{1,\dots,N\}$, consisting of pairwise disjoint sets. A sampling $\tilde{\mathcal{P}}_k$ subordinate to $\mathcal{U}^{\alpha}_{\tau}[\mathcal{P}_{\Lambda^k}]$ is then $(\alpha,\sigma)$-separated for small enough $\sigma > 0$. Hence, $\tilde{\mathcal{P}}_\Lambda := \bigcup_k \tilde{\mathcal{P}}_k$ is relatively $\alpha$-separated. Moreover, for large enough $\rho > 0$, the sampling $\tilde{\mathcal{P}}_\Lambda$ is also $(\alpha,\rho)$-dense. This proves that it is $\alpha$-well-spread, and thus, as a special case, also $\mathcal{P}_\Lambda$ is $\alpha$-well-spread. □

### 4.4.2 Associated Sequence Spaces

Let $\mathcal{P}_\Lambda = (\Lambda, \mathcal{P})$ be an $\alpha$-well-spread sampling of $\mathbb{X}$. Then we can use $\mathcal{P}_\Lambda$ as a means to associate sequence spaces to $L^s_{p,q}(\mathbb{X})$ and $\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$. The definition and analysis of these spaces is the topic of this subsection.

Let us first introduce the abstract concept of a sequence space $Y^\flat\langle\mathcal{U}\rangle$ associated to a rich solid QBF-space $Y$ on $\mathbb{X}$ and an admissible collection $\mathcal{U} = \{U_\lambda\}_{\lambda\in\Lambda}$ of subsets of $\mathbb{X}$. Hereby we do not require $\mathcal{U}$ to be a covering as in [74, Def. 2.6].

**Definition 4.4.9** (compare [74, Def. 2.6])**.** Let $Y$ be a rich solid QBF-space on $\mathbb{X}$ and let $\mathcal{U} = \{U_\lambda\}_{\lambda\in\Lambda}$ be an admissible collection of subsets of $\mathbb{X}$. The sequence space $Y^\flat := Y^\flat\langle\mathcal{U}\rangle$ is defined as the space

$$Y^\flat\langle\mathcal{U}\rangle := \Big\{ \{c_\lambda\}_{\lambda\in\Lambda} \subset \mathbb{C} \ : \ \sum_{\lambda\in\Lambda} |c_\lambda| \mathcal{X}_{U_\lambda} \in Y \Big\},$$

equipped with the quasi-norm

$$\|\{c_\lambda\}|Y^\flat\langle\mathcal{U}\rangle\| := \Big\| \sum_{\lambda\in\Lambda} |c_\lambda| \mathcal{X}_{U_\lambda} \Big| Y \Big\|.$$

Note that, since $\mathbb{X}$ is $\sigma$-compact and $\mathcal{U}$ is admissible, the indices $\Lambda$ necessarily constitute a countable set. Further, since the collection $\mathcal{U}$ is locally finite, the sum $\sum_{\lambda\in\Lambda} |c_\lambda|\mathcal{X}_{U_\lambda}$ is always well-defined in a point-wise sense.

A sequence $\{c_\lambda\}_{\lambda\in\Lambda}$ can be viewed as a function on the index set $\Lambda$. Taking this perspective, the spaces $Y^\flat\langle\mathcal{U}\rangle$ are function spaces on $\Lambda$. Moreover, we can think of $\Lambda$ as being equipped with the discrete topology and the counting measure. Then the notion of a QBF-space on $\Lambda$ and the corresponding terminology are available (see [74]). For better distinction, we refer to the latter as *quasi-Banach sequence spaces*, abbreviated *QBS-spaces*.

The space $Y^\flat\langle\mathcal{U}\rangle$ inherits many properties from $Y$. It was shown in [74] that it constitutes a rich solid QBS-space with the same quasi-norm constant $C_Y$ as $Y$.

**Proposition 4.4.10** ([74, Prop. 2.7])**.** *The sequence space $Y^\flat\langle\mathcal{U}\rangle$ is a rich solid QBS-space with the same quasi-norm constant $C_Y$ as $Y$.*

*Proof.* We refer to [74, Prop. 2.7]. □

Depending on $\alpha \in [0,1]$ and $\tau > 0$, we now partition $\mathbb{X}$ as follows. We put $\tilde{\tau} := 2\pi/\lceil 2\pi/\tau \rceil$, which is an integer fraction of $2\pi$, and define for every $J = (j,\ell) \in \mathbb{N}_0^2$

$$\mathbb{X}^{\alpha,\tau}_J := \Big\{ \mathbf{x} = (x,\eta,t) \in \mathbb{X} \ : \ t \in 2^{-j\tau}[1, 2^\tau) \, , \eta \in 2^{-\lceil j\tau(1-\alpha)\rceil}(\ell\tilde{\tau} + [0,\tilde{\tau})) \Big\}.$$

Then $\mathbb{X} = \bigcup_{J\in\mathbb{N}_0^2} \mathbb{X}_J^{\alpha,\tau}$ is a partition of $\mathbb{X}$, whereby many of the sets $\mathbb{X}_J^{\alpha,\tau}$ are empty. Given a sampling $\mathcal{P} : \Lambda \to \mathbb{X}$ with $\lambda \mapsto \mathbf{x}_\lambda$, the above partition further induces a corresponding partition of the index set $\Lambda$, namely $\Lambda = \bigcup_{J\in\mathbb{N}_0^2} \Lambda_J^{\alpha,\tau}$ with

$$\Lambda_J^{\alpha,\tau} := \Lambda_J^{\alpha,\tau}[\mathcal{P}] := \Big\{\lambda \in \Lambda \;:\; \mathbf{x}_\lambda \in \mathbb{X}_J^{\alpha,\tau}\Big\}.$$

Now we are ready for the next definition.

**Definition 4.4.11.** Let $\alpha \in [0,1]$, $s \in \mathbb{R}$, $0 < p,q < \infty$. Let $\mathcal{P}_\Lambda = \{\mathbf{x}_\lambda\}_{\lambda\in\Lambda}$ be a relatively $\alpha$-separated family of points in $\mathbb{X}$. We then define the sequence space

$$\ell_{p,q}^{\alpha,s}[\mathcal{P}_\Lambda] := \Big\{\{c_\lambda\}_{\lambda\in\Lambda} \;:\; \Big\|\{c_\lambda\}_\lambda|\ell_{p,q}^{\alpha,s}[\mathcal{P}_\Lambda]\Big\| := \Big(\sum_{J\in\mathbb{N}_0^2}\Big(\sum_{\lambda\in\Lambda_J^{\alpha,1}[\mathcal{P}]} |\mathbf{v}_{p,q}^{\alpha,s}(\mathbf{x}_\lambda)c_\lambda|^p\Big)^{q/p}\Big)^{1/q} < \infty\Big\},$$

where $\mathbf{v}_{p,q}^{\alpha,s}$ is the weight from (4.24) given by

$$\mathbf{v}_{p,q}^{\alpha,s}(\mathbf{x}) := t^{-s+(1+\alpha)/p+(1-\alpha)/q}\,, \quad \mathbf{x} = (x,\eta,t) \in \mathbb{X}\,.$$

As a consequence of Proposition 4.4.10 and the characterization established below by Theorem 4.4.12, these spaces are rich solid QBS-spaces.

**Theorem 4.4.12.** *Let $\alpha \in [0,1]$ and let $\mathcal{P}_\Lambda = \{\mathbf{x}_\lambda\}_{\lambda\in\Lambda}$ be a family of points in $\mathbb{X}$ which is relatively $\alpha$-separated. Assume further that $s \in \mathbb{R}$ and $0 < p,q < \infty$. For every fixed $\tau > 0$ we then have the equivalence*

$$L_{p,q}^s(\mathbb{X})^\flat\langle\mathcal{U}_\tau^\alpha[\mathcal{P}_\Lambda]\rangle \asymp \ell_{p,q}^{\alpha,s}[\mathcal{P}_\Lambda].$$

*Proof.* Recall the definition of the sets $U_\tau^\alpha(\mathbf{x})$ from (4.44), namely

$$U_\tau^\alpha(\mathbf{x}) = \begin{cases}(x + R_\eta^{-1}A_{\alpha,t}Q^\tau) \times (\eta + t^{1-\alpha}I^\tau)_{2\pi} \times (tJ^\tau \cap (0,1)) & ,\, \mathbf{x} \in \mathbb{X}_0\,,\\ (x + R_\eta^{-1}Q^\tau) \times (\eta + I^\tau)_{2\pi} \times (\{1\}) & ,\, \mathbf{x} \in \mathbb{X}_1\,,\end{cases}$$

where $\mathbf{x} = (x,\eta,t) \in \mathbb{X}$ and $Q^\tau = [-\tau,\tau]^2$, $I^\tau = [-\tau,\tau]$, $J^\tau = [2^{-\tau},2^\tau]$. For the associated characteristic functions, we will subsequently use the short-hand notation $\mathcal{X}_{\mathbf{x}}^{\alpha,\tau} := \mathcal{X}_{U_\tau^\alpha(\mathbf{x})}$ as introduced in (4.21). Further, recall that one can decompose these functions in the form $\mathcal{X}_{\mathbf{x}}^{\alpha,\tau}(y,\theta,u) = \chi_{x,\eta,t}^{\alpha,\tau}(y)\cdot\chi_{\eta,t}^{\alpha,\tau}(\theta)\cdot\chi_t^\tau(u)$ with characteristic functions as in (4.22).

We will handle the proof for the homogeneous and the inhomogeneous parts of the sequence spaces separately. Accordingly, we decompose $\Lambda = \Lambda_0 \cup \Lambda_1$ into

$$\Lambda_0 := \{\lambda \in \Lambda \;:\; \mathbf{x}_\lambda \in \mathbb{X}_0\} \quad \text{and} \quad \Lambda_1 := \{\lambda \in \Lambda \;:\; \mathbf{x}_\lambda \in \mathbb{X}_1\}.$$

We further introduce the convenient notation

$$\|\{c_\lambda\}_\lambda\|_{\mathrm{hom}}^{(\tau)} := \|\{c_\lambda\}_{\lambda\in\Lambda_0}\Big|(L_{p,q}^s)^\flat\langle\mathcal{U}_\tau^\alpha[\mathcal{P}_{\Lambda_0}]\rangle\Big\|, \quad \|\{c_\lambda\}_\lambda\|_{\mathrm{in}}^{(\tau)} := \|\{c_\lambda\}_{\lambda\in\Lambda_1}\Big|(L_{p,q}^s)^\flat\langle\mathcal{U}_\tau^\alpha[\mathcal{P}_{\Lambda_1}]\rangle\Big\|.$$

Since the intersection number of $\mathcal{U}_\tau^\alpha[\mathcal{P}_{\Lambda_0}]$ is finite, we can rewrite the homogeneous part of the quasi-norm in the form

$$\begin{aligned}
\|\{c_\lambda\}_\lambda\|_{\mathrm{hom}}^{(\tau)} &= \Big( \int_0^1 \int_0^{2\pi} u^{-sq} \Big\| \sum_{\lambda\in\Lambda_0} |c_\lambda| \mathcal{X}_{\mathbf{x}_\lambda}^{\alpha,\tau}(\cdot,\theta,u) |L_p\Big\|^q \frac{d\theta\, du}{u}\Big)^{1/q} \\
&\asymp \Big( \int_0^1 \int_0^{2\pi} u^{-sq} \Big( \sum_{\lambda\in\Lambda_0} |c_\lambda|^p \chi_{t_\lambda}^{\tau}(u) \chi_{\eta_\lambda,t_\lambda}^{\alpha,\tau}(\theta) \|\chi_{x_\lambda,\eta_\lambda,t_\lambda}^{\alpha,\tau}|L_1\|\Big)^{q/p} \frac{d\theta\, du}{u}\Big)^{1/q} \\
&\asymp \Big( \int_0^1 \int_0^{2\pi} u^{-sq} \Big( \sum_{J\in\mathbb{N}\times\mathbb{N}_0} 2^{-j\sigma(1+\alpha)} \underbrace{\sum_{\lambda\in\Lambda_J^{\alpha,\sigma}} |c_\lambda|^p \chi_{t_\lambda}^{\tau}(u) \chi_{\eta_\lambda,t_\lambda}^{\alpha,\tau}(\theta)}_{(*)} \Big)^{q/p} \frac{d\theta\, du}{u}\Big)^{1/q},
\end{aligned} \tag{4.46}$$

where we used that for every fixed $\tau > 0$

$$\|\chi_{x,\eta,t}^{\alpha,\tau}|L_1\| \asymp |A_{\alpha,t}Q^\tau| \asymp t^{1+\alpha} \quad \text{holds uniformly in } (x,\eta,t)\in\mathbb{X}.$$

We next estimate $(*)$ from above and below.

First, we note that for $\lambda\in\Lambda_J^{\alpha,\sigma}$ and $J=(j,\ell)\in\mathbb{N}_0^2$ we have by definition

$$t_\lambda \in 2^{-j\sigma}[1,2^\sigma) \quad \text{and} \quad \eta_\lambda \in 2^{-\lceil j\sigma(1-\alpha)\rceil}(\ell\tilde\sigma + [0,\tilde\sigma)),$$

where $\tilde\sigma := 2\pi/\lceil 2\pi/\sigma\rceil$. We deduce

$$t_\lambda J^\tau \subseteq 2^{-j\sigma}[2^{-\tau},2^{\tau+\sigma}) \quad \text{and} \quad t_\lambda^{1-\alpha} I^\tau \subseteq 2^{-j\sigma(1-\alpha)} 2^{\sigma(1-\alpha)} I^\tau,$$

which yields

$$t_\lambda J^\tau \cap (0,1] \subseteq J_j^{+,\tau,\sigma} := 2^{-j\sigma}[2^{-\tau},2^{\tau+\sigma}) \cap (0,1].$$

Further, since $2^{-\lceil j\sigma(1-\alpha)\rceil} \le 2^{-j\sigma(1-\alpha)}$ and $\tilde\sigma\le\sigma$, it holds

$$\eta_\lambda \in 2^{-\lceil j\sigma(1-\alpha)\rceil}(\ell\tilde\sigma + [0,\tilde\sigma)) \subseteq 2^{-\lceil j\sigma(1-\alpha)\rceil}\ell\tilde\sigma + 2^{-j\sigma(1-\alpha)}[0,\sigma].$$

This implies

$$\begin{aligned}
\eta_\lambda + t_\lambda^{1-\alpha} I^\tau &\subseteq 2^{-\lceil j\sigma(1-\alpha)\rceil}\ell\tilde\sigma + 2^{-j\sigma(1-\alpha)}\big(2^{\sigma(1-\alpha)}[-\tau,\tau] + [0,\sigma]\big) \\
&\subseteq 2^{-\lceil j\sigma(1-\alpha)\rceil}\ell\tilde\sigma + 2^{-j\sigma(1-\alpha)}2^{\sigma(1-\alpha)}[-\tau,\tau+\sigma],
\end{aligned}$$

and hence

$$(\eta_\lambda + t_\lambda^{1-\alpha} I^\tau)_{2\pi} \subseteq I_{j,\ell}^{+,\tau,\sigma} := \big(2^{-\lceil j\sigma(1-\alpha)\rceil}\ell\tilde\sigma + 2^{-(j-1)\sigma(1-\alpha)}[-\tau,\tau+\sigma]\big)_{2\pi}.$$

Altogether, this proves for every $J\in\mathbb{N}_0^2$ and $(\theta,u)\in\mathbb{T}\times(0,1]$ the inequality

$$\sum_{\lambda\in\Lambda_J^{\alpha,\sigma}} |c_\lambda|^p \chi_{t_\lambda}^{\tau}(u)\chi_{\eta_\lambda,t_\lambda}^{\alpha,\tau}(\theta) \le \sum_{\lambda\in\Lambda_J^{\alpha,\sigma}} |c_\lambda|^p \chi_J^{+,\tau,\sigma}(\theta,u), \tag{4.47}$$

where $\chi_J^{+,\tau,\sigma}$ is the characteristic function of the set $I_{j,\ell}^{+,\tau,\sigma}\times J_j^{+,\tau,\sigma}$.

Similarly, we obtain an estimate from below, this time with $\sigma = \tau$ however. It holds

$$\sum_{\lambda \in \Lambda_J^{\alpha,\sigma}} |c_\lambda|^p \chi_{t_\lambda}^{\sigma}(u) \chi_{\eta_\lambda, t_\lambda}^{\alpha,\sigma}(\theta) \geq \sum_{\lambda \in \Lambda_J^{\alpha,\sigma}} |c_\lambda|^p \chi_J^{-,\sigma}(\theta, u) \tag{4.48}$$

with the characteristic function $\chi_J^{-,\sigma}$ of the set

$$I_{j,\ell}^{-,\sigma} \times J_j^{-,\sigma} := \Big( (2^{-\lceil j\sigma(1-\alpha) \rceil} \ell\tilde{\sigma} + 2^{-j\sigma(1-\alpha)}[0,\sigma))_{2\pi} \Big) \times \Big( 2^{-j\sigma}[1, 2^\sigma) \cap (0,1] \Big).$$

Indeed, we see that for $\lambda \in \Lambda_J^{\alpha,\sigma}$ and $J = (j,\ell) \in \mathbb{N}_0^2$,

$$t_\lambda J^\sigma \supseteq 2^{-j\sigma}[1, 2^\sigma) \quad \text{and} \quad t_\lambda^{1-\alpha} I^\sigma \supseteq 2^{-j\sigma(1-\alpha)} I^\sigma.$$

Further, as above,

$$\eta_\lambda \in 2^{-\lceil j\sigma(1-\alpha) \rceil}(\ell\tilde{\sigma} + [0, \tilde{\sigma})) \subseteq 2^{-\lceil j\sigma(1-\alpha) \rceil} \ell\tilde{\sigma} + 2^{-j\sigma(1-\alpha)}[0, \sigma].$$

Hence

$$\eta_\lambda + t_\lambda^{1-\alpha} I^\sigma \supseteq 2^{-\lceil j\sigma(1-\alpha) \rceil} \ell\tilde{\sigma} + 2^{-j\sigma(1-\alpha)}[0, \sigma],$$

and the estimate (4.48) follows.

In the sequel, it is essential that the functions $\chi_J^{+,\tau,\sigma}$ and $\chi_J^{-,\sigma}$ are merely dependent on the indices $J \in \mathbb{N} \times \mathbb{N}_0$, and not on the particular $\lambda \in \Lambda_J^{\alpha,\tau}$ any more. We calculate

$$\int_0^1 \int_0^{2\pi} \chi_J^{-,\sigma}(\theta, u) \frac{d\theta\, du}{u} \asymp \int_0^1 \int_0^{2\pi} \chi_J^{+,\tau,\sigma}(\theta, u) \frac{d\theta\, du}{u} \asymp 2^{-j\sigma(1-\alpha)}$$

for fixed $\tau, \sigma > 0$.

Using (4.47) and (4.48), we can then prove

$$\|\{c_\lambda\}_\lambda\|_{\text{hom}}^{(\tau)} \lesssim \Bigg( \sum_{J \in \mathbb{N} \times \mathbb{N}_0} \Big( \sum_{\lambda \in \Lambda_J^{\alpha,\sigma}} |\mathbf{v}_{p,q}^{\alpha,s}(\mathbf{x}_\lambda) c_\lambda|^p \Big)^{q/p} \Bigg)^{1/q} \lesssim \|\{c_\lambda\}_\lambda\|_{\text{hom}}^{(\sigma)}, \tag{4.49}$$

whereby the implicit constants are dependent on $\tau$ and $\sigma$.

For the first inequality, we plug (4.47) into (4.46) and obtain

$$
\begin{aligned}
\|\{c_\lambda\}_\lambda\|_{\mathrm{hom}}^{(\tau)} &\lesssim \Big(\int_0^1\int_0^{2\pi} u^{-sq}\Big(\sum_{J\in\mathbb{N}\times\mathbb{N}_0} 2^{-j\sigma(1+\alpha)}\chi_J^{+,\tau,\sigma}(\theta,u)\sum_{\lambda\in\Lambda_J^{\alpha,\sigma}}|c_\lambda|^p\Big)^{q/p}\frac{d\theta\,du}{u}\Big)^{1/q}\\
&\asymp \Big(\int_0^1\int_0^{2\pi} u^{-sq}\sum_{J\in\mathbb{N}\times\mathbb{N}_0} 2^{-j\sigma(1+\alpha)q/p}\chi_J^{+,\tau,\sigma}(\theta,u)\Big(\sum_{\lambda\in\Lambda_J^{\alpha,\sigma}}|c_\lambda|^p\Big)^{q/p}\frac{d\theta\,du}{u}\Big)^{1/q}\\
&\asymp \Big(\sum_{J\in\mathbb{N}\times\mathbb{N}_0} 2^{j\sigma sq}2^{-j\sigma(1+\alpha)q/p}\Big[\int_0^1\int_0^{2\pi}\chi_J^{+,\tau,\sigma}(\theta,u)\frac{d\theta\,du}{u}\Big]\Big(\sum_{\lambda\in\Lambda_J^{\alpha,\sigma}}|c_\lambda|^p\Big)^{q/p}\Big)^{1/q}\\
&\asymp \Big(\sum_{J\in\mathbb{N}\times\mathbb{N}_0} 2^{j\sigma sq}2^{-j\sigma(1+\alpha)q/p}2^{-j\sigma(1-\alpha)}\Big(\sum_{\lambda\in\Lambda_J^{\alpha,\sigma}}|c_\lambda|^p\Big)^{q/p}\Big)^{1/q}\\
&= \Big(\sum_{J\in\mathbb{N}\times\mathbb{N}_0}\Big(\sum_{\lambda\in\Lambda_J^{\alpha,\sigma}} 2^{j\sigma(s-(1+\alpha)/p-(1-\alpha)/q)p}|c_\lambda|^p\Big)^{q/p}\Big)^{1/q}\\
&\asymp \Big(\sum_{J\in\mathbb{N}\times\mathbb{N}_0}\Big(\sum_{\lambda\in\Lambda_J^{\alpha,\sigma}}|\mathbf{v}_{p,q}^{\alpha,s}(\mathbf{x}_\lambda)c_\lambda|^p\Big)^{q/p}\Big)^{1/q}.
\end{aligned}
$$

Plugging (4.48) into (4.46), we obtain the second inequality

$$
\begin{aligned}
\|\{c_\lambda\}_\lambda\|_{\mathrm{hom}}^{(\sigma)} &\gtrsim \Big(\int_0^1\int_0^{2\pi} u^{-sq}\Big(\sum_{J\in\mathbb{N}\times\mathbb{N}_0} 2^{-j\sigma(1+\alpha)}\chi_J^{-,\sigma}(\theta,u)\sum_{\lambda\in\Lambda_J^{\alpha,\sigma}}|c_\lambda|^p\Big)^{q/p}\frac{d\theta\,du}{u}\Big)^{1/q}\\
&= \Big(\int_0^1\int_0^{2\pi} u^{-sq}\sum_{J\in\mathbb{N}\times\mathbb{N}_0} 2^{-j\sigma(1+\alpha)q/p}\chi_J^{-,\sigma}(\theta,u)\Big(\sum_{\lambda\in\Lambda_J^{\alpha,\sigma}}|c_\lambda|^p\Big)^{q/p}\frac{d\theta\,du}{u}\Big)^{1/q}\\
&\asymp \Big(\sum_{J\in\mathbb{N}\times\mathbb{N}_0} 2^{j\sigma sq}2^{-j\sigma(1+\alpha)q/p}\Big[\int_0^1\int_0^{2\pi}\chi_J^{-,\sigma}(\theta,u)\frac{d\theta\,du}{u}\Big]\Big(\sum_{\lambda\in\Lambda_J^{\alpha,\sigma}}|c_\lambda|^p\Big)^{q/p}\Big)^{1/q}\\
&\asymp \Big(\sum_{J\in\mathbb{N}\times\mathbb{N}_0} 2^{j\sigma sq}2^{-j\sigma(1+\alpha)q/p}2^{-j\sigma(1-\alpha)}\Big(\sum_{\lambda\in\Lambda_J^{\alpha,\sigma}}|c_\lambda|^p\Big)^{q/p}\Big)^{1/q}\\
&\asymp \Big(\sum_{J\in\mathbb{N}\times\mathbb{N}_0}\Big(\sum_{\lambda\in\Lambda_J^{\alpha,\sigma}}|\mathbf{v}_{p,q}^{\alpha,s}(\mathbf{x}_\lambda)c_\lambda|^p\Big)^{q/p}\Big)^{1/q}.
\end{aligned}
$$

For arbitrary $\tau,\sigma>0$ we can finally deduce, using the symmetry of (4.49),

$$\|\{c_\lambda\}_\lambda\|_{\mathrm{hom}}^{(\tau)} \asymp \|\{c_\lambda\}_\lambda\|_{\mathrm{hom}}^{(\sigma)},$$

with implicit constants depending on those parameters. A similar estimate as (4.49) holds true for the inhomogeneous parts, as well as $\|\{c_\lambda\}_\lambda\|_{\mathrm{in}}^{(\tau)} \asymp \|\{c_\lambda\}_\lambda\|_{\mathrm{in}}^{(\sigma)}$.

This shows that for every $\tau>0$ the spaces $(L_{p,q}^s)^\flat\langle\mathcal{U}_\tau^\alpha[\mathcal{P}_\Lambda]\rangle$ contain the same sequences. Furthermore, the quasi-norms of these spaces are all equivalent. Finally, choosing $\tau=\sigma=1$ in (4.49) and the analogue for the inhomogeneous part yields the assertion. □

For the Wiener spaces $\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$ we can derive the same associated sequence spaces.

**Corollary 4.4.13.** *Under the same assumptions as in Theorem 4.4.12, we have*

$$\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})^\flat\langle\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]\rangle \asymp \ell^{\alpha,s}_{p,q}[\mathcal{P}_\Lambda].$$

*Proof.* In view of Lemma 4.2.5(v) (see also (4.20)), we have with $\sigma := h(\tau,1)$

$$\mathbf{W}^\alpha \mathcal{X}_{U^\alpha_\tau(\mathbf{x}_\lambda)} \le \mathcal{X}_{U^\alpha_\sigma(\mathbf{x}_\lambda)}$$

for all $\mathbf{x}_\lambda \in \mathcal{P}_\Lambda$. Hence we can estimate

$$\begin{aligned}\|\{c_\lambda\}|(L^s_{p,q})^\flat\langle\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]\rangle\| &\le \|\{c_\lambda\}|(\mathbb{L}^{\alpha,s}_{p,q})^\flat\langle\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]\rangle\| = \|\sum_\lambda |c_\lambda|\mathcal{X}_{U^\alpha_\tau(\mathbf{x}_\lambda)}|\mathbb{L}^{\alpha,s}_{p,q}\| \\ &\le \|\sum_\lambda |c_\lambda|\mathcal{X}_{U^\alpha_\sigma(\mathbf{x}_\lambda)}|L^s_{p,q}\| = \|\{c_\lambda\}|(L^s_{p,q})^\flat\langle\mathcal{U}^\alpha_\sigma[\mathcal{P}_\Lambda]\rangle\|.\end{aligned}$$

Since the quasi-norms on the left- and right-hand side are equivalent, according to Theorem 4.4.12, the assertion follows. □

As the last result of this subsection, we prove that the definition of the spaces $\ell^{\alpha,s}_{p,q}[\mathcal{P}_\Lambda]$ is rather robust with respect to the utilized sampling $\mathcal{P}_\Lambda$.

**Corollary 4.4.14.** *Let $\mathcal{P}_\Lambda = \{\mathbf{x}_\lambda\}_{\lambda\in\Lambda}$ be an $\alpha$-well-spread family of points in $\mathbb{X}$, and assume that $\mathcal{S} : \Lambda \to \mathbb{X}$ is a sampling subject to the condition $\mathcal{S}(\lambda) \in U^\alpha_\tau(\mathbf{x}_\lambda)$ for all $\lambda \in \Lambda$, where $\tau > 0$ is arbitrary but fixed. Then $\ell^{\alpha,s}_{p,q}[\mathcal{P}_\Lambda] \asymp \ell^{\alpha,s}_{p,q}[\mathcal{S}_\Lambda]$.*

*Proof.* Using Lemma 4.2.5(iii) and (iv), we obtain

$$U^\alpha_\tau(\mathbf{x}_\lambda) \subseteq U^\alpha_\sigma(S(\lambda)) \subseteq U^\alpha_{\tau'}(\mathbf{x}_\lambda)$$

with $\sigma := g(\tau,\tau)$ and $\tau' := f(\tau,\sigma)$. With the help of Theorem 4.4.12, we then deduce

$$\ell^{\alpha,s}_{p,q}[\mathcal{P}_\Lambda] \asymp \big(L^s_{p,q}(\mathbb{X})\big)^\flat\langle\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]\rangle \lesssim \big(L^s_{p,q}(\mathbb{X})\big)^\flat\langle\mathcal{U}^\alpha_\sigma[\mathcal{S}_\Lambda]\rangle \asymp \ell^{\alpha,s}_{p,q}[\mathcal{S}_\Lambda],$$

and the opposite direction

$$\ell^{\alpha,s}_{p,q}[\mathcal{S}_\Lambda] \asymp \big(L^s_{p,q}(\mathbb{X})\big)^\flat\langle\mathcal{U}^\alpha_\sigma[\mathcal{S}_\Lambda]\rangle \lesssim \big(L^s_{p,q}(\mathbb{X})\big)^\flat\langle\mathcal{U}^\alpha_{\tau'}[\mathcal{P}_\Lambda]\rangle \asymp \ell^{\alpha,s}_{p,q}[\mathcal{P}_\Lambda].$$

The proof is finished. □

### 4.4.3 Discrete Characterizations: Atomic Decompositions and Quasi-Banach Frames

In this subsection we will finally apply the abstract discretization results [74, Thm. 2.48] and [74, Thm. 2.50] to our specific coorbit space setting, leading to Theorems 4.4.19 and 4.4.21 below. In order to do this we need some preparation. Let us start with some definitions.

**Definition 4.4.15** (compare [110, Def. 3.9])**.** Suppose $Y$ is a quasi-Banach space. A family $\{h_\lambda\}_{\lambda\in\Lambda}$ of bounded linear functionals on $Y$ is called a *quasi-Banach frame* for $Y$, if there exists a QBS-space $Y^\flat = Y^\flat(\Lambda)$ and a bounded linear reconstruction operator $\Omega : Y^\flat \to Y$ such that the following holds true,

i) The associated analysis operator $H : f \mapsto \{h_\lambda(f)\}_{\lambda\in\Lambda}$ is bounded from $Y$ to $Y^\flat$,

ii) It holds $\Omega(H(f)) = f$ for all $f \in Y$.

Note that if $\{h_\lambda\}_{\lambda\in\Lambda}$ is a quasi-Banach frame for $Y$, then there exist *frame bounds* $0 < C_1 \le C_2 \le \infty$ such that

$$C_1\|f|Y\| \le \|H(f)|Y^\flat\| \le C_2\|f|Y\|.$$

A somewhat dual notion to a frame is the notion of an atomic decomposition.

**Definition 4.4.16** (compare [110, Def. 3.8])**.** A family $\{g_\lambda\}_{\lambda\in\Lambda}$ in a quasi-Banach space $Y$ is called an *atomic decomposition* for $Y$, if there exists a quasi-Banach frame $\{h_\lambda\}_{\lambda\in\Lambda}$ for $Y$ with associated QBS-space $Y^\flat$ such that:

i) The associated synthesis operator $\Omega : \{c_\lambda\}_{\lambda\in\Lambda} \mapsto \sum_{\lambda\in\Lambda} c_\lambda g_\lambda$ is bounded from $Y^\flat$ to $Y$,

ii) The reconstruction formula $f = \sum_{\lambda\in\Lambda} h_\lambda(f) g_\lambda$ holds true for all $f \in Y$.

### Frame Sampling

Our first discretization result, Theorem 4.4.19, yields atomic decompositions and discrete quasi-Banach frames for the coorbit space $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q})$ by suitably sampling the continuous $\alpha$-curvelet frame $\mathfrak{C}_\alpha$. Its proof is based on the analysis of the so-called oscillation kernel associated to $\mathfrak{C}_\alpha$.

**Definition 4.4.17** ([74])**.** Let $\mathcal{U} = \{U_\lambda\}_{\lambda\in\Lambda}$ be an admissible covering of $\mathbb{X}$ and let $\Gamma : \mathbb{X}\times\mathbb{X} \to \mathbb{S}^1$ be a phase function. We define the *oscillation kernel* associated to $\mathcal{U}$ and the $\alpha$-curvelet frame $\mathfrak{C}_\alpha = \{\psi_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ from Section 3.1 by

$$\mathrm{osc}_{\mathcal{U},\Gamma}(\mathbf{x},\mathbf{y}) := \sup_{\mathbf{z}\in U_{\mathbf{y}}} |\mathcal{G}[\mathfrak{C}_\alpha](\mathbf{x},\mathbf{y}) - \Gamma(\mathbf{y},\mathbf{z})\mathcal{G}[\mathfrak{C}_\alpha](\mathbf{x},\mathbf{z})| = \sup_{\mathbf{z}\in U_{\mathbf{y}}} |\langle \psi_{\mathbf{x}}, \psi_{\mathbf{y}} - \Gamma(\mathbf{y},\mathbf{z})\psi_{\mathbf{z}}\rangle|,$$

where $\mathbf{x},\mathbf{y} \in \mathbb{X}$ and $U_{\mathbf{y}} := \bigcup_{\lambda:\mathbf{y}\in U_\lambda} U_\lambda$. Further, we put $\mathrm{osc}^*_{\mathcal{U},\Gamma}(\mathbf{x},\mathbf{y}) := \mathrm{osc}_{\mathcal{U},\Gamma}(\mathbf{y},\mathbf{x})$.

We next want to apply [74, Thm. 2.48]. For this, we need to verify Property $D(\delta,\nu,Y)$ for $\mathfrak{C}_\alpha$. The following definition is in line with [74, Def. 2.43].

**Definition 4.4.18** (compare [74, Def. 2.43])**.** Let $Y$ be a rich solid QBF-space on $\mathbb{X}$ and let $\mathfrak{C}_\alpha$ be the Parseval frame of $\alpha$-curvelets from Section 3.1. We say that $\mathfrak{C}_\alpha$ possesses Property $D(\delta,\nu,Y)$ for a weight $\nu \ge 1$ and some $\delta > 0$ if it has Property $F(\nu,Y)$(see Definition 4.3.6) and if there exists an admissible covering $\mathcal{U}$ and a phase function $\Gamma : X\times X \to \mathbb{S}^1$ so that

(i) $|\mathcal{G}[\mathfrak{C}_\alpha]|$, $\mathrm{osc}_{\mathcal{U},\Gamma}$, $\mathrm{osc}^*_{\mathcal{U},\Gamma} \in \mathcal{B}_{m_\nu,Y}$.

(ii) $\|\mathrm{osc}_{\mathcal{U},\Gamma}|\mathcal{B}_{m_\nu,Y}\| < \delta$ and $\|\mathrm{osc}^*_{\mathcal{U},\Gamma}|\mathcal{B}_{m_\nu,Y}\| < \delta$.

Now we are ready to apply [74, Thm. 2.48]. We obtain the theorem below.

**Theorem 4.4.19.** *Let $\alpha \in [0,1]$ and let $\mathfrak{C}_\alpha := \{\psi_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ be the continuous Parseval frame of $\alpha$-curvelets constructed in Section 3.1. Further, let $s \in \mathbb{R}$ and $0 < p,q < \infty$ be fixed. Then there exists $\tau_0 = \tau_0[\alpha,s,p,q] > 0$ such that for every $\alpha$-well-spread sampling $\mathcal{P}_\Lambda$ of $\mathbb{X}$ of density $\tau \le \tau_0$ the sampled system $\{\psi_\lambda := \psi_{\mathbf{x}_\lambda} : \lambda \in \Lambda\}$ is a discrete $\alpha$-curvelet frame for $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q})$ which possesses a corresponding dual frame $\{\tilde{\psi}_\lambda : \lambda \in \Lambda\}$ such that:*

*i) (Analysis) For $f \in (\mathcal{H}^\nu_1)^\urcorner$ with $\nu = \nu^{\alpha,s}_{p,q}$ as in (4.35) we have*

$$f \in \mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q}) \quad \Leftrightarrow \quad \{\langle f,\psi_\lambda\rangle\}_\lambda \in \ell^{\alpha,s}_{p,q}[\mathcal{P}_\Lambda] \quad \Leftrightarrow \quad \{\langle f,\tilde{\psi}_\lambda\rangle\}_\lambda \in \ell^{\alpha,s}_{p,q}[\mathcal{P}_\Lambda].$$

*In case $f \in \mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q})$ the quasi-norms are equivalent, i.e.,*

$$\|f|\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q})\| \asymp \|\{\langle f,\psi_\lambda\rangle\}_\lambda|\ell^{\alpha,s}_{p,q}[\mathcal{P}_\Lambda]\| \asymp \|\{\langle f,\tilde{\psi}_\lambda\rangle\}_\lambda|\ell^{\alpha,s}_{p,q}[\mathcal{P}_\Lambda]\| .$$

*ii) (Synthesis) For each sequence $\{c_\lambda\}_\lambda \in \ell^{\alpha,s}_{p,q}[\mathcal{P}_\Lambda]$ the sums*

$$\sum_{\lambda\in\Lambda} c_\lambda \psi_\lambda \quad \textit{and} \quad \sum_{\lambda\in\Lambda} c_\lambda \tilde{\psi}_\lambda$$

*converge unconditionally in $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q})$. Moreover, the assigned synthesis operators are bounded from $\ell^{\alpha,s}_{p,q}[\mathcal{P}_\Lambda]$ to $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q})$.*

*iii) (Reconstruction) For all $f \in \mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}^{\alpha,s}_{p,q})$ we have*

$$f = \sum_{\lambda\in\Lambda} \langle f,\psi_\lambda\rangle \tilde{\psi}_\lambda = \sum_{\lambda\in\Lambda} \langle f,\tilde{\psi}_\lambda\rangle \psi_\lambda .$$

*Proof.* We have shown in Proposition 4.2.9 that $Y := \mathbb{L}^{\alpha,s}_{p,q}$ is a rich solid QBF-space on $\mathbb{X}$. Let $C_Y$ denote the associated quasi-norm constant and $\nu := \nu^{\alpha,s}_{p,q}$ the associated weight defined in (4.35). We know from Proposition 3.1.3 and Theorem 4.5.5 that $\mathcal{G}[\mathfrak{C}_\alpha] \in \mathcal{B}_{m_\nu,Y}$. Hence, we can choose a number $\delta = \delta[\alpha,s,p,q] > 0$ which satisfies

$$\delta\big((1+C_Y)\||\mathcal{G}[\mathfrak{C}_\alpha]||\mathcal{B}_{m_\nu,Y}\| + \delta C_Y\big)C_Y \le 1,$$

the condition required in [74, Thm. 2.48].

Now assume that $\mathcal{P}_\Lambda : \Lambda \to \mathbb{X}$ defines an $\alpha$-well-spread and $(\alpha,\tau)$-dense point family $\{\mathbf{x}_\lambda\}_{\lambda\in\Lambda}$ in $\mathbb{X}$. According to Proposition 4.4.8, the collection $\mathcal{U} := \mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]$ is then an admissible covering. Further, due to Lemma 4.2.5(iv), if $\sigma \ge g(\tau,\tau)$ we have

$$U_{\mathbf{y}} := \bigcup_{\lambda:\mathbf{y}\in U^\alpha_\tau(\mathbf{x}_\lambda)} U^\alpha_\tau(\mathbf{x}_\lambda) \subseteq U^\alpha_\sigma(\mathbf{y}) \quad \text{for all } \mathbf{y} \in \mathbb{X}.$$

It follows, with $\Gamma \equiv 1$, that $\operatorname{osc}_{\mathcal{U},\Gamma}(\mathbf{x},\mathbf{y}) \le \operatorname{osc}_\sigma(\mathbf{x},\mathbf{y})$ and $\operatorname{osc}^*_{\mathcal{U},\Gamma}(\mathbf{x},\mathbf{y}) \le \operatorname{osc}^*_\sigma(\mathbf{x},\mathbf{y})$, where $\operatorname{osc}_\sigma$ is the oscillation kernel from Definition 4.5.6.

According to Theorem 4.5.7, there further exists $\sigma_0 > 0$ such that

$$\|\operatorname{osc}_\sigma|\mathcal{B}_{m_\nu,Y}\| < \delta \quad \text{and} \quad \|\operatorname{osc}^*_\sigma|\mathcal{B}_{m_\nu,Y}\| < \delta \quad \text{for all } \sigma \le \sigma_0.$$

We now choose $\tau_0 > 0$ such that $\sigma_0 \ge g(\tau_0,\tau_0)$. Then for all $\tau \le \tau_0$ and $\mathcal{U} = \mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]$

$$\|\operatorname{osc}_{\mathcal{U},\Gamma}|\mathcal{B}_{m_\nu,Y}\| < \delta \quad \text{and} \quad \|\operatorname{osc}^*_{\mathcal{U},\Gamma}|\mathcal{B}_{m_\nu,Y}\| < \delta.$$

This shows that $\mathfrak{C}_\alpha$ possesses Property $D(\delta,\nu,Y)$ with respect to $\mathcal{U} = \mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]$ if $\tau \le \tau_0$. Hence we can apply [74, Thm. 2.48] with sampling points $\mathbf{x}_\lambda \in U_\lambda = U^\alpha_\tau(\mathbf{x}_\lambda)$. Finally, note that it holds $\ell^{\alpha,s}_{p,q}[\mathcal{P}_\Lambda] \asymp \mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})^\flat\langle\mathcal{U}^\alpha_\tau[\mathcal{P}_\Lambda]\rangle$ due to Corollary 4.4.13. Taking into account Corollary 4.2.12 and the fact that the finite sequences are dense in $\ell^{\alpha,s}_{p,q}[\mathcal{P}_\Lambda]$, since $p,q < \infty$, the assertions follow. ☐

**Frame Expansion**

Another useful discretization result of the abstract coorbit theory is [74, Thm. 2.50]. Subsequently, we consider a discrete frame $\mathfrak{M}_\alpha^\bullet = \{m_\lambda\}_{\lambda\in\Lambda}$ of $\alpha$-molecules, indexed by some countable index set $\Lambda$, such that the associated parametrization $\Phi : \Lambda \to \mathbb{X}$ yields an $\alpha$-well-spread family $\Phi_\Lambda$ of points in $\mathbb{X}$. In Theorem 4.4.21, we will derive a sufficient condition, using the abstract result [74, Thm. 2.50], when such a frame constitutes a quasi-Banach frame as well as an atomic decomposition for the coorbit $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})$.

The proof is based on the analysis of so-called cross-Gramian maximal kernels defined as follows.

**Definition 4.4.20** (compare [74, eq. (2.13)])**.** Let $\mathfrak{M}_\alpha = \{m_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ and $\widetilde{\mathfrak{M}}_\alpha = \{\tilde{m}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ be two systems of $\alpha$-molecules of order $(L, M, N_1, N_2)$ with respect to the canonical parametrization. Let $\mathcal{U} = \{U_\lambda\}_{\lambda\in\Lambda}$ be an admissible covering of $\mathbb{X}$. The associated *cross-Gramian maximal kernel* is defined by

$$\mathcal{M}_{\mathcal{U}}[\widetilde{\mathfrak{M}}_\alpha, \mathfrak{M}_\alpha](\mathbf{x}, \mathbf{y}) := \sup_{\mathbf{z}\in U_{\mathbf{y}}} |\mathcal{G}[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha](\mathbf{x}, \mathbf{z})| = \sup_{\mathbf{z}\in U_{\mathbf{y}}} |\langle m_{\mathbf{x}}, \tilde{m}_{\mathbf{z}}\rangle|, \quad \mathbf{x}, \mathbf{y} \in \mathbb{X}.$$

where $U_{\mathbf{y}} := \bigcup_{\lambda:\mathbf{y}\in U_\lambda} U_\lambda$. Further, we put $\mathcal{M}_{\mathcal{U}}[\widetilde{\mathfrak{M}}_\alpha, \mathfrak{M}_\alpha]^*(\mathbf{x}, \mathbf{y}) := \mathcal{M}_{\mathcal{U}}[\widetilde{\mathfrak{M}}_\alpha, \mathfrak{M}_\alpha](\mathbf{y}, \mathbf{x})$.

Now we can formulate the next theorem and prove it with [74, Thm. 2.50] (see also [111, Thm. 3.14]). Recall that $\mathfrak{C}_\alpha = \{\psi_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ denotes the continuous Parseval frame of $\alpha$-curvelets from Section 3.1.

**Theorem 4.4.21.** *Let $\alpha \in [0, 1]$ and let $\mathfrak{M}_\alpha^\bullet = \{m_\lambda\}_{\lambda\in\Lambda}$ and $\widetilde{\mathfrak{M}}_\alpha^\bullet = \{\tilde{m}_\lambda\}_{\lambda\in\Lambda}$ be two discrete $\alpha$-molecule frames in $L_2(\mathbb{R}^2)$ of order $(L, M, N_1, N_2)$ and with respective parametrizations $\Phi : \Lambda \to \mathbb{X}$ and $\tilde{\Phi} : \Lambda \to \mathbb{X}$. Further, assume that $\mathfrak{M}_\alpha^\bullet$ and $\widetilde{\mathfrak{M}}_\alpha^\bullet$ are dual to each other, i.e.,*

$$f = \sum_{\lambda\in\Lambda} \langle f, \tilde{m}_\lambda\rangle m_\lambda = \sum_{\lambda\in\Lambda} \langle f, m_\lambda\rangle \tilde{m}_\lambda \quad \textit{for all } f \in L_2(\mathbb{R}^2). \tag{4.50}$$

*Provided that the following conditions are fulfilled:*

*i) There exist $\tau > 0$ and an $\alpha$-well-spread point family $\mathcal{P}_\Lambda = \{\mathbf{x}_\lambda\}_{\lambda\in\Lambda}$ in $\mathbb{X}$ such that*

$$\Phi(\lambda) \in U_\tau^\alpha(\mathbf{x}_\lambda) \quad \textit{and} \quad \tilde{\Phi}(\lambda) \in U_\tau^\alpha(\mathbf{x}_\lambda) \quad \textit{for all } \lambda \in \Lambda,$$

*ii) The order $(L, M, N_1, N_2)$ of $\mathfrak{M}_\alpha^\bullet$ and $\widetilde{\mathfrak{M}}_\alpha^\bullet$ satisfies condition (4.36) with respect to $s \in \mathbb{R}$ and $0 < p, q < \infty$,*

*then both frames $\mathfrak{M}_\alpha^\bullet$ and $\widetilde{\mathfrak{M}}_\alpha^\bullet$ are contained in $\mathcal{H}_1^\nu$, where $\nu := \nu_{p,q}^{\alpha,s}$ as in (4.35), and the expansion (4.50) is valid for all $f \in \mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})$ with quasi-norm convergence in $\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})$. Here $\langle\cdot,\cdot\rangle$ must be interpreted as the duality product of the pairing $(\mathcal{H}_1^\nu)^\urcorner \times \mathcal{H}_1^\nu$.*

*Furthermore, for every $f \in (\mathcal{H}_1^\nu)^\urcorner$ we have*

$$f \in \mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s}) \quad \Leftrightarrow \quad \{\langle f, m_\lambda\rangle\}_\lambda \in \ell_{p,q}^{\alpha,s}[\mathcal{P}_\Lambda] \quad \Leftrightarrow \quad \{\langle f, \tilde{m}_\lambda\rangle\}_\lambda \in \ell_{p,q}^{\alpha,s}[\mathcal{P}_\Lambda],$$

*and – in case $f \in \mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})$ – it holds*

$$\|f|\mathsf{Co}(\mathfrak{C}_\alpha, \mathbb{L}_{p,q}^{\alpha,s})\| \asymp \|\{\langle f, m_\lambda\rangle\}_\lambda|\ell_{p,q}^{\alpha,s}[\mathcal{P}_\Lambda]\| \asymp \|\{\langle f, \tilde{m}_\lambda\rangle\}_\lambda|\ell_{p,q}^{\alpha,s}[\mathcal{P}_\Lambda]\|.$$

*Proof.* We want to apply [74, Thm. 2.50]. To this end, note that according to Lemma 4.4.6 the intersection number $N$ of $\mathcal{U}_\tau^\alpha[\mathcal{P}_\Lambda]$ is finite, where $\tau > 0$ stems from condition (i). By Lemma 4.4.2 we can thus split the family $\mathcal{P}_\Lambda$ into at most $N$ subfamilies $\mathcal{P}_{\Lambda^k}$, where $k \in \{1, \dots, r\}$ and $r \leq N$, which are each $(\alpha, \tau)$-separated. The associated splitting of the index set $\Lambda$ shall be denoted by $\Lambda^k$. We obtain corresponding collections $\mathcal{U}_\tau^\alpha[\mathcal{P}_{\Lambda^k}]$, $k \in \{1, \dots, r\}$, consisting of pairwise disjoint sets.

Next, we build $r$ continuous families of $\alpha$-molecules. We define for $k \in \{1, \dots, r\}$:

$$\mathfrak{M}_\alpha^{(k)} := \{m_{\mathbf{x}}^{(k)} : \mathbf{x} \in \mathbb{X}\} \quad \text{with} \quad m_{\mathbf{x}}^{(k)} := \begin{cases} m_\lambda & \text{if } \mathbf{x} = \Phi(\lambda), \lambda \in \Lambda^k, \\ 0 & \text{else,} \end{cases}$$

$$\tilde{\mathfrak{M}}_\alpha^{(k)} := \{\tilde{m}_{\mathbf{x}}^{(k)} : \mathbf{x} \in \mathbb{X}\} \quad \text{with} \quad \tilde{m}_{\mathbf{x}}^{(k)} := \begin{cases} \tilde{m}_\lambda & \text{if } \mathbf{x} = \tilde{\Phi}(\lambda), \lambda \in \Lambda^k, \\ 0 & \text{else.} \end{cases}$$

By definition, these families are nonzero merely on a discrete subset of $\mathbb{X}$, namely the sampling points determined by $\Phi$ and $\tilde{\Phi}$. Further, they are systems of $\alpha$-molecules with respect to the canonical parametrization of the same order $(L, M, N_1, N_2)$ as $\mathfrak{M}_\alpha^\bullet$ and $\widetilde{\mathfrak{M}}_\alpha^\bullet$.

We proceed by extending each point family $\mathcal{P}_{\Lambda^k}$ to an $\alpha$-well-spread family $\mathcal{P}_k^{\text{ext}}$ that is still $(\alpha, \tau)$-separated. This is possible by Lemma 4.4.5. More concretely, looking into the proof of this lemma, we can assume $\mathcal{P}_k^{\text{ext}}$ to be $(\alpha, \sigma)$-dense for some $\sigma \geq h(\tau, \tau)$, where $h$ is the function from Lemma 4.2.5(v).

Let us denote the extended index sets by $\Lambda_k^{\text{ext}}$, and for the extended point families $\mathcal{P}_k^{\text{ext}}$ let us write $\mathcal{P}_k^{\text{ext}} = \{\mathbf{x}_\lambda^k : \lambda \in \Lambda_k^{\text{ext}}\}$ with $\mathbf{x}_\lambda^k := \mathbf{x}_\lambda$ for $\lambda \in \Lambda^k$. By Proposition 4.4.8, we obtain associated admissible coverings $\mathcal{U}_k := \mathcal{U}_\sigma^\alpha[\mathcal{P}_k^{\text{ext}}] = \{U_\lambda^k : \lambda \in \Lambda_k^{\text{ext}}\}$ where $U_\lambda^k := U_\sigma^\alpha(\mathbf{x}_\lambda^k)$. Furthermore, these coverings are moderate (see [46, page 260] for a definition) since $\mu(U_\lambda^k) \asymp 1$ by Corollary 4.2.12. In particular, [74, eq. 2.25] is satisfied.

In view of Lemma 4.2.5(iv), if $\rho \geq g(\sigma, \sigma)$ we have

$$U_{\mathbf{y}}^k := \bigcup_{\lambda : \mathbf{y} \in U_\lambda^k} U_\lambda^k \subseteq U_\rho^\alpha(\mathbf{y}) \quad \text{for all } \mathbf{y} \in \mathbb{X}.$$

Hence, we can define and estimate the kernel functions $\mathcal{K}_k$ and $\tilde{\mathcal{K}}_k$, $k \in \{1, \dots, r\}$, as follows,

$$\mathcal{K}_k(\mathbf{x}, \mathbf{y}) := \mathcal{M}_{\mathcal{U}_k}[\mathfrak{M}_\alpha^{(k)}, \mathfrak{C}_\alpha](\mathbf{x}, \mathbf{y}) \leq \mathcal{M}_\rho^\alpha[\mathfrak{M}_\alpha^{(k)}, \mathfrak{C}_\alpha](\mathbf{x}, \mathbf{y}),$$
$$\tilde{\mathcal{K}}_k(\mathbf{x}, \mathbf{y}) := \mathcal{M}_{\mathcal{U}_k}[\tilde{\mathfrak{M}}_\alpha^{(k)}, \mathfrak{C}_\alpha](\mathbf{x}, \mathbf{y}) \leq \mathcal{M}_\rho^\alpha[\tilde{\mathfrak{M}}_\alpha^{(k)}, \mathfrak{C}_\alpha](\mathbf{x}, \mathbf{y}),$$

with $\mathcal{M}_\rho^\alpha[\mathfrak{M}_\alpha^{(k)}, \mathfrak{C}_\alpha]$ and $\mathcal{M}_\rho^\alpha[\tilde{\mathfrak{M}}_\alpha^{(k)}, \mathfrak{C}_\alpha]$ being defined as in (4.54).

According to Theorem 4.5.5 and due to condition (ii), these kernel functions as well as their involutions $\mathcal{M}_\rho^\alpha[\mathfrak{M}_\alpha^{(k)}, \mathfrak{C}_\alpha]^*$ and $\mathcal{M}_\rho^\alpha[\tilde{\mathfrak{M}}_\alpha^{(k)}, \mathfrak{C}_\alpha]^*$, are elements of the algebra $\mathcal{B}_{m_\nu, Y}$ for $\nu := \nu_{p,q}^{\alpha,s}$ and $Y := \mathbb{L}_{p,q}^{\alpha,s}$. By solidity, the kernels $\mathcal{K}_k$, $\tilde{\mathcal{K}}_k$, $\mathcal{K}_k^*$, and $\tilde{\mathcal{K}}_k^*$ thus also belong to $\mathcal{B}_{m_\nu, Y}$.

We now choose for each $k \in \{1, \dots, r\}$ and each $\lambda \in \Lambda_k^{\text{ext}}$ sampling points $\mathbf{x}_{k,\lambda} \in U_\lambda^k$ and $\tilde{\mathbf{x}}_{k,\lambda} \in U_\lambda^k$ as follows:

$$\begin{aligned} &\mathbf{x}_{k,\lambda} := \Phi(\lambda) \text{ and } \tilde{\mathbf{x}}_{k,\lambda} := \tilde{\Phi}(\lambda) && \text{for } \lambda \in \Lambda^k, \\ &\mathbf{x}_{k,\lambda} := \tilde{\mathbf{x}}_{k,\lambda} \in U_\tau^\alpha(\mathbf{x}_\lambda^k) \text{ arbitrary} && \text{for } \lambda \in \Lambda_k^{\text{ext}} \backslash \Lambda^k. \end{aligned}$$

Then we get from (4.50) the validity of

$$f = \sum_{k=1}^{r} \sum_{\lambda \in \Lambda_k^{\text{ext}}} \langle f, \tilde{m}^{(k)}_{\tilde{\mathbf{x}}_{k,\lambda}} \rangle m^{(k)}_{\mathbf{x}_{k,\lambda}} = \sum_{k=1}^{r} \sum_{\lambda \in \Lambda_k^{\text{ext}}} \langle f, m^{(k)}_{\mathbf{x}_{k,\lambda}} \rangle \tilde{m}^{(k)}_{\tilde{\mathbf{x}}_{k,\lambda}} \quad \text{for all } f \in L_2(\mathbb{R}^2).$$

Hence, the prerequisites to apply [74, Thm. 2.50] are (almost) fulfilled, albeit not precisely since we have used possibly different coverings $\mathcal{U}_k$ for each pair of kernels $\mathcal{K}_k$ and $\tilde{\mathcal{K}}_k$. Moreover, the primal and dual sampling points $\mathbf{x}_{k,\lambda} \in U^k_\lambda$ and $\tilde{\mathbf{x}}_{k,\lambda} \in U^k_\lambda$ might not coincide and may also differ for different $k$.

Hence, we need to apply the original theorem [74, Thm. 2.50] in a slightly generalized form. This is possible, since revisiting the proof of [74, Thm. 2.50] (see [111, Thm. 3.14]), it becomes clear that the statement of [74, Thm. 2.50] still holds true under these generalized assumptions. An application thus yields the assertion. To see this, we remark that as a consequence of Corollary 4.4.13

$$\sum_{k=1}^{r} \|\{\langle f, m^{(k)}_{\mathbf{x}_{k,\lambda}} \rangle\}_{\lambda \in \Lambda^k} | \mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})^\flat \langle \mathcal{U}_k \rangle\| \asymp \|\{\langle f, m_\lambda \rangle\}_{\lambda \in \Lambda} | \ell^{\alpha,s}_{p,q}[\mathcal{P}_\Lambda]\|.$$

Also observe that the finite sequences are dense in $\ell^{\alpha,s}_{p,q}[\mathcal{P}_\Lambda]$. □

Note that the frames $\mathfrak{M}^\bullet_\alpha$ and $\widetilde{\mathfrak{M}}^\bullet_\alpha$ in Theorem 4.4.21 need not coincide, which extends the range of applicability of the result significantly. For example, frames of compactly supported shearlets, where no tight frame constructions are known, might be possible choices for $\mathfrak{M}^\bullet_\alpha$. A drawback for the application however is the required knowledge on the dual frame, which is not available for many concrete constructions.

## 4.5 Appendix: *Kernel Analysis*

In this appendix the technical details are provided needed for the application of the abstract theory of coorbit spaces from [74] to our concrete setting of $\alpha$-molecule coorbits. The abstract theory relies heavily on the analysis of certain kernel functions and their mapping properties. In the following exposition we pursue the required analysis for the specific kernels associated with the continuous $\alpha$-molecule transform.

Thereby, in our concrete setting, a *kernel function*, or a *kernel* for short, refers to a measurable function $K : \mathbb{X} \times \mathbb{X} \to \mathbb{C}$. For convenience, we collect all such kernels in the set

$$\boldsymbol{\mathcal{K}} := \Big\{ K : \mathbb{X} \times \mathbb{X} \to \mathbb{C} \ : \ K \ (\mu \otimes \mu)\text{-measurable} \Big\}$$

and identify those which coincide apart from a null set. Clearly, equipped with point-wise addition and scalar multiplication, $\boldsymbol{\mathcal{K}}$ is a $\mathbb{C}$-linear space. Moreover, this space is closed under *involution*, i.e., the operation given by

$$K \mapsto K^* \quad \text{where} \quad K^*(\mathbf{x}, \mathbf{y}) := \overline{K(\mathbf{y}, \mathbf{x})}\,, \quad \mathbf{x}, \mathbf{y} \in \mathbb{X}.$$

The significance of the kernel space $\boldsymbol{\mathcal{K}}$ stems from the fact that its elements naturally act on functions with domain $\mathbb{X}$. This shall be understood in the following sense: A kernel $K \in \boldsymbol{\mathcal{K}}$ maps a measurable function $F : \mathbb{X} \to \mathbb{C}$ to a function $KF := K[F]$ via

$$K[F](\mathbf{x}) := \int_{\mathbb{X}} K(\mathbf{x}, \mathbf{y}) F(\mathbf{y}) \, d\mu(\mathbf{y})\,, \quad \mathbf{x} \in \mathbb{X},$$

whenever the integral on the right-hand side is well-defined for almost every $\mathbf{x} \in \mathbb{X}$. Note however, that for a given kernel $K \in \boldsymbol{\mathcal{K}}$ this may not be the case for all measurable functions $F : \mathbb{X} \to \mathbb{C}$.

The operator associated with a kernel $K$ will sometimes be denoted by $K^{op}$, for better distinction, but whenever the meaning is clear we will use the same notation $K$. The composition $K^{op} \circ L^{op}$ of two kernel operators leads to a corresponding multiplication operation on the kernel space $\boldsymbol{\mathcal{K}}$ given by

$$(K \circ L)(\mathbf{x}, \mathbf{y}) := \int_{\mathbb{X}} K(\mathbf{x}, \mathbf{z}) L(\mathbf{z}, \mathbf{y})\, d\mu(\mathbf{z}), \quad \mathbf{x}, \mathbf{y} \in \mathbb{X}. \tag{4.51}$$

But one has to be careful since again this multiplication is not a well-defined operation for all kernel pairs $(K, L)$ in $\boldsymbol{\mathcal{K}} \times \boldsymbol{\mathcal{K}}$.

**The algebra $\mathcal{A}_{m_\nu}$** A natural subset of $\boldsymbol{\mathcal{K}}$, where the multiplication is well-defined for all kernel pairs, is the space (see [46, page 249])

$$\mathcal{A} := \Big\{ K \in \boldsymbol{\mathcal{K}} \ : \ \|K|\mathcal{A}\| < \infty \Big\},$$

where for $K \in \boldsymbol{\mathcal{K}}$ the symbol $\|K|\mathcal{A}\|$ denotes the norm

$$\|K|\mathcal{A}\| := \max \Big\{ \operatorname*{ess\,sup}_{\mathbf{x} \in \mathbb{X}} \int_{\mathbf{y} \in \mathbb{X}} |K(\mathbf{x}, \mathbf{y})|\, d\mu(\mathbf{y}), \operatorname*{ess\,sup}_{\mathbf{y} \in \mathbb{X}} \int_{\mathbf{x} \in \mathbb{X}} |K(\mathbf{x}, \mathbf{y})|\, d\mu(\mathbf{x}) \Big\}.$$

The space $\mathcal{A}$ is even a Banach algebra as we will see in Proposition 4.5.1 below. Moreover, it is solid and closed under involution. The solidity of a kernel space is thereby defined analogously to the notion on a QBF-space, i.e., a subspace $\boldsymbol{\mathcal{L}} \subseteq \boldsymbol{\mathcal{K}}$ with quasi-norm $\|\cdot|\boldsymbol{\mathcal{L}}\|$ is said to be *solid* if for every kernel $K \in \boldsymbol{\mathcal{K}}$ we have the implication

$$|K| \le |L| \text{ for some } L \in \boldsymbol{\mathcal{L}} \Rightarrow K \in \boldsymbol{\mathcal{L}} \text{ and } \|K|\boldsymbol{\mathcal{L}}\| \le \|L|\boldsymbol{\mathcal{L}}\|\,.$$

Before we come to Proposition 4.5.1, let us introduce a more general weighted version of $\mathcal{A}$ which plays an essential role in coorbit theory.

Given a weight function $\nu : \mathbb{X} \to \mathbb{R}_+$ on $\mathbb{X}$, which is without saying always assumed to be measurable, we can associate a bivariate weight $m_\nu : \mathbb{X} \times \mathbb{X} \to [1, \infty)$ via

$$m_\nu(\mathbf{x}, \mathbf{y}) := \max \Big\{ \frac{\nu(\mathbf{x})}{\nu(\mathbf{y})}, \frac{\nu(\mathbf{y})}{\nu(\mathbf{x})} \Big\}, \quad \mathbf{x}, \mathbf{y} \in \mathbb{X}.$$

The weighted kernel algebra $\mathcal{A}_{m_\nu}$ (see [46, page 250] and [74, eq. (2.8)]) is then defined by

$$\mathcal{A}_{m_\nu} := \Big\{ K \in \boldsymbol{\mathcal{K}} \ : \ \|K|\mathcal{A}_{m_\nu}\| < \infty \Big\} \quad \text{with norm} \quad \|K|\mathcal{A}_{m_\nu}\| := \|K m_\nu|\mathcal{A}\|.$$

For a constant weight $\nu$ we get $m_\nu \equiv 1$ and thus retrieve the unweighted algebra $\mathcal{A} = \mathcal{A}_{m_\nu}$. In general, we have $m_\nu \ge 1$ wherefore $\mathcal{A}_{m_\nu}$ is always continuously embedded into $\mathcal{A}$.

An important structural result for $\mathcal{A}_{m_\nu}$ is stated in the following proposition.

**Proposition 4.5.1.** *For each weight $\nu : \mathbb{X} \to \mathbb{R}_+$ the kernel space $\mathcal{A}_{m_\nu}$ is a Banach algebra, solid, and closed under involution.*

Since a proof of this fact is not contained in [46, 111, 74] we decided to include one in this thesis.

*Proof.* We first prove that $\mathcal{A}_{m_\nu}$ is a Banach space with norm $\|\cdot|\mathcal{A}_{m_\nu}\|$. Apart from the completeness, everything is straightforward to verify. Hence let us just concentrate on the completeness and consider a Cauchy sequence $(K_n)_{n\in\mathbb{N}}$ in $\mathcal{A}_{m_\nu}$.

Without loss of generality we can assume $\|K_{n+1}-K_n|\mathcal{A}_{m_\nu}\| \le 2^{-n}$ for every $n \in \mathbb{N}$. We then define the auxiliary kernels $L_m$, $m \in \mathbb{N}$, given by

$$L_m(\mathbf{x},\mathbf{y}) := \sum_{n=1}^{m} \left|(K_{n+1}(\mathbf{x},\mathbf{y}) - K_n(\mathbf{x},\mathbf{y}))m_\nu(\mathbf{x},\mathbf{y})\right|, \quad \mathbf{x},\mathbf{y} \in \mathbb{X}.$$

Using monotone convergence, we obtain

$$\|\lim_{m\to\infty} L_m(\mathbf{x},\cdot)|L_1\| = \lim_{m\to\infty} \|L_m(\mathbf{x},\cdot)|L_1\| \le \sum_{n=1}^{\infty} \|K_{n+1}-K_n|\mathcal{A}_{m_\nu}\| \le 1 \quad \text{for a.e. } \mathbf{x} \in \mathbb{X}.$$

Hence, at every position $\mathbf{x} \in \mathbb{X}$ apart from a null set, the sequence $\sum_{n=1}^{\infty} |(K_{n+1}(\mathbf{x},\mathbf{y}) - K_n(\mathbf{x},\mathbf{y}))m_\nu(\mathbf{x},\mathbf{y})|$ converges for almost all $\mathbf{y} \in \mathbb{X}$. Since everywhere $m_\nu(\mathbf{x},\mathbf{y}) \neq 0$, also $\sum_{n=1}^{\infty} |(K_{n+1}(\mathbf{x},\mathbf{y}) - K_n(\mathbf{x},\mathbf{y}))|$ converges for almost all $\mathbf{x},\mathbf{y} \in \mathbb{X}$. This allows to define a kernel $K$ as the pointwise limit

$$K(\mathbf{x},\mathbf{y}) := \lim_{n\to\infty} K_n(\mathbf{x},\mathbf{y}), \quad \mathbf{x},\mathbf{y} \in \mathbb{X},$$

since those limits exist almost everywhere. This kernel is also characterized by the property

$$\begin{aligned} K(\mathbf{x},\cdot)m_\nu(\mathbf{x},\cdot) &= L_1\text{-}\lim\nolimits_{n\to\infty} K_n(\mathbf{x},\cdot)m_\nu(\mathbf{x},\cdot) \quad \text{for a.e. } \mathbf{x} \in \mathbb{X}, \\ K(\cdot,\mathbf{y})m_\nu(\cdot,\mathbf{y}) &= L_1\text{-}\lim\nolimits_{n\to\infty} K_n(\cdot,\mathbf{y})m_\nu(\cdot,\mathbf{y}) \quad \text{for a.e. } \mathbf{y} \in \mathbb{X}. \end{aligned}$$

This follows from the fact that $(K_n(\mathbf{x},\cdot)m_\nu(\mathbf{x},\cdot))_{n\in\mathbb{N}}$ and $(K_n(\cdot,\mathbf{y})m_\nu(\cdot,\mathbf{y}))_{n\in\mathbb{N}}$ are Cauchy sequences in $L_1$ for almost every $\mathbf{x} \in \mathbb{X}$ and $\mathbf{y} \in \mathbb{X}$, respectively.

The validity of $K \in \mathcal{A}_{m_\nu}$ and $K_n \to K$ in $\mathcal{A}_{m_\nu}$ is now a consequence of the observation

$$\begin{aligned} \operatorname*{ess\,sup}_{\mathbf{x}\in\mathbb{X}} \|(K_n(\mathbf{x},\cdot) - K(\mathbf{x},\cdot))m_\nu(\mathbf{x},\cdot)|L_1\| &= \operatorname*{ess\,sup}_{\mathbf{x}\in\mathbb{X}} \lim_{m\to\infty} \|(K_n(\mathbf{x},\cdot) - K_m(\mathbf{x},\cdot))m_\nu(\mathbf{x},\cdot)|L_1\| \\ &\le \liminf_{m\to\infty} \operatorname*{ess\,sup}_{\mathbf{x}\in\mathbb{X}} \|(K_n(\mathbf{x},\cdot) - K_m(\mathbf{x},\cdot))m_\nu(\mathbf{x},\cdot)|L_1\| \\ &\le \liminf_{m\to\infty} \|K_n - K_m|\mathcal{A}_{m_\nu}\| \to 0 \quad (n\to\infty) \end{aligned}$$

and the analogous result

$$\operatorname*{ess\,sup}_{\mathbf{y}\in\mathbb{X}} \|(K_n(\cdot,\mathbf{y}) - K(\cdot,\mathbf{y}))m_\nu(\cdot,\mathbf{y})|L_1\| \le \liminf_{m\to\infty} \|K_n - K_m|\mathcal{A}_{m_\nu}\| \to 0 \quad (n\to\infty).$$

This proves the completeness of $\mathcal{A}_{m_\nu}$ and establishes $\mathcal{A}_{m_\nu}$ as a Banach space. The solidity of $\mathcal{A}_{m_\nu}$ and the closedness under involution are clear.

Let us finally turn to the multiplicative structure. First observe that

$$m_\nu(\mathbf{x},\mathbf{y}) \le m_\nu(\mathbf{x},\mathbf{z})m_\nu(\mathbf{z},\mathbf{y}) \quad \text{for all } \mathbf{x},\mathbf{y},\mathbf{z} \in \mathbb{X}.$$

Hence, we can estimate for almost every $\mathbf{x} \in \mathbb{X}$

$$\begin{aligned}
&\int_{\mathbb{X}} \int_{\mathbb{X}} |K(\mathbf{x},\mathbf{z})L(\mathbf{z},\mathbf{y})| m_\nu(\mathbf{x},\mathbf{y}) \, d\mu(\mathbf{z}) \, d\mu(\mathbf{y}) \\
&\qquad \le \int_{\mathbb{X}} |K(\mathbf{x},\mathbf{z})| m_\nu(\mathbf{x},\mathbf{z}) \Big( \int_{\mathbb{X}} |L(\mathbf{z},\mathbf{y})| m_\nu(\mathbf{z},\mathbf{y}) \, d\mu(\mathbf{y}) \Big) \, d\mu(\mathbf{z}) \\
&\qquad\quad \le \|L|\mathcal{A}_{m_\nu}\| \int_{\mathbb{X}} |K(\mathbf{x},\mathbf{z})| m_\nu(\mathbf{x},\mathbf{z}) \, d\mu(\mathbf{z}) \le \|K|\mathcal{A}_{m_\nu}\| \|L|\mathcal{A}_{m_\nu}\|.
\end{aligned}$$

As a consequence of this estimate and the corresponding dual result

$$\int_{\mathbb{X}} \int_{\mathbb{X}} |K(\mathbf{x},\mathbf{z})L(\mathbf{z},\mathbf{y})| m_\nu(\mathbf{x},\mathbf{y}) \, d\mu(\mathbf{z}) \, d\mu(\mathbf{x}) \le \|K|\mathcal{A}_{m_\nu}\| \|L|\mathcal{A}_{m_\nu}\|,$$

we obtain the point-wise well-definedness of the product kernel $(K \circ L)(\mathbf{x},\mathbf{y})$ at almost every $(\mathbf{x},\mathbf{y}) \in \mathbb{X} \times \mathbb{X}$ and the estimate

$$\|K \circ L|\mathcal{A}_{m_\nu}\| \le \|K|\mathcal{A}_{m_\nu}\| \|L|\mathcal{A}_{m_\nu}\|.$$

The proof is finished. □

Next, we are interested in mapping properties of kernels belonging to $\mathcal{A}_{m_\nu}$. Using Schur's test and the Riesz-Thorin theorem, it can be shown that kernels in $\mathcal{A}$ operate continuously on the Lebesgue spaces $L_p(\mathbb{X})$ if $1 \le p \le \infty$. More general, as observed in [46], kernels in the weighted algebra $\mathcal{A}_{m_\nu}$ operate continuously on the weighted spaces $L_p^\nu(\mathbb{X})$ and $L_p^{1/\nu}(\mathbb{X})$.

**Lemma 4.5.2** (see [46, page 250]). *Let $p \in [1,\infty]$ and $\nu : \mathbb{X} \to \mathbb{R}_+$ be a weight on $\mathbb{X}$. Then all kernels $K \in \mathcal{A}_{m_\nu}$ operate continuously on $L_p^\nu(\mathbb{X})$ and $L_p^{1/\nu}(\mathbb{X})$ with $\|K^{op}|L_p^\nu \to L_p^\nu\| \le \|K|\mathcal{A}_{m_\nu}\|$ and $\|K^{op}|L_p^{1/\nu} \to L_p^{1/\nu}\| \le \|K|\mathcal{A}_{m_\nu}\|$.*

*Proof.* Let $\tilde{K} := |K| m_\nu$. Then, by Schur's test and complex interpolation, $\tilde{K}$ operates continuously on $L_p$, $1 \le p \le \infty$, with $\|\tilde{K}^{op}|L_p \to L_p\| \le \|\tilde{K}|\mathcal{A}\| = \|K|\mathcal{A}_{m_\nu}\|$. Further, the kernels $K_1$ and $K_2$ defined by $K_1(\mathbf{x},\mathbf{y}) := \nu(\mathbf{x})K(\mathbf{x},\mathbf{y})/\nu(\mathbf{y})$ and $K_2(\mathbf{x},\mathbf{y}) := \nu(\mathbf{y})K(\mathbf{x},\mathbf{y})/\nu(\mathbf{x})$ for $\mathbf{x},\mathbf{y} \in \mathbb{X}$ are majorized by $\tilde{K}$, i.e., $|K_1| \le \tilde{K}$ and $|K_2| \le \tilde{K}$. In view of Lemma 4.5.3 and the solidity of $L_p$, these kernels hence also induce continuous operations on $L_p$ with $\|K_i^{op} : L_p \to L_p\| \le \|K|\mathcal{A}_{m_\nu}\|$ for $i = 1, 2$. This translates to the assertion of the lemma. □

**The algebra $\mathcal{B}_{m_\nu,Y}$** Another way to define meaningful subspaces of the kernel space $\boldsymbol{\mathcal{K}}$ is to distinguish those kernels which operate continuously on some given function space $Y$. To be more concrete, let us assume that $Y$ is a QBF-space on $\mathbb{X}$. The space $\mathbf{L}(Y)$ of all bounded linear operators on $Y$ is then a quasi-Banach space with the same quasi-norm constant $C_Y$ as $Y$. Moreover, equipped with the operator-quasi-norm and the composition operation as multiplication, this space becomes a quasi-Banach algebra since in particular

$$\|K \circ L|Y \to Y\| \le \|K|Y \to Y\| \|L|Y \to Y\| \quad \text{for every } K, L \in \mathbf{L}(Y).$$

Also observe that, if $T_n \to T$ in $\mathbf{L}(Y)$, we have $\lim_{n\to\infty} \|T - T_n|Y \to Y\| = 0$ and thus the estimate

$$C_Y^{-1}\|T|Y \to Y\| \le \liminf_{n\to\infty} \Big(\|T - T_n|Y \to Y\| + \|T_n|Y \to Y\|\Big) = \liminf_{n\to\infty} \|T_n|Y \to Y\|. \tag{4.52}$$

In case of kernel operations, things are a bit more complicated, since in general not all elements of $\mathbf{L}(Y)$ stem from associated kernel functions.

A useful auxiliary result concerning the operations of kernels on solid QBF-spaces is the following lemma.

**Lemma 4.5.3** ([74, Lem. 2.45])**.** *Let $Y$ be a solid QBF-space on $\mathbb{X}$, and let $K : \mathbb{X}\times\mathbb{X} \to \mathbb{C}$ be a kernel such that $|K|$ operates continuously on $Y$. Further, let $L : \mathbb{X}\times\mathbb{X} \to \mathbb{C}$ be a kernel satisfying $|L| \le |K|$ almost everywhere. Then $L$ acts continuously on $Y$ with the estimate $\|L^{op}|Y \to Y\| \le \||K|^{op}|Y \to Y\|$.*

*Proof.* Let $F \in Y$. Then by the solidity of $Y$ we have $|F| \in Y$ with $\|F|Y\| = \||F||Y\|$. It follows $|K|^{op}[|F|] \in Y$, and since

$$\int_{\mathbb{X}} |L(\mathbf{x},\mathbf{y})F(\mathbf{y})|\, d\mu(\mathbf{y}) \le \int_{\mathbb{X}} |K(\mathbf{x},\mathbf{y})F(\mathbf{y})|\, d\mu(\mathbf{y}) = |K|^{op}[|F|](\mathbf{x})$$

for almost every $\mathbf{x} \in \mathbb{X}$, the operator

$$L^{op}[F](\mathbf{x}) = \int_{\mathbb{X}} L(\mathbf{x},\mathbf{y})F(\mathbf{y})\, d\mu(\mathbf{y})$$

is well-defined since the integral converges absolutely at these $\mathbf{x} \in \mathbb{X}$. Moreover, by solidity of $Y$, we have $L^{op}[F] \in Y$ and finally, we conclude

$$\|L^{op}[F]|Y\| \le \||L|^{op}[|F|]|Y\| \le \||\mathcal{K}|^{op}[|F|]|Y\| \le \||\mathcal{K}|^{op}|Y \to Y\|\|F|Y\|.$$

Hence $L$ defines a continuous operator $L^{op} : Y \to Y$ with $\|L^{op}|Y \to Y\| \le \||K|^{op}|Y \to Y\|$. ☐

For the choice $L = K$ in Lemma 4.5.3 we can deduce that especially $K$ acts on $Y$ with $\|K^{op}|Y \to Y\| \le \||K|^{op}|Y \to Y\|$. Notice however that $\|K^{op}|Y \to Y\| = \||K|^{op}|Y \to Y\|$ need not be true.

Now we can give the definition of the algebra $\mathcal{B}_{m_\nu,Y}$ which plays an important role in coorbit theory. It is the space

$$\mathcal{B}_{m_\nu,Y} := \Big\{K \in \boldsymbol{\mathcal{K}} \;:\; \|K|\mathcal{A}_{m_\nu}\| < \infty\,,\, |K|^{op} : Y \to Y \text{ operates continuously}\Big\}$$

equipped with the quasi-norm

$$\|K|\mathcal{B}_{m_\nu,Y}\| := \max\Big\{\|K|\mathcal{A}_{m_\nu}\|, \||K|^{op}|Y \to Y\|\Big\}.$$

Note that, motivated by Lemma 4.5.3, we pursued a small modification in the definition of $\mathcal{B}_{m_\nu,Y}$ compared to [74] or [111, eq. (3.4)]. We require $|K|^{op}$ to be a continuous operator on $Y$, which is a slightly stricter condition than just requiring this for $K^{op}$. The algebra $\mathcal{B}_{m_\nu,Y}$ is then always a solid space in the sense defined above. Moreover, this modification leads to a straight-forward proof that $\mathcal{B}_{m_\nu,Y}$ is a quasi-Banach algebra.

**Proposition 4.5.4.** *Let $Y$ be a rich solid QBF-space on $\mathbb{X}$ and let $\nu : \mathbb{X} \to \mathbb{R}_+$ be a weight function. Then $\mathcal{B}_{m_\nu,Y}$ is a solid quasi-Banach algebra.*

*Proof.* It is clear that $\mathcal{B}_{m_\nu,Y}$ is an algebra. Further, $\|\cdot|\mathcal{B}_{m_\nu,Y}\|$ is a quasi-norm on $\mathcal{B}_{m_\nu,Y}$ with quasi-norm constant $C_Y$ inherited from $Y$. The solidity of $\mathcal{B}_{m_\nu,Y}$ follows directly from Lemma 4.5.3 and the solidity of $\mathcal{A}_{m_\nu}$. Further $\|K \circ L|\mathcal{B}_{m_\nu,Y}\| \leq \|K|\mathcal{B}_{m_\nu,Y}\|\|L|\mathcal{B}_{m_\nu,Y}\|$.

It remains to prove the completeness. First note that a Cauchy sequence $(K_n)_{n\in\mathbb{N}}$ in $\mathcal{B}_{m_\nu,Y}$ is also Cauchy sequences in $\mathcal{A}_{m_\nu}$. By Proposition 4.5.1 we hence obtain a unique kernel $K \in \mathcal{A}_{m_\nu}$ as the $\mathcal{A}_{m_\nu}$-limit. Moreover, by possibly taking a suitable subsequence, we can without loss of generality assume that $K_n \to K$ pointwise almost everywhere (see proof of Proposition 4.5.1). In the sequel we will use this assumption.

Using the notation $K_0 := 0$ for the zero kernel, we will subsequently show that $|K-K_n|^{op}$ is a well-defined element of $\mathbf{L}(Y)$ for all $n \in \mathbb{N}_0$ and that $|K - K_n|^{op} \to 0$ in $\mathbf{L}(Y)$. With this we then directly obtain $K \in \mathcal{B}_{m_\nu,Y}$ and $K_n \to K$ in $\mathcal{B}_{m_\nu,Y}$, finishing the proof.

Let us turn to the operator side. For each fixed $n \in \mathbb{N}_0$, the operator sequence $(|K_m - K_n|^{op})_{m\in\mathbb{N}}$ is a Cauchy sequence in $\mathbf{L}(Y)$ due to the estimate $||K_m - K_n| - |K_{\tilde{m}} - K_n|| \leq |K_{\tilde{m}} - K_m|$, the solidity of $Y$ and Lemma 4.5.3. Let $\hat{K}_n \in \mathbf{L}(Y)$ denote the respective limits and take an arbitrary $F \in Y$. Using Fatou's lemma, we can estimate for almost every $\mathbf{x} \in \mathbb{X}$

$$\begin{aligned}\int_{\mathbb{X}} |(K - K_n)(\mathbf{x},\mathbf{y})F(\mathbf{y})|\, d\mu(\mathbf{y}) &= \int_{\mathbb{X}} \lim_{m\to\infty} |(K_m - K_n)(\mathbf{x},\mathbf{y})F(\mathbf{y})|\, d\mu(\mathbf{y}) \\ &\leq \liminf_{m\to\infty} \int_{\mathbb{X}} |(K_m - K_n)(\mathbf{x},\mathbf{y})F(\mathbf{y})|\, d\mu(\mathbf{y}) \\ &= \liminf_{m\to\infty} |K_m - K_n|^{op}[|F|](\mathbf{x}) \leq \hat{K}_n[|F|](\mathbf{x}).\end{aligned}$$

Here the last inequality is due to Lemma 4.2.1.

Since $\hat{K}_n[|F|] \in Y$, we deduce that $|K - K_n|^{op}F$ is well-defined pointwise almost everywhere for every $F \in Y$. Further, by solidity, $|K - K_n|^{op}F \in Y$ since $||K - K_n|^{op}F| \leq |K - K_n|^{op}[|F|] \leq \hat{K}_n[|F|]$. Moreover, the operators $|K - K_n|^{op}$ are contained in $\mathbf{L}(Y)$ since

$$\||K - K_n|^{op}F|Y\| \leq \|\hat{K}_n[|F|]|Y\| \leq \|\hat{K}_n|Y \to Y\|\|F|Y\|. \tag{4.53}$$

For the particular choice $n = 0$, this shows $|K|^{op} \in \mathbf{L}(Y)$ and hence $K \in \mathcal{B}_{m_\nu,Y}$. Further, using (4.52), we get for $n \geq 1$

$$C_Y^{-1}\|\hat{K}_n|Y \to Y\| \leq \liminf_{m\to\infty} \||K_m - K_n|^{op}|Y \to Y\| \leq \liminf_{m\to\infty} \|K_m - K_n|\mathcal{B}_{m_\nu,Y}\|,$$

which with (4.53) finally implies

$$\||K - K_n|^{op}|Y \to Y\| \leq \|\hat{K}_n|Y \to Y\| \leq C_Y \liminf_{m\to\infty} \|K_m - K_n|\mathcal{B}_{m_\nu,Y}\| \to 0 \quad (n \to \infty).$$

Together with $\|K - K_n|\mathcal{A}_{m_\nu}\| \to 0$, this establishes $K_n \to K$ in $\mathcal{B}_{m_\nu,Y}$, finishing the proof. □

With this structural result on $\mathcal{B}_{m_\nu,Y}$ our general introduction to kernel functions ends. In the remainder, we are interested in the concrete kernels occurring in the context of coorbit theory. Thereby our main aim are simple criteria to decide whether these kernels belong to $\mathcal{A}_{m_\nu}$ or $\mathcal{B}_{m_\nu,Y}$.

### 4.5.1 The (cross-)Gramian Kernels

Let $\alpha \in [0,1]$, and let $\mathfrak{M}_\alpha = \{m_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ and $\widetilde{\mathfrak{M}}_\alpha = \{\tilde{m}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ be two systems of $\alpha$-molecules in $L_2(\mathbb{R}^2)$ with respect to the canonical parametrization. The associated *cross-Gramian kernel* is given by

$$\mathcal{G}[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha](\mathbf{x},\mathbf{y}) := \overline{\langle m_{\mathbf{x}}, \tilde{m}_{\mathbf{y}} \rangle} = \langle \tilde{m}_{\mathbf{y}}, m_{\mathbf{x}} \rangle\,, \quad \mathbf{x},\mathbf{y} \in \mathbb{X}.$$

If both systems $\mathfrak{M}_\alpha$ and $\widetilde{\mathfrak{M}}_\alpha$ coincide, we simply speak of the *Gramian kernel* associated to $\mathfrak{M}_\alpha$ and denote it by $\mathcal{G}[\mathfrak{M}_\alpha]$.

Properties of such kernels play an essential role in the theory of $\alpha$-molecule coorbit spaces. Their respective maximal versions (compare with [74, eq. (2.13)]) are given as follows,

$$\mathcal{M}^\alpha_\tau[\widetilde{\mathfrak{M}}_\alpha, \mathfrak{M}_\alpha](\mathbf{x},\mathbf{y}) := \sup_{\mathbf{z}\in U^\alpha_\tau(\mathbf{y})} |\mathcal{G}[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha](\mathbf{x},\mathbf{z})| = \sup_{\mathbf{z}\in U^\alpha_\tau(\mathbf{y})} |\langle m_{\mathbf{x}}, \tilde{m}_{\mathbf{z}} \rangle|\,, \quad \mathbf{x},\mathbf{y}\in\mathbb{X}, \tag{4.54}$$

where $\tau \geq 0$ is a parameter and $U^\alpha_\tau(\mathbf{y})$ are subsets of $\mathbb{X}$ of the form (4.15). They are referred to as the *cross-Gramian maximal kernels* associated to $\mathfrak{M}_\alpha$ and $\widetilde{\mathfrak{M}}_\alpha$. If $\mathfrak{M}_\alpha = \widetilde{\mathfrak{M}}_\alpha$ we use the notation $\mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha] := \mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \mathfrak{M}_\alpha]$. Note that in the definition of $\mathcal{M}^\alpha_\tau$ the strict supremum and not the essential supremum is taken.

One of the main results of this appendix is Theorem 4.5.5 below. It states that, if the order of the $\alpha$-molecule systems $\mathfrak{M}_\alpha$ and $\widetilde{\mathfrak{M}}_\alpha$ is sufficiently high, the associated cross-Gramian maximal kernels belong to $\mathcal{A}_{m_\nu}$ or even $\mathcal{B}_{m_\nu,Y}$.

**Theorem 4.5.5.** *Let $\alpha \in [0,1]$, and let $\mathfrak{M}_\alpha$ and $\widetilde{\mathfrak{M}}_\alpha$ be two systems of $\alpha$-molecules of order $(L, M, N_1, N_2)$ with respect to the canonical parametrization. Assume that for some $\rho \geq 0$*

$$L > 2(\rho+2)\,, \quad M > 3(\rho+2) - \frac{3-\alpha}{2}\,, \quad N_1 > \rho + 2 + \frac{1+\alpha}{2}\,, \quad N_2 > 2(\rho+2)\,. \tag{4.55}$$

*Then, for arbitrary $\tau \geq 0$, the following statements on the associated cross-Gramian maximal kernels $\mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha]$ from (4.54) and their involutions $\mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha]^*$ hold true:*

*i) If $\rho \geq |\gamma|$, $\gamma \in \mathbb{R}$, we have with the weight $\nu = \nu_\gamma$ from (4.31)*

$$\mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha] \in \mathcal{A}_{m_\nu} \qquad \text{and} \qquad \mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha]^* \in \mathcal{A}_{m_\nu}.$$

*ii) Let $0 < p,q < \infty$, $r := \min\{1,p,q\}$, and $s \in \mathbb{R}$. If $\rho \geq \max\{|s| + 2(1/r - 1), |\tilde{s}|\}$, where $\tilde{s} := s - (1+\alpha)/p - (1-\alpha)/q$, then, for $Y := \mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$ and associated weight $\nu := \nu^{\alpha,s}_{p,q}$ as in (4.35), it holds*

$$\mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha] \in \mathcal{B}_{m_\nu,Y} \qquad \text{and} \qquad \mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha]^* \in \mathcal{B}_{m_\nu,Y}.$$

*Proof.* Condition (4.55) allows to choose $N > \rho + 2$ such that condition (4.57) in Proposition 4.5.8 is fulfilled. For such $N$, according to Proposition 4.5.8, it then holds

$$\mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha](\mathbf{x},\mathbf{y}) \lesssim \mathcal{G}_N(\mathbf{x},\mathbf{y}) \quad \text{and} \quad \mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha]^*(\mathbf{x},\mathbf{y}) \lesssim \mathcal{G}_N(\mathbf{x},\mathbf{y}).$$

If $\rho \geq |\gamma|$ we have $N > 2 + |\gamma|$ and thus $\mathcal{G}_N \in \mathcal{A}_{m_\nu}$ with $\nu = \nu_\gamma$ by Proposition 4.5.13. Since $\mathcal{A}_{m_\nu}$ is solid, this proves (i).

If $\rho \geq \max\{|s| + 2(1/r - 1), |\tilde{s}|\}$, we obtain, since now $N > 2 + \rho \geq 2 + |\tilde{s}|$,

$$\mathcal{G}_N \in \mathcal{A}_{m_\nu} \quad \text{with } \nu = \nu_{\tilde{s}}.$$

Further, since $N > 2 + \rho \geq 2/r + |s|$, according to Proposition 4.5.16

$$\mathcal{G}_N : \mathbb{L}^{\alpha,s}_{p,q} \to \mathbb{L}^{\alpha,s}_{p,q} \quad \text{operates continuously.}$$

Hence, $\mathcal{G}_N \in \mathcal{B}_{m_\nu,Y}$ and by solidity of $\mathcal{B}_{m_\nu,Y}$ statement (ii) follows. □

Due to the solidity of $\mathcal{A}_{m_\nu}$ and $\mathcal{B}_{m_\nu,Y}$, Theorem 4.5.5 also has implications for the corresponding cross-Gramian kernels $\mathcal{G}[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha]$, since

$$\mathcal{M}^\alpha_0[\widetilde{\mathfrak{M}}_\alpha, \mathfrak{M}_\alpha] = |\mathcal{G}[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha]|.$$

### 4.5.2 The Oscillation Kernels

Another important class of kernels occurring in the proof of Theorem 4.4.19 are the oscillation kernels from Definition 4.4.17 associated to the continuous $\alpha$-curvelet frame $\mathfrak{C}_\alpha$ and admissible coverings of $\mathbb{X}$. In the following definition, we introduce a continuous variant.

**Definition 4.5.6.** Let $\alpha \in [0, 1]$. For $\tau \geq 0$ we define the oscillation kernel $\mathrm{osc}_\tau$ associated to the continuous $\alpha$-curvelet frame $\mathfrak{C}_\alpha = \{\psi_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ by

$$\mathrm{osc}_\tau(\mathbf{x}, \mathbf{y}) := \sup_{\mathbf{z}\in U^\alpha_\tau(\mathbf{y})} |\mathcal{G}[\mathfrak{C}_\alpha](\mathbf{x}, \mathbf{y}) - \mathcal{G}[\mathfrak{C}_\alpha](\mathbf{x}, \mathbf{z})| = \sup_{\mathbf{z}\in U^\alpha_\tau(\mathbf{y})} |\langle \psi_{\mathbf{x}}, \psi_{\mathbf{y}} - \psi_{\mathbf{z}}\rangle|, \quad \mathbf{x}, \mathbf{y} \in \mathbb{X}\,.$$

Further, we put $\mathrm{osc}^*_\tau(\mathbf{x}, \mathbf{y}) := \mathrm{osc}_\tau(\mathbf{y}, \mathbf{x})$.

Let $\nu := \nu^{\alpha,s}_{p,q}$ and $Y := \mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$. As a direct consequence of Theorem 4.5.5, Proposition 3.1.3, and the solidity of $\mathcal{B}_{m_\nu,Y}$, the estimates

$$|\mathrm{osc}_\tau| \leq |\mathcal{G}[\mathfrak{C}_\alpha]| + |\mathcal{M}^\alpha_\tau[\mathfrak{C}_\alpha]| \quad \text{and} \quad |\mathrm{osc}^*_\tau| \leq |\mathcal{G}[\mathfrak{C}_\alpha]^*| + |\mathcal{M}^\alpha_\tau[\mathfrak{C}_\alpha]^*| \tag{4.56}$$

yield

$$\mathrm{osc}_\tau \in \mathcal{B}_{m_\nu,Y} \quad \text{and} \quad \mathrm{osc}^*_\tau \in \mathcal{B}_{m_\nu,Y} \quad \text{for all } \tau \geq 0.$$

However, we can even prove a more sophisticated result.

**Theorem 4.5.7.** *Let $\alpha \in [0, 1]$. Let $Y := \mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$ and $\nu := \nu^{\alpha,s}_{p,q}$ be the associated weight defined in* (4.35)*. Then, for $\tau \geq 0$, the kernels $\mathrm{osc}_\tau$ and $\mathrm{osc}^*_\tau$ defined in Definition 4.5.6 belong to $\mathcal{B}_{m_\nu,Y}$, and they satisfy*

$$\|\mathrm{osc}_\tau : \mathcal{B}_{m_\nu,Y}\| \leq C_\tau(2^\tau - 1) \quad \textit{and} \quad \|\mathrm{osc}^*_\tau : \mathcal{B}_{m_\nu,Y}\| \leq C_\tau(2^\tau - 1)$$

*with a value $C_\tau > 0$ that increases monotonically with $\tau$.*

*Proof.* Choose $N \in \mathbb{N}$ such that $N > \max\{|s|+2/r, 2+|s-(1+\alpha)/p-(1-\alpha)/q|\}$. Then $\mathcal{G}_N \in \mathcal{B}_{m_\nu,Y}$ due to Proposition 4.5.13 and Proposition 4.5.16. According to Proposition 4.5.9, we further have

$$\mathrm{osc}_\tau(\mathbf{x}, \mathbf{y}) \leq C_{N,\tau}(2^\tau - 1)\mathcal{G}_N(\mathbf{x}, \mathbf{y}) \quad \text{and} \quad \mathrm{osc}^*_\tau(\mathbf{x}, \mathbf{y}) \leq C_{N,\tau}(2^\tau - 1)\mathcal{G}_N(\mathbf{x}, \mathbf{y}).$$

Since $\mathcal{B}_{m_\nu,Y}$ is solid by Proposition 4.5.4, we deduce

$$\begin{aligned} \|\mathrm{osc}_\tau : \mathcal{B}_{m_\nu,Y}\| &\leq C_{N,\tau}(2^\tau - 1)\|\mathcal{G}_N : \mathcal{B}_{m_\nu,Y}\| \lesssim C_{N,\tau}(2^\tau - 1), \\ \|\mathrm{osc}^*_\tau : \mathcal{B}_{m_\nu,Y}\| &\leq C_{N,\tau}(2^\tau - 1)\|\mathcal{G}_N : \mathcal{B}_{m_\nu,Y}\| \lesssim C_{N,\tau}(2^\tau - 1). \end{aligned}$$
□

The proof of Theorem 4.5.5 and Theorem 4.5.7 rests on two bounding results, namely Proposition 4.5.8 and Proposition 4.5.9, which will be proved in the next subsection.

### 4.5.3 The Bounding Kernels

Motivated by Theorem 2.2.2, for each $N > 0$ we now introduce a non-negative kernel function $\mathcal{G}_N$ by

$$\mathcal{G}_N : \mathbb{X} \times \mathbb{X} \to [0, \infty), \quad \mathcal{G}_N(\mathbf{x}, \mathbf{y}) := \omega_\alpha(\mathbf{x}, \mathbf{y})^{-N}.$$

Clearly, due to $1 \leq \omega_\alpha < \infty$, these functions satisfy $0 < \mathcal{G}_N \leq 1$ for all $N > 0$. More precisely, they equal 1 on the diagonal and decay away from it, with a rate controlled by the parameter $N$.

Recalling Theorem 2.2.2, the kernels $\mathcal{G}_N$ are naturally suited for bounding the cross-Gramian kernels associated to canonically parameterized $\alpha$-molecule systems. A precise statement is formulated in the following proposition.

**Proposition 4.5.8.** *Let $\alpha \in [0, 1]$. Let $\mathfrak{M}_\alpha$ and $\widetilde{\mathfrak{M}}_\alpha$ be two systems of $\alpha$-molecules of order $(L, M, N_1, N_2)$ with respect to the canonical parametrization. Further assume that for $N > 0$*

$$L \geq 2N, \quad M > 3N - \frac{3-\alpha}{2}, \quad N_1 \geq N + \frac{1+\alpha}{2}, \quad N_2 \geq 2N. \tag{4.57}$$

*Then, for each $\tau \geq 0$, the cross-Gramian maximal kernel $\mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha]$ and its involution $\mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha]^*$ satisfy*

$$\mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha](\mathbf{x}, \mathbf{y}) \leq C_{N,\tau}\mathcal{G}_N(\mathbf{x}, \mathbf{y}) \quad \text{and} \quad \mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha]^*(\mathbf{x}, \mathbf{y}) \leq C_{N,\tau}\mathcal{G}_N(\mathbf{x}, \mathbf{y}),$$

*with a constant $C_{N,\tau} > 0$ independent of $\mathbf{x}, \mathbf{y} \in \mathbb{X}$ which grows with larger $\tau$.*

*Proof.* An application of Theorem 2.2.2 yields for every $\mathbf{x}, \mathbf{y} \in \mathbb{X}$

$$\mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha](\mathbf{x}, \mathbf{y}) = \sup_{\mathbf{z} \in U^\alpha_\tau(\mathbf{y})} |\langle m_\mathbf{x}, \tilde{m}_\mathbf{z} \rangle| \leq C_N \sup_{\mathbf{z} \in U^\alpha_\tau(\mathbf{y})} \omega_\alpha(\mathbf{x}, \mathbf{z})^{-N},$$

with a constant $C_N > 0$ depending only on $N$. Further, by Corollary 2.2.22 we have the estimate

$$\sup_{\mathbf{z} \in U^\alpha_\tau(\mathbf{y})} \omega_\alpha(\mathbf{x}, \mathbf{z})^{-N} \leq C_\tau \omega_\alpha(\mathbf{x}, \mathbf{y})^{-N} = C_\tau \mathcal{G}_N(\mathbf{x}, \mathbf{y}),$$

where $C_\tau \geq 1$ increases with $\tau \geq 0$. For the involution, we argue as follows,

$$\mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha]^*(\mathbf{x}, \mathbf{y}) = \mathcal{M}^\alpha_\tau[\mathfrak{M}_\alpha, \widetilde{\mathfrak{M}}_\alpha](\mathbf{y}, \mathbf{x}) \lesssim \mathcal{G}_N(\mathbf{y}, \mathbf{x}) \lesssim \mathcal{G}_N(\mathbf{x}, \mathbf{y}),$$

where the last estimate is due to the quasi-symmetry of $\omega_\alpha$ (see Theorem 2.2.12). $\square$

The kernels $\mathcal{G}_N$ can also be used to bound the oscillation kernels $\operatorname{osc}_\tau$ defined for $\tau \geq 0$ in Definition 4.5.6. In view of (4.56), we directly obtain

$$\operatorname{osc}_\tau \lesssim \mathcal{G}_N \quad \text{and} \quad \operatorname{osc}^*_\tau \lesssim \mathcal{G}_N,$$

which is true for arbitrary $N > 0$ due to the regularity of the $\alpha$-curvelet frame $\mathfrak{C}_\alpha$.

However, we can prove the much stronger result given below.

**Proposition 4.5.9.** *For every $N > 0$ and every $\tau \geq 0$, there exists a constant $C_{N,\tau} > 0$ such that*

$$\operatorname{osc}_\tau(\mathbf{x},\mathbf{y}) \leq C_{N,\tau}(2^\tau - 1)\mathcal{G}_N(\mathbf{x},\mathbf{y}) \quad \textit{and} \quad \operatorname{osc}^*_\tau(\mathbf{x},\mathbf{y}) \leq C_{N,\tau}(2^\tau - 1)\mathcal{G}_N(\mathbf{x},\mathbf{y})$$

*holds for all $\mathbf{x},\mathbf{y} \in \mathbb{X}$, whereby $C_{N,\tau}$ increases monotonically with $\tau$.*

*Proof.* By definition, the oscillation kernel $\operatorname{osc}_\tau$, where $\tau \geq 0$, has the form

$$\operatorname{osc}_\tau(\mathbf{x},\mathbf{y}) = \sup_{\mathbf{z}\in U^\alpha_\tau(\mathbf{y})} |\langle \psi_\mathbf{x}, \psi_\mathbf{y} - \psi_\mathbf{z}\rangle| , \quad \mathbf{x},\mathbf{y} \in \mathbb{X},$$

whereby $\psi_\mathbf{x} \in \mathfrak{C}_\alpha$ are the $\alpha$-curvelets defined in Section 3.1. If $\tau = 0$, we have $\operatorname{osc}_\tau = 0$ and the statement of the proposition is obviously true.

Let us turn to the case $\tau > 0$. For $\mathbf{y} = (y,\theta,u) \in \mathbb{X}$ and $\mathbf{z} = (z,\kappa,v) \in U^\alpha_\tau(\mathbf{y})$, we first split

$$\psi_\mathbf{y} - \psi_\mathbf{z} = (\psi_\mathbf{y} - \psi_{z,\theta,u}) + (\psi_{z,\theta,u} - \psi_{z,\kappa,u}) + (\psi_{z,\kappa,u} - \psi_\mathbf{z}) =: I_1(\mathbf{y},\mathbf{z}) + I_2(\mathbf{y},\mathbf{z}) + I_3(\mathbf{y},\mathbf{z}),$$

leading to the estimate

$$|\langle \psi_\mathbf{x}, \psi_\mathbf{y} - \psi_\mathbf{z}\rangle| \leq |\langle \psi_\mathbf{x}, I_1(\mathbf{y},\mathbf{z})\rangle| + |\langle \psi_\mathbf{x}, I_2(\mathbf{y},\mathbf{z})\rangle| + |\langle \psi_\mathbf{x}, I_3(\mathbf{y},\mathbf{z})\rangle|.$$

For the proof of the assertion, it then suffices to verify

$$\sup_{\mathbf{z}\in U^\alpha_\tau(\mathbf{y})} |\langle \psi_\mathbf{x}, I_j(\mathbf{y},\mathbf{z})\rangle| \lesssim (2^\tau - 1)\mathcal{G}_N(\mathbf{x},\mathbf{y}) \quad \text{for } j \in \{1,2,3\} \tag{4.58}$$

with an implicit constant depending on $N$.

To obtain these estimates, the regularity of the continuous $\alpha$-curvelet frame $\mathfrak{C}_\alpha = \{\psi_\mathbf{x}\}_{\mathbf{x}\in\mathbb{X}}$ needs to be taken into account. Recall that, on the Fourier side, for every $\mathbf{x} = (x,\eta,t) \in \mathbb{X}$ the curvelets $\psi_\mathbf{x} = \psi_{x,\eta,t} \in \mathfrak{C}_\alpha$ have the form

$$\widehat{\psi}_{x,\eta,t}(\xi) = t^{(1+\alpha)/2} \exp(-2\pi i\langle x,\xi\rangle) W_{\eta,t}(\xi)$$

with the wedge functions $W_{\eta,t} \in C^\infty_c(\mathbb{R}^2)$ from (3.2) given by

$$W_{\eta,t}(\xi(r,\phi)) = \begin{cases} U(tr)V(t^{-(1-\alpha)}\{\phi+\eta\}_{2\mathbf{T}}) & , \mathbf{x} = (x,\eta,t) \in \mathbb{X}_0 , \\ U_1(r)V_1(\{\phi\}_{2\mathbf{T}}) & , \mathbf{x} = (x,\eta,1) \in \mathbb{X}_1 . \end{cases}$$

Hereby, the points $\xi(r,\phi) = (r\cos(\phi), r\sin(\phi)) \in \mathbb{R}^2$ are determined by their polar coordinates $r \in [0,\infty)$ and $\phi \in [0,2\pi)$. For the definition of the functions $U$, $U_1$, $V$, and $V_1$ we refer to Section 3.1.

**Step 1:** In a first step, we differentiate $\hat{\psi}_{x,\eta,t}$ with respect to the parameters $(x,\eta,t) \in \mathbb{X}$. This is possible due to the regularity of the system $\mathfrak{C}_\alpha = \{\psi_\mathbf{x}\}_{\mathbf{x}\in\mathbb{X}}$.

In the sequel, the differentiation operators $\partial_{x_1}$ and $\partial_{x_2}$ shall act on the respective location parameters $x = (x_1,x_2) \in \mathbb{R}^2$. The symbol $\nabla_x := (\partial_{x_1},\partial_{x_2})$ will be used for the corresponding nabla operator. For a fixed orientation $\eta \in \mathbb{T}$, with associated orientation vector $e_\eta = (\cos(\eta), -\sin(\eta)) \in \mathbb{R}^2$, we further introduce the rotated versions

$$\begin{aligned} \partial^\eta_{x_1} &:= [R_\eta(\partial_{x_1},\partial_{x_2})^T]_1 = \langle e_\eta, \nabla_x\rangle = \cos\eta\cdot\partial_{x_1} - \sin\eta\cdot\partial_{x_2}, \\ \partial^\eta_{x_2} &:= [R_\eta(\partial_{x_1},\partial_{x_2})^T]_2 = \langle R_{\frac{\pi}{2}} e_\eta, \nabla_x\rangle = \sin\eta\cdot\partial_{x_1} + \cos\eta\cdot\partial_{x_2}. \end{aligned}$$

Here, the brackets $[\cdot]_1$ and $[\cdot]_2$ evaluate the first and second component, respectively.

At a fixed position $\mathbf{x} = (x, \eta, t) \in \mathbb{X}$, we obtain for $\xi \in \mathbb{R}^2$

$$\begin{aligned}\partial^{\eta}_{x_1} \widehat{\psi}_{x,\eta,t}(\xi) &= \partial^{\eta}_{x_1} \Big( t^{(1+\alpha)/2} W_{\eta,t}(\xi) \exp(-2\pi i \langle x, \xi \rangle) \Big) \\ &= -2\pi i [R_\eta \xi]_1 t^{(1+\alpha)/2} W_{\eta,t}(\xi) \exp(-2\pi i \langle x, \xi \rangle) \\ &= t^{-1} \big( t^{(1+\alpha)/2} W^{[1]}_{\eta,t}(\xi) \exp(-2\pi i \langle x, \xi \rangle) \big) =: t^{-1} \widehat{\psi}^{[1]}_{x,\eta,t}(\xi)\end{aligned}$$

with $$W^{[1]}_{\eta,t}(\xi) := -2\pi i t [R_\eta \xi]_1 W_{\eta,t}(\xi).$$

Similarly, we calculate

$$\begin{aligned}\partial^{\eta}_{x_2} \widehat{\psi}_{x,\eta,t}(\xi) &= \partial^{\eta}_{x_2} \Big( t^{(1+\alpha)/2} W_{\eta,t}(\xi) \exp(-2\pi i \langle x, \xi \rangle) \Big) \\ &= -2\pi i [R_\eta \xi]_2 t^{(1+\alpha)/2} W_{\eta,t}(\xi) \exp(-2\pi i \langle x, \xi \rangle) \\ &= t^{-\alpha} \big( t^{(1+\alpha)/2} W^{[2]}_{\eta,t}(\xi) \exp(-2\pi i \langle x, \xi \rangle) \big) =: t^{-\alpha} \widehat{\psi}^{[2]}_{x,\eta,t}(\xi)\end{aligned}$$

with $$W^{[2]}_{\eta,t}(\xi) := -2\pi i t^{\alpha} [R_\eta \xi]_2 W_{\eta,t}(\xi).$$

We proceed with the differentiation $\partial_\eta$ with respect to the parameter $\eta \in \mathbb{T}$. Here we obtain for $\mathbf{x} = (x, \eta, t) \in \mathbb{X}_0$

$$\begin{aligned}\partial_\eta \widehat{\psi}_{x,\eta,t}(\xi) &= \partial_\eta \Big( t^{(1+\alpha)/2} U(tr) V(t^{-(1-\alpha)} \{\phi + \eta\}_{2\mathbf{T}}) \exp(-2\pi i \langle x, \xi \rangle) \Big) \\ &= t^{-(1-\alpha)} \Big( t^{(1+\alpha)/2} U(tr) V'(t^{-(1-\alpha)} \{\phi + \eta\}_{2\mathbf{T}}) \exp(2\pi i \langle x, \xi \rangle) \Big).\end{aligned}$$

For $\mathbf{x} = (x, \eta, 1) \in \mathbb{X}_1$ the derivative vanishes,

$$\partial_\eta \widehat{\psi}_{x,\eta,1}(\xi) = \partial_\eta \Big( U_1(r) V_1(\{\phi\}_{2\mathbf{T}}) \exp(-2\pi i \langle x, \xi \rangle) \Big) = 0.$$

These results motivate the definition

$$W^{[\eta]}_{\eta,t}(\xi(r,\phi)) := \begin{cases} U(tr) V'(t^{-(1-\alpha)} \{\phi + \eta\}_{2\mathbf{T}}) & , \mathbf{x} = (x, \eta, t) \in \mathbb{X}_0, \\ 0 & , \mathbf{x} = (x, \eta, 1) \in \mathbb{X}_1, \end{cases}$$

and further

$$\widehat{\psi}^{[\eta]}_{x,\eta,t}(\xi) := t^{(1+\alpha)/2} W^{[\eta]}_{\eta,t}(\xi) \exp(2\pi i \langle x, \xi \rangle).$$

Then we can write

$$\partial_\eta \widehat{\psi}_{x,\eta,t}(\xi) = t^{-(1-\alpha)} \widehat{\psi}^{[\eta]}_{x,\eta,t}(\xi).$$

For the differentiation operator with respect to the scale variable $t \in (0, 1)$ we shall subsequently use the symbol $\partial_t$. We calculate for $\mathbf{x} = (x, \eta, t) \in \mathbb{X}_0$

$$\begin{aligned}\partial_t \widehat{\psi}_{x,\eta,t}(\xi) &= \partial_t \Big( t^{(1+\alpha)/2} U(tr) V(t^{-(1-\alpha)} \{\phi + \eta\}_{2\mathbf{T}}) \exp(2\pi i \langle x, \xi \rangle) \Big) \\ &= t^{-1} \cdot \Big( t^{(1+\alpha)/2} W^{[t]}_{\eta,t}(\xi) \exp(2\pi i \langle x, \xi \rangle) \Big) =: t^{-1} \widehat{\psi}^{[t]}_{x,\eta,t}(\xi)\,,\end{aligned}$$

where $W_{\eta,t}^{[t]}(\xi)$ is the sum

$$W_{\eta,t}^{[t]}(\xi) := W_{\eta,t}^{[t,0]}(\xi) + W_{\eta,t}^{[t,1]}(\xi). + W_{\eta,t}^{[t,2]}(\xi).$$

With $U^{[t]} \in C_c^\infty([0,\infty))$ and $V^{[t]} \in C_c^\infty([-\pi,\pi])$ given by

$$U^{[t]}(r) := rU'(r) \quad \text{and} \quad V^{[t]}(\phi) := \phi V'(\phi),$$

the summands of this sum are the functions

$$\begin{aligned}
W_{\eta,t}^{[t,0]}(\xi(r,\phi)) &:= \frac{1+\alpha}{2} W_{\eta,t}(\xi(r,\phi)), \\
W_{\eta,t}^{[t,1]}(\xi(r,\phi)) &:= U^{[t]}(tr)V(t^{-(1-\alpha)}\{\phi+\eta\}_{2\mathbf{T}}), \\
W_{\eta,t}^{[t,2]}(\xi(r,\phi)) &:= (\alpha-1)U(tr)V^{[t]}(t^{-(1-\alpha)}\{\phi+\eta\}_{2\mathbf{T}}).
\end{aligned}$$

Finally, for $\mathbf{x} = (x,\eta,1) \in \mathbb{X}_1$, we introduce the functions $\psi_{x,\eta,1}^{[t]} := 0$.

By parameter differentiation we have thus derived new function systems from $\mathfrak{C}_\alpha$, namely

$$\mathfrak{C}_\alpha^{[1]} := \{\psi_{\mathbf{x}}^{[1]}\}_{\mathbf{x}\in\mathbb{X}}, \quad \mathfrak{C}_\alpha^{[2]} := \{\psi_{\mathbf{x}}^{[2]}\}_{\mathbf{x}\in\mathbb{X}}, \quad \mathfrak{C}_\alpha^{[\eta]} := \{\psi_{\mathbf{x}}^{[\eta]}\}_{\mathbf{x}\in\mathbb{X}}, \quad \mathfrak{C}_\alpha^{[t]} := \{\psi_{\mathbf{x}}^{[t]}\}_{\mathbf{x}\in\mathbb{X}}.$$

**Step 2:** Next, we verify that these systems are instances of continuous $\alpha$-molecules of order $(\infty,\infty,\infty,\infty)$ with respect to the canonical parametrization, as $\mathfrak{C}_\alpha = \{\psi_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ itself.

The reason for this is that the modified wedge functions $W_{\eta,t}^{[1]}$, $W_{\eta,t}^{[2]}$, $W_{\eta,t}^{[\eta]}$, $W_{\eta,t}^{[t,0]}$, $W_{\eta,t}^{[t,1]}$, and $W_{\eta,t}^{[t,2]}$ are all built in the same way as the original functions $W_{\eta,t}$. Indeed, for $W_{\eta,t}^{[\eta]}$, $W_{\eta,t}^{[t,0]}$, $W_{\eta,t}^{[t,1]}$, and $W_{\eta,t}^{[t,2]}$ this is already clear from the above representations. Concerning $W_{\eta,t}^{[1]}$, and $W_{\eta,t}^{[2]}$, in polar representation $\xi(r,\phi) = (r\cos(\phi), r\sin(\phi))$, we have with $U^{[1]}(r) := rU(r)$ and $V^{[1]}(\phi) := V(\phi)$

$$W_{\eta,t}^{[1]}(\xi(r,\phi)) := -2\pi i \cos(\{\phi+\eta\}_{2\mathbf{T}}) U^{[1]}(tr) V^{[1]}(t^{-(1-\alpha)}\{\phi+\eta\}_{2\mathbf{T}}).$$

Further, with $U^{[2]}(r) = rU(r)$ and $V^{[2]}(\phi) = \phi V(\phi)$, it holds

$$W_{\eta,t}^{[2]}(\xi(r,\phi)) := -2\pi i \frac{\sin(\{\phi+\eta\}_{2\mathbf{T}})}{\{\phi+\eta\}_{2\mathbf{T}}} U^{[2]}(tr) V^{[2]}(t^{-(1-\alpha)}\{\phi+\eta\}_{2\mathbf{T}})$$

The remaining arguments are then analogous to those used in the proof of Proposition 3.1.3. They are based on the smoothness and support properties of the functions $U^{[1]}$, $V^{[1]}$, $U^{[2]}$, $V^{[2]}$, $U^{[t]}$, $V^{[t]}$, and $V'$, which are similar to the properties of $U$ and $V$.

**Step 3:** In the final step, we provide the desired estimates in (4.58) for $j \in \{1,2,3\}$.

Let us first assume $\mathbf{y} = (y,\theta,u) \in \mathbb{X}_0$ and $\mathbf{z} = (z,\kappa,v) \in U_\tau^\alpha(\mathbf{y})$. In this case, $\mathbf{z} \in \mathbb{X}_0$ and it holds (see definition of $U_\tau^\alpha(\mathbf{y})$ in (4.15))

$$\tilde{z} := z - y \in R_\theta^{-1} A_{\alpha,u} Q^\tau, \quad \tilde{\kappa} := \{\kappa-\theta\}_{2\mathbf{T}} \in u^{1-\alpha} I^\tau, \quad \tilde{v} := v/u \in J^\tau.$$

Using the fundamental theorem of calculus, we obtain for $\xi \in \mathbb{R}^2$

$$\begin{aligned}
\hat{I}_1(\xi) = \hat{\psi}_{y,\theta,u}(\xi) - \hat{\psi}_{y+\tilde{z},\theta,u}(\xi) &= -\int_0^1 \langle \nabla_y \hat{\psi}_{y+(1-a)\tilde{z},\theta,u}(\xi), \tilde{z}\rangle \, da \\
&= -u^{-1} \int_0^1 \hat{\psi}_{y+a\tilde{z},\theta,u}^{[1]}(\xi)[R_\theta\tilde{z}]_1 \, da - u^{-\alpha} \int_0^1 \hat{\psi}_{y+a\tilde{z},\theta,u}^{[2]}(\xi)[R_\theta\tilde{z}]_2 \, da.
\end{aligned}$$

Similarly, with $\tilde{\kappa} = \{\kappa - \theta\}_{2\mathbf{T}}$, we get for fixed $\xi \in \mathbb{R}^2$

$$\hat{I}_2(\xi) = \hat{\psi}_{z,\theta,u}(\xi) - \hat{\psi}_{z,(\theta+\tilde{\kappa})_{2\pi},u}(\xi) = \int_{\tilde{\kappa}}^{0} \partial_\eta \hat{\psi}_{z,(\theta+a)_{2\pi},u}(\xi)\, da = u^{-(1-\alpha)} \int_{\tilde{\kappa}}^{0} \hat{\psi}^{[\eta]}_{z,(\theta+a)_{2\pi},u}(\xi)\, da.$$

Hereby $(\theta + a)_{2\pi}$ is the value of $\theta + a$ modulo $2\pi$. Further, it holds for $\tilde{v} = v/u$ and $\xi \in \mathbb{R}^2$

$$\hat{I}_3(\xi) = \hat{\psi}_{z,\kappa,u}(\xi) - \hat{\psi}_{z,\kappa,\tilde{v}u}(\xi) = \int_{\tilde{v}u}^{u} \partial_t \hat{\psi}_{z,\kappa,a}(\xi)\, da = u^{-1} \int_{\tilde{v}u}^{u} \hat{\psi}^{[t]}_{z,\kappa,a}(\xi)\, da = \int_{\tilde{v}}^{1} \hat{\psi}^{[t]}_{z,\kappa,au}(\xi)\, da.$$

Using the Plancherel theorem and the theorem of Fubini-Tonelli, we deduce for arbitrary $\mathbf{x} \in \mathbb{X}$, $\mathbf{y} \in \mathbb{X}_0$, and $\mathbf{z} \in U^\alpha_\tau(\mathbf{y})$

$$\begin{aligned}
\Big|\langle \psi_{\mathbf{x}}, I_1(\mathbf{y},\mathbf{z})\rangle\Big| &= \Big| \frac{[R_\theta \tilde{z}]_1}{u} \Big\langle \hat{\psi}_{\mathbf{x}}, \int_0^1 \hat{\psi}^{[1]}_{y+a\tilde{z},\theta,u}\, da \Big\rangle + \frac{[R_\theta \tilde{z}]_2}{u^\alpha} \Big\langle \hat{\psi}_{\mathbf{x}}, \int_0^1 \hat{\psi}^{[2]}_{y+a\tilde{z},\theta,u}\, da \Big\rangle \Big| \\
&\le \frac{|[R_\theta \tilde{z}]_1|}{u} \int_0^1 \Big|\langle \hat{\psi}_{\mathbf{x}}, \hat{\psi}^{[1]}_{y+a\tilde{z},\theta,u}\rangle\Big|\, da + \frac{|[R_\theta \tilde{z}]_2|}{u^\alpha} \int_0^1 \Big|\langle \hat{\psi}_{\mathbf{x}}, \hat{\psi}^{[2]}_{y+a\tilde{z},\theta,u}\rangle\Big|\, da \\
&\le \frac{|[R_\theta \tilde{z}]_1|}{u} \mathcal{M}^\alpha_\tau[\mathfrak{C}^{[1]}_\alpha, \mathfrak{C}_\alpha](\mathbf{x},\mathbf{y}) + \frac{|[R_\theta \tilde{z}]_2|}{u^\alpha} \mathcal{M}^\alpha_\tau[\mathfrak{C}^{[2]}_\alpha, \mathfrak{C}_\alpha](\mathbf{x},\mathbf{y})\,.
\end{aligned}$$

From $\tilde{z} \in R_\theta^{-1} A_{\alpha,u} Q^\tau$ we deduce $R_\theta \tilde{z} \in A_{\alpha,u} Q^\tau$ and hence $|[R_\theta \tilde{z}]_1| \le \tau u$ and $|[R_\theta \tilde{z}]_2| \le \tau u^\alpha$. Invoking Proposition 4.5.8, we arrive at the estimate

$$\Big|\langle \psi_{\mathbf{x}}, I_1(\mathbf{y},\mathbf{z})\rangle\Big| \le C_{N,\tau} \tau \mathcal{G}_N(\mathbf{x},\mathbf{y})\,,$$

from which the desired estimate for $j = 1$ in (4.58) follows due to $\tau \le 2^\tau - 1$.

Similarly, we estimate the two other terms corresponding to $j \in \{2,3\}$ in (4.58). Since $\tilde{\kappa} \in u^{1-\alpha} I^\tau$, we have $|\tilde{\kappa}| \le \tau u^{1-\alpha}$ and thus

$$\begin{aligned}
\Big|\langle \psi_{\mathbf{x}}, I_2(\mathbf{y},\mathbf{z})\rangle\Big| &= \Big|\langle \hat{\psi}_{\mathbf{x}}, \hat{I}_2(\mathbf{y},\mathbf{z})\rangle\Big| = u^{-(1-\alpha)} \Big|\Big\langle \hat{\psi}_{\mathbf{x}}, \int_0^{\tilde{\kappa}} \hat{\psi}^{[\eta]}_{z,(\theta+a)_{2\pi},u}(\xi)\, da \Big\rangle\Big| \\
&\le u^{-(1-\alpha)} \int_0^{|\tilde{\kappa}|} \Big|\langle \hat{\psi}_{\mathbf{x}}, \hat{\psi}^{[\eta]}_{z,(\theta+a)_{2\pi},u}\rangle\Big|\, da \le u^{-(1-\alpha)} \int_0^{|\tilde{\kappa}|} \mathcal{M}^\alpha_\tau[\mathfrak{C}^{[\eta]}_\alpha, \mathfrak{C}_\alpha](\mathbf{x},\mathbf{y})\, da \\
&\le C_{N,\tau} u^{-(1-\alpha)} |\tilde{\kappa}| \mathcal{G}_N(\mathbf{x},\mathbf{y}) \le C_{N,\tau} \tau \mathcal{G}_N(\mathbf{x},\mathbf{y}).
\end{aligned}$$

Finally, due to $\tilde{v} \in J^\tau = [2^{-\tau}, 2^\tau]$ we have $|1 - \tilde{v}| \le 2^\tau - 1$. This yields

$$\begin{aligned}
\Big|\langle \psi_{\mathbf{x}}, I_3(\mathbf{y},\mathbf{z})\rangle\Big| &= \Big|\langle \hat{\psi}_{\mathbf{x}}, \hat{I}_3(\mathbf{y},\mathbf{z})\rangle\Big| = \Big|\Big\langle \hat{\psi}_{\mathbf{x}}, \int_{\tilde{v}}^1 \hat{\psi}^{[t]}_{z,\kappa,au}(\xi)\, da \Big\rangle\Big| \\
&\le \Big| \int_{\tilde{v}}^1 \Big|\langle \hat{\psi}_{\mathbf{x}}, \hat{\psi}^{[t]}_{z,\kappa,au}\rangle\Big|\, da \Big| \le \Big| \int_{\tilde{v}}^1 \mathcal{M}^\alpha_\tau[\mathfrak{C}^{[t]}_\alpha, \mathfrak{C}_\alpha](\mathbf{x},\mathbf{y})\, da \Big| \\
&\le C_{N,\tau} |1 - \tilde{v}| \mathcal{G}_N(\mathbf{x},\mathbf{y}) \le C_{N,\tau} (2^\tau - 1) \mathcal{G}_N(\mathbf{x},\mathbf{y}).
\end{aligned}$$

It remains to handle the case $\mathbf{y} \in \mathbb{X}_1$. Here the estimates are trivial for $j \in \{2,3\}$ since $I_2(\mathbf{y},\mathbf{z}) = I_3(\mathbf{y},\mathbf{z}) = 0$. For $j = 1$ the arguments are the same as before.

Concerning the estimate of the involuted kernel $\mathrm{osc}^*_\tau$, we can argue with the quasi-symmetry of $\mathcal{G}_N$ which directly follows from the quasi-symmetry of $\omega_\alpha$ (see Theorem 2.2.12). $\square$

In the following, we will deduce sufficient conditions for $\mathcal{G}_N$ to belong to $\mathcal{A}_{m_\nu}$ and $\mathcal{B}_{m_\nu,Y}$. Together with solidity arguments based on Propositions 4.5.8 and 4.5.9, those are the foundation for the proof of Theorem 4.5.5 and Theorem 4.5.7.

### 4.5.4 The Convolution-Type Auxiliary Kernels

For the subsequent investigation, it is useful to introduce another scale of kernels with a convolution-type structure. For $N > 0$, let us introduce

$$\mathcal{H}_N : \mathbb{X} \times \mathbb{X} \to [0, \infty), \quad \mathcal{H}_N(\mathbf{x}, \mathbf{y}) := \max\left\{\frac{t}{u}, \frac{u}{t}\right\}^{-N} (1 + d'_\alpha(\mathbf{x}, \mathbf{y}))^{-N},$$

where, with $e_1$ and $e_2$ denoting the respective unit vectors of $\mathbb{R}^2$,

$$d'_\alpha(\mathbf{x}, \mathbf{y}) := \frac{|\{\eta - \theta\}|^2}{\max\{t, u\}^{2(1-\alpha)}} + \frac{|\langle R_\eta^{-1} e_2, x - y\rangle|^2}{\max\{t, u\}^{2\alpha}} + \frac{|\langle R_\eta^{-1} e_1, x - y\rangle|}{\max\{t, u\}}.$$

Hereby, the bracket $\{\cdot\} = \{\cdot\}_{\mathbf{T}}$ denotes the projective bracket defined in (2.9). For convenience, we will often use the short-hand notation $(t, u)_+ := \max\{t, u\}$ for $t, u \in \mathbb{R}$ in the sequel.

The kernels $\mathcal{H}_N$ can also be written in the form

$$\mathcal{H}_N(\mathbf{x}, \mathbf{y}) = H_N\big(A_{\alpha,t}^{-1} R_\eta (y - x), t^{-(1-\alpha)}\{\theta - \eta\}, u/t\big)$$

with non-negative functions $H_N : \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+ \to [0, \infty)$ given by

$$H_N(a, b, c) := \max\left\{c, c^{-1}\right\}^{-N} \left(1 + \frac{|b|^2}{\max\{1, c\}^{2(1-\alpha)}} + \frac{|[a]_2|^2}{\max\{1, c\}^{2\alpha}} + \frac{|[a]_1|}{\max\{1, c\}}\right)^{-N}. \tag{4.59}$$

Clearly, from the definition, $0 < H_N(a, b, c) \le 1$ for all $(a, b, c) \in \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+$.

An important relation of $\mathcal{H}_N$ to the kernels $\mathcal{G}_N$ and $\mathcal{G}_N^*$ are the following estimates.

**Lemma 4.5.10.** *For every $N > 0$ we have*

$$\mathcal{G}_N(\mathbf{x}, \mathbf{y}) \lesssim \mathcal{H}_N(\mathbf{x}, \mathbf{y}) \quad \textit{and} \quad \mathcal{G}_N^*(\mathbf{x}, \mathbf{y}) \lesssim \mathcal{H}_N(\mathbf{x}, \mathbf{y}), \quad \mathbf{x}, \mathbf{y} \in \mathbb{X}.$$

*Proof.* Recall the simplified index distance $\omega_\alpha^{\mathrm{sim}}$ from Definition 2.2.3 given by

$$\omega_\alpha^{\mathrm{sim}}(\mathbf{x}, \mathbf{y}) = \max\left\{\frac{t}{u}, \frac{u}{t}\right\}(1 + d_\alpha^{\mathrm{sim}}(\mathbf{x}, \mathbf{y}))$$

with

$$d_\alpha^{\mathrm{sim}}(\mathbf{x}, \mathbf{y}) = \frac{|\{\eta - \theta\}|^2}{\max\{t, u\}^{2(1-\alpha)}} + \frac{|x - y|^2}{\max\{t, u\}^{2\alpha}} + \frac{|\langle R_\eta^{-1} e_1, x - y\rangle|}{\max\{t, u\}}.$$

It was proved in Lemma 2.2.4 that $\omega_\alpha(\mathbf{x}, \mathbf{y}) \gtrsim \omega_\alpha^{\mathrm{sim}}(\mathbf{x}, \mathbf{y})$ and $d_\alpha(\mathbf{x}, \mathbf{y}) \gtrsim d_\alpha^{\mathrm{sim}}(\mathbf{x}, \mathbf{y})$. Since $|\langle R_\eta^{-1} e_2, x - y\rangle| \le |x - y|$ we can deduce

$$d_\alpha(\mathbf{x}, \mathbf{y}) \gtrsim d_\alpha^{\mathrm{sim}}(\mathbf{x}, \mathbf{y}) \ge d'_\alpha(\mathbf{x}, \mathbf{y}).$$

The assertion follows. $\square$

As a consequence of these estimates, a sufficient condition for $\mathcal{H}_N$ to belong to $\mathcal{A}_{m_\nu}$ and $\mathcal{B}_{m_\nu,Y}$ allows to deduce such conditions for $\mathcal{G}_N$ and $\mathcal{G}_N^*$. The advantage of the kernels $\mathcal{H}_N$ is their convolution-type structure. Due to this structure, the analysis of $\mathcal{H}_N$ can be reduced to an analysis of the corresponding function $H_N$ given by (4.59), which is a great simplification.

One could pose the question, why the kernels $\mathcal{G}_N$ are defined after all, if we could directly use $\mathcal{H}_N$ as bounding kernels. The reason for this is that the kernels $\mathcal{H}_N$ do not possess the same symmetry and stability properties as $\mathcal{G}_N$ (see Corollary 2.2.9 and Corollary 2.2.22).

In the sequel, the integrability properties of $H_N$ stated in the next lemma are essential.

**Lemma 4.5.11.** *Let* $0 < r \leq 1$, $\gamma \in \mathbb{R}$, *and* $N > 2/r + \max\{\gamma, 0\}$. *Then*

$$\int_{\mathbb{R}_+}\int_{\mathbb{R}}\int_{\mathbb{R}^2} H_N(y,\theta,u)^r \max\{u,u^{-1}\}^{r\gamma}(u+1)^2 \frac{dy\,d\theta\,du}{u^3} < \infty\,,$$

$$\int_{\mathbb{R}}\int_{\mathbb{R}^2} H_N(y,\theta,1)^r\,dy\,d\theta < \infty\,.$$

*Proof.* For fixed $u \in \mathbb{R}_+$, let us first consider the inner integral

$$I(u) := \int_{\mathbb{R}}\int_{\mathbb{R}^2} H_N(y,\theta,u)^r\,dy\,d\theta.$$

Generally, we have for quantities $R > 0$, $c > 0$, $\gamma > 0$, and $N > \frac{1}{\gamma}$ the formula

$$\begin{aligned}
\int_0^\infty (R + c^{-\gamma}a^\gamma)^{-N}\,da &\asymp \int_0^\infty (R^{1/\gamma} + c^{-1}a)^{-\gamma N}\,da = c \cdot \int_{R^{1/\gamma}}^\infty a^{-\gamma N}\,da \\
&= \frac{c}{1-\gamma N}\Big[a^{-\gamma N+1}\Big]_{R^{1/\gamma}}^\infty = \frac{c}{\gamma N - 1} R^{-(N-1/\gamma)} \asymp c \cdot R^{-(N-1/\gamma)}\,.
\end{aligned}$$

Applying this formula iteratively, we can calculate $I(u)$. For $N > 2/r$ and $u \in \mathbb{R}_+$ we obtain

$$\begin{aligned}
I(u) &= \max\{u,u^{-1}\}^{-Nr}\int_{\mathbb{R}}\int_{\mathbb{R}^2}\Big(1 + \frac{|\theta|^2}{\max\{1,u\}^{2(1-\alpha)}} + \frac{|[y]_2|^2}{\max\{1,u\}^{2\alpha}} + \frac{|[y]_1|}{\max\{1,u\}}\Big)^{-Nr} dy\,d\theta \\
&\asymp \max\{u,u^{-1}\}^{-Nr}\max\{1,u\}^2.
\end{aligned}$$

Specifically for $u = 1$, this implies $I(1) \asymp 1 < \infty$ proving the second assertion.

To obtain the first assertion, we evaluate the outer integral over the scales, namely

$$\begin{aligned}
\int_{\mathbb{R}_+} I(u)\max\{u,u^{-1}\}^{r\gamma}(u+1)^2\frac{du}{u^3} &\asymp \int_0^\infty \max\{u,u^{-1}\}^{-Nr+\gamma r}\max\{1,u\}^2(u+1)^2\,\frac{du}{u^3} \\
&= \int_0^1 u^{Nr-\gamma r-2}(u+1)^2\,\frac{du}{u} + \int_1^\infty u^{-Nr+\gamma r+2}\frac{(u+1)^2}{u^2}\,\frac{du}{u} \\
&\asymp \int_0^1 u^{Nr-\gamma r-2}\,\frac{du}{u} + \int_1^\infty u^{-(Nr-\gamma r-2)}\,\frac{du}{u} < \infty\,.
\end{aligned}$$

Precisely if $Nr - \gamma r - 2 > 0$, or equivalently $N > 2/r + \gamma$, the last two integrals converge. Since the assumption $N > 2/r + \max\{\gamma, 0\}$ ensures both $N > 2/r$ as well as $N > 2/r + \gamma$, the proof is finished. □

Another important auxiliary result concerning $H_N$ is the following lemma.

**Lemma 4.5.12.** *It holds, uniformly in $(x,\eta,t) \in \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+$,*

$$\inf_{(\kappa,v)\in I\times J} H_N(x,\eta+t^{1-\alpha}\kappa,tv) \asymp H_N(x,\eta,t) \asymp \sup_{(\kappa,v)\in I\times J} H_N(x,\eta+t^{1-\alpha}\kappa,tv),$$

*where $I := [-1,1]$ and $J := [\frac{1}{2},2]$.*

*Proof.* Let us define $\tilde{H}_N(x,\eta,t) := \sup_{(\kappa,v)\in I\times J} H_N(x,\eta+t^{1-\alpha}\kappa,tv)$ and $H'_N(x,\eta,t) := \inf_{(\kappa,v)\in I\times J} H_N(x,\eta+t^{1-\alpha}\kappa,tv)$. Clearly, $H'_N(x,\eta,t) \le H_N(x,\eta,t) \le \tilde{H}_N(x,\eta,t)$. For the opposite estimates, we first note that for $c \in \mathbb{R}_+$

$$\inf_{v\in cJ} \max\{v,v^{-1}\} \asymp \sup_{v\in cJ} \max\{v,v^{-1}\} \quad \text{and} \quad \inf_{v\in cJ} \max\{1,v\} \asymp \sup_{v\in cJ} \max\{1,v\}.$$

Further, the estimate $|\eta+t^{1-\alpha}\kappa|^2 \le 2|\eta|^2 + 2\max\{1,t\}^{2(1-\alpha)}|\kappa|^2$ yields

$$\sup_{\kappa\in I} \frac{|\eta+t^{1-\alpha}\kappa|^2}{\max\{1,t\}^{2(1-\alpha)}} \le \frac{2|\eta|^2}{\max\{1,t\}^{2(1-\alpha)}} + 2.$$

Similarly, $|\eta|^2 \le 2|\eta-t^{1-\alpha}\kappa|^2 + 2\max\{1,t\}^{2(1-\alpha)}|\kappa|^2$ leads to

$$\inf_{\kappa\in I} \frac{|\eta+t^{1-\alpha}\kappa|^2}{\max\{1,t\}^{2(1-\alpha)}} \ge \frac{|\eta|^2}{2\max\{1,t\}^{2(1-\alpha)}} - 1.$$

Altogether, all these ingredients yield $\tilde{H}_N(x,\eta,t) \lesssim H_N(x,\eta,t) \lesssim H'_N(x,\eta,t)$. □

After this preparation, we are now ready to prove conditions on $N$ ensuring the membership of $\mathcal{H}_N$, and in turn $\mathcal{G}_N$ and $\mathcal{G}_N^*$, in the algebras $\mathcal{A}_{m_\nu}$ and $\mathcal{B}_{m_\nu,Y}$.

### 4.5.5 Kernel Criteria for $\mathcal{G}_N \in \mathcal{A}_{m_\nu}$ and $\mathcal{G}_N \in \mathcal{B}_{m_\nu,Y}$

In this last subsection of the appendix we aim for an easily applicable criterion to be able to decide whether a kernel $\mathcal{G}_N$ belongs to the algebra $\mathcal{B}_{m_\nu,Y}$, where either $Y := L^s_{p,q}(\mathbb{X})$ or $Y := \mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$ and $\nu$ is the respective associated weight.

As a first step, we prove a simple criterion which ensures that $\mathcal{G}_N$ belongs to the algebra $\mathcal{A}_{m_\nu}$ with a weight $\nu = \nu_\gamma$ of the form (4.31) with $\gamma \in \mathbb{R}$. The following proposition shows that there exists a threshold for $N$ above which $\mathcal{G}_N \in \mathcal{A}_{m_\nu}$ is guaranteed. This is not surprising since a large $N$ promotes a fast off-diagonal decay of $\mathcal{G}_N$.

**Proposition 4.5.13.** *Let $\gamma \in \mathbb{R}$ and let $\nu = \nu_\gamma$ be the weight defined in* (4.31)*. If $N > 2+|\gamma|$ then $\mathcal{G}_N$ is an element of $\mathcal{A}_{m_\nu}$.*

*Proof.* According to Lemma 4.5.10 we have

$$\mathcal{G}_N(\mathbf{x},\mathbf{y}) \lesssim \mathcal{H}_N(\mathbf{x},\mathbf{y}) = H_N\big(A^{-1}_{\alpha,t} R_\eta(y-x), t^{-(1-\alpha)}\{\theta-\eta\}, u/t\big),$$

where $H_N$ is the function from (4.59). An application of Lemma 4.5.15, which is proved below, then yields the assertion. We only need to verify that the quantities $A_i$, $i \in \{1,2,3\}$, associated to $H_N$ are finite if $N > 2 + |\gamma|$.

For the quantities $A_1$ and $A_3$ this follows directly from Lemma 4.5.11 which asserts

$$A_1 = \int\limits_{\mathbb{R}_+}\int\limits_{\mathbb{R}}\int\limits_{\mathbb{R}^2} H_N(y,\theta,u)\max\{u,u^{-1}\}^{|\gamma|}(u+1)^2\frac{dy\,d\theta\,du}{u^3} < \infty\,,$$

$$A_3 = \int\limits_{\mathbf{T}}\int\limits_{\mathbb{R}^2} H_N(y,\theta,1)\,dy\,d\theta \le \int\limits_{\mathbb{R}}\int\limits_{\mathbb{R}^2} H_N(y,\theta,1)\,dy\,d\theta < \infty\,.$$

Using Lemma 4.5.14 and Lemma 4.5.12, we can further show that $A_2 \lesssim A_1$. Indeed, noting that $dt/t$ is the Haar measure of the multiplicative group $\mathbb{R}_+$, we have with the window $J := [\frac{1}{2}, 2] \subset \mathbb{R}_+$

$$\begin{aligned}
A_2 &\le \operatorname*{ess\,sup}_{u\in\mathbb{R}_+}\Big(\operatorname*{ess\,sup}_{v\in uJ}\int_{\mathbb{R}}\int_{\mathbb{R}^2}|H_N(y,\theta,v)|\max\{v,v^{-1}\}^{|\gamma|}(1+v)^{-2}\,dy\,d\theta\Big)\\
&\lesssim \int_{\mathbb{R}_+}\Big(\int_{\mathbb{R}}\int_{\mathbb{R}^2}\operatorname*{ess\,sup}_{v\in uJ}|H_N(y,\theta,v)|\max\{v,v^{-1}\}^{|\gamma|}\,dy\,d\theta\Big)\frac{du}{u}\\
&\lesssim \int_{\mathbb{R}_+}\int_{\mathbb{R}}\int_{\mathbb{R}^2}|H_N(y,\theta,u)|\max\{u,u^{-1}\}^{|\gamma|}\,\frac{dy\,d\theta\,du}{u} \le A_1.
\end{aligned}$$

As a consequence, also $A_2 < \infty$ holds true if $N > 2 + |\gamma|$. □

In the proof of the previous proposition, we have implicitly used an embedding result for Wiener amalgam spaces. It can be termed as an estimate of the corresponding quasi-norms. Given an arbitrary measure space $(\mathbb{X}, \mu)$ and $0 < p \le \infty$, let $L_p(\mathbb{X})$ denote the usual Lebesgue space on $\mathbb{X}$ with quasi-norm $\|\cdot|L_p\|$. Further, assume that $\{W_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{X}}$ is a family of measurable windows $W_{\mathbf{x}} \subseteq \mathbb{X}$ such that the associated dual windows $\widetilde{W_{\mathbf{x}}} \subseteq \mathbb{X}$ defined by the relation $\mathcal{X}_{\widetilde{W_{\mathbf{x}}}}(\mathbf{y}) = \mathcal{X}_{W_{\mathbf{y}}}(\mathbf{x})$ are also measurable. Then we have the following estimate.

**Lemma 4.5.14.** *If there is $m > 0$ such that $\mu(\widetilde{W_{\mathbf{x}}}) \ge m$ holds independently of $\mathbf{x} \in \mathbb{X}$, then we have for $0 < p \le q \le \infty$ and every measurable function $f : \mathbb{X} \to \mathbb{C}$ the estimate*

$$\|Kf|L_q\| \le m^{1/q-1/p}\|Kf|L_p\|,$$

*where $Kf(\mathbf{x}) := \operatorname{ess\,sup}_{\mathbf{y}\in\mathbb{X}}|f(\mathbf{y})|\mathcal{X}_{W_{\mathbf{x}}}(\mathbf{y})$ is a Wiener control function as in* (4.14).

*Proof.* We restrict the proof to the case $q < \infty$, with obvious modifications if $q = \infty$.

First, we observe that $\|Kf|L_\infty\| = \|f|L_\infty\|$. Second, we see that

$$\begin{aligned}
\|Kf|L_q\|^q &= \int\limits_{\mathbb{X}} \operatorname*{ess\,sup}_{\mathbf{y}\in\mathbb{X}}|f(\mathbf{y})\mathcal{X}_{W_{\mathbf{x}}}(\mathbf{y})|^q\,d\mu(\mathbf{x})\\
&\ge \operatorname*{ess\,sup}_{\mathbf{y}\in\mathbb{X}}\int\limits_{\mathbb{X}}|f(\mathbf{y})|^q\mathcal{X}_{W_{\mathbf{x}}}(\mathbf{y})\,d\mu(\mathbf{x}) \ge m\|Kf|L_\infty\|^q.
\end{aligned}$$

Since $|Kf|/\|Kf|L_\infty\| \le 1$ almost everywhere, we can deduce for $p \le q$

$$m \le \frac{\|Kf|L_q\|^q}{\|Kf|L_\infty\|^q} \le \frac{\|Kf|L_p\|^p}{\|Kf|L_\infty\|^p}.$$

Since $1/q \le 1/p$, this further implies

$$\frac{\|Kf|L_q\|}{\|Kf|L_\infty\|m^{1/q}} \le \frac{\|Kf|L_p\|}{\|Kf|L_\infty\|m^{1/p}},$$

and therefore $\|Kf|L_q\| \le m^{1/q-1/p}\|Kf|L_p\|$. □

Note that the proven estimate $\|Kf|L_q\| \lesssim \|Kf|L_p\|$ for $0 < p \le q \le \infty$ resembles the relation $\|\cdot\|_{\ell^q} \le \|\cdot\|_{\ell^p}$ of the discrete Lebesgue quasi-norms. Hence, we can record that the embedding properties of the Wiener amalgams are analogous to those of the corresponding sequence spaces.

Another essential ingredient in the proof of Proposition 4.5.13 is the following technical lemma.

**Lemma 4.5.15.** *Let $\alpha \in [0,1]$, and assume that $K : \mathbb{X} \times \mathbb{X} \to \mathbb{C}$ is a kernel function and $H : \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+ \to [0,\infty)$ a measurable function such that*

$$|K(\mathbf{x},\mathbf{y})| \le H\big(A_{\alpha,t}^{-1} R_\eta(y-x), t^{-(1-\alpha)}\{\theta-\eta\}, u/t\big).$$

*Further let $\gamma \in \mathbb{R}$ be fixed and let us put $\mathbf{T} := [-\pi/2, \pi/2)$ and $(u,1)_+ := \max\{u,1\}$. Then the finiteness of the quantities*

$$\begin{aligned}
A_1 &:= \int_{\mathbb{R}_+}\int_{\mathbb{R}}\int_{\mathbb{R}^2} H(y,\theta,u)\max\{u,u^{-1}\}^{|\gamma|}(1+u)^2 \frac{dy\,d\theta\,du}{u^3},\\
A_2 &:= \operatorname*{ess\,sup}_{u\in\mathbb{R}_+} \int_{(u,1)_+^{1-\alpha}\mathbf{T}}\int_{\mathbb{R}^2} H(y,\theta,u)\max\{u,u^{-1}\}^{|\gamma|}(1+u)^{-2}\,dy\,d\theta,\\
A_3 &:= \int_{\mathbf{T}}\int_{\mathbb{R}^2} H(y,\theta,1)\,dy\,d\theta,
\end{aligned}$$

*implies*

$$K \in \mathcal{A}_{m_\nu} \quad \text{with} \quad \|K|\mathcal{A}_{m_\nu}\| \le 2\big(A_1 + \max\{4A_2, A_3\}\big),$$

*with $m_\nu$ denoting the bivariate weight associated to the weight $\nu = \nu_\gamma$ defined in* (4.31).

*Proof.* For fixed $\mathbf{x} = (x,\eta,t) \in \mathbb{X}$ we have

$$\int_{\mathbb{X}} |K(\mathbf{x},\mathbf{y})| m_\nu(\mathbf{x},\mathbf{y})\,d\mu(\mathbf{y}) \le 2\big(I_0(\mathbf{x}) + I_1(\mathbf{x})\big)$$

with the integrals

$$\begin{aligned}
I_0(\mathbf{x}) &:= \frac{1}{2}\int_0^1\int_{\mathbb{T}}\int_{\mathbb{R}^2} H\big(A_{\alpha,t}^{-1}R_\eta(y-x), t^{-(1-\alpha)}\{\theta-\eta\}, u/t\big)(t/u,u/t)_+^{|\gamma|}\frac{dy\,d\theta\,du}{u^3},\\
I_1(\mathbf{x}) &:= \frac{1}{2}\int_{\mathbb{T}}\int_{\mathbb{R}^2} H\big(A_{\alpha,t}^{-1}R_\eta(y-x), t^{-(1-\alpha)}\{\theta-\eta\}, 1/t\big)t^{-|\gamma|}\,dy\,d\theta.
\end{aligned}$$

Eliminating the brackets $\{\cdot\}$ in these integrals, they simplify to

$$\begin{aligned}
I_0(\mathbf{x}) &= \int_0^1\int_{\mathbf{T}}\int_{\mathbb{R}^2} H\big(A_{\alpha,t}^{-1}R_\eta(y-x), t^{-(1-\alpha)}\theta, u/t\big)(t/u,u/t)_+^{|\gamma|}\frac{dy\,d\theta\,du}{u^3},\\
I_1(\mathbf{x}) &= \int_{\mathbf{T}}\int_{\mathbb{R}^2} H\big(A_{\alpha,t}^{-1}R_\eta(y-x), t^{-(1-\alpha)}\theta, 1/t\big)t^{-|\gamma|}\,dy\,d\theta.
\end{aligned}$$

Substituting $y \mapsto R_\eta^{-1} A_{\alpha,t} y + x$ and $\theta \mapsto t^{1-\alpha}\theta$, we obtain

$$I_0(\mathbf{x}) = t^2 \int_0^1 \int_{t^{-(1-\alpha)}\mathbf{T}} \int_{\mathbb{R}^2} H(y,\theta,u/t)(t/u,u/t)_+^{|\gamma|} \frac{dy\, d\theta\, du}{u^3},$$
$$I_1(\mathbf{x}) = t^2 \int_{t^{-(1-\alpha)}\mathbf{T}} \int_{\mathbb{R}^2} H(y,\theta,1/t) t^{-|\gamma|}\, dy\, d\theta\,.$$

The substitution $u \mapsto tu$ further yields

$$I_0(\mathbf{x}) = \int_0^{1/t} \int_{t^{-(1-\alpha)}\mathbf{T}} \int_{\mathbb{R}^2} H(y,\theta,u)(u,1/u)_+^{|\gamma|} \frac{dy\, d\theta\, du}{u^3}\,.$$

Now we observe

$$\operatorname*{ess\,sup}_{\mathbf{x}\in\mathbb{X}} I_1(\mathbf{x}) = \max\Big\{ \operatorname*{ess\,sup}_{\mathbf{x}\in\mathbb{X}_0} I_1(\mathbf{x}), \operatorname*{ess\,sup}_{\mathbf{x}\in\mathbb{X}_1} I_1(\mathbf{x}) \Big\}$$

with the terms

$$\operatorname*{ess\,sup}_{\mathbf{x}\in\mathbb{X}_1} I_1(\mathbf{x}) = \int_{\mathbf{T}} \int_{\mathbb{R}^2} H(y,\theta,1)\, dy\, d\theta = A_3,$$
$$\operatorname*{ess\,sup}_{\mathbf{x}\in\mathbb{X}_0} I_1(\mathbf{x}) = \operatorname*{ess\,sup}_{t>1} t^{|\gamma|} \int_{t^{1-\alpha}\mathbf{T}} \int_{\mathbb{R}^2} H(y,\theta,t)\, \frac{dy\, d\theta}{t^2} \leq 4A_2\,.$$

Further, we have

$$\operatorname*{ess\,sup}_{\mathbf{x}\in\mathbb{X}} I_0(\mathbf{x}) = \max\Big\{ \operatorname*{ess\,sup}_{\mathbf{x}\in\mathbb{X}_0} I_0(\mathbf{x}), \operatorname*{ess\,sup}_{\mathbf{x}\in\mathbb{X}_1} I_0(\mathbf{x}) \Big\}$$
$$\leq \int_0^\infty \int_{\mathbb{R}} \int_{\mathbb{R}^2} H(y,\theta,u)(u,1/u)_+^{|\gamma|} \frac{dy\, d\theta\, du}{u^3} \leq A_1\,.$$

Analogously, for fixed $\mathbf{y} = (y,\theta,u) \in \mathbb{X}$ we get

$$\int_{\mathbb{X}} |K(\mathbf{x},\mathbf{y})| m_\nu(\mathbf{x},\mathbf{y})\, d\mu(\mathbf{x}) \leq 2\big(\tilde{I}_0(\mathbf{y}) + \tilde{I}_1(\mathbf{y})\big)$$

with the integrals

$$\tilde{I}_0(\mathbf{y}) := \frac{1}{2} \int_0^1 \int_{\mathbb{T}} \int_{\mathbb{R}^2} H(A_{\alpha,t}^{-1} R_\eta (y-x), t^{-(1-\alpha)}\{\theta-\eta\}, u/t)(t/u,u/t)_+^{|\gamma|} \frac{dx\, d\eta\, dt}{t^3},$$
$$\tilde{I}_1(\mathbf{y}) := \frac{1}{2} \int_{\mathbb{T}} \int_{\mathbb{R}^2} H(R_\eta(y-x), \{\theta-\eta\}, u) u^{-|\gamma|}\, dx\, d\eta.$$

We now first use the substitution $\eta \mapsto \theta - \eta$. Then, as before, we eliminate the brackets,

$$\begin{aligned}
\tilde{I}_0(\mathbf{y}) &= \frac{1}{2}\int_0^1 \int_{\mathbb{T}} \int_{\mathbb{R}^2} H\big(A_{\alpha,t}^{-1} R_{\theta-\eta}(y-x), t^{-(1-\alpha)}\{\eta\}, u/t\big)(t/u, u/t)_+^{|\gamma|} \frac{dx\, d\eta\, dt}{t^3} \\
&= \int_0^1 \int_{\mathbf{T}} \int_{\mathbb{R}^2} H\big(A_{\alpha,t}^{-1} R_{\theta-\eta}(y-x), t^{-(1-\alpha)}\eta, u/t\big)(t/u, u/t)_+^{|\gamma|} \frac{dx\, d\eta\, dt}{t^3}, \\
\tilde{I}_1(\mathbf{y}) &= \frac{1}{2}\int_{\mathbb{T}} \int_{\mathbb{R}^2} H\big(R_{\theta-\eta}(y-x), \{\eta\}, u\big) u^{-|\gamma|}\, dx\, d\eta \\
&= \int_{\mathbf{T}} \int_{\mathbb{R}^2} H\big(R_{\theta-\eta}(y-x), \eta, u\big) u^{-|\gamma|}\, dx\, d\eta.
\end{aligned}$$

The substitutions $x \mapsto y - R_{\theta-\eta}^{-1} A_{\alpha,t} x$ and $\eta \mapsto t^{1-\alpha}\eta$ yield

$$\begin{aligned}
\tilde{I}_0(\mathbf{y}) &= \int_0^1 \int_{t^{-(1-\alpha)}\mathbf{T}} \int_{\mathbb{R}^2} H(x, \eta, u/t)(t/u, u/t)_+^{|\gamma|} \frac{dx\, d\eta\, dt}{t}, \\
\tilde{I}_1(\mathbf{y}) &= \int_{\mathbf{T}} \int_{\mathbb{R}^2} H(x, \eta, u) u^{-|\gamma|}\, dx\, d\eta\,.
\end{aligned}$$

Finally, the substitution $t \mapsto u/t$ gives

$$\tilde{I}_0(\mathbf{y}) = \int_u^\infty \int_{(t/u)^{1-\alpha}\mathbf{T}} \int_{\mathbb{R}^2} H(x, \eta, t)(t, 1/t)_+^{|\gamma|} \frac{dx\, d\eta\, dt}{t}\,.$$

We obtain

$$\operatorname*{ess\,sup}_{\mathbf{y}\in\mathbb{X}} \tilde{I}_1(\mathbf{y}) = \max\Big\{ \operatorname*{ess\,sup}_{\mathbf{y}\in\mathbb{X}_0} \tilde{I}_1(\mathbf{y}), \operatorname*{ess\,sup}_{\mathbf{y}\in\mathbb{X}_1} \tilde{I}_1(\mathbf{y}) \Big\}$$

with

$$\begin{aligned}
&\operatorname*{ess\,sup}_{\mathbf{y}\in\mathbb{X}_1} \tilde{I}_1(\mathbf{y}) = \int_{\mathbf{T}} \int_{\mathbb{R}^2} H(x, \eta, 1)\, dx\, d\eta = A_3, \\
&\operatorname*{ess\,sup}_{\mathbf{y}\in\mathbb{X}_0} \tilde{I}_1(\mathbf{y}) = \operatorname*{ess\,sup}_{0<u<1} \int_{\mathbf{T}} \int_{\mathbb{R}^2} H(x, \eta, u) u^{-|\gamma|}\, dx\, d\eta \le 4A_2\,.
\end{aligned}$$

For $\operatorname*{ess\,sup}_{\mathbf{y}\in\mathbb{X}} \tilde{I}_0(\mathbf{y}) = \max\Big\{ \operatorname*{ess\,sup}_{\mathbf{y}\in\mathbb{X}_0} \tilde{I}_0(\mathbf{y}), \operatorname*{ess\,sup}_{\mathbf{y}\in\mathbb{X}_1} \tilde{I}_0(\mathbf{y}) \Big\}$ we get

$$\operatorname*{ess\,sup}_{\mathbf{y}\in\mathbb{X}} \tilde{I}_0(\mathbf{y}) \le \int_{\mathbb{R}_+} \int_{\mathbb{R}} \int_{\mathbb{R}^2} H(x, \eta, t)(t, 1/t)_+^{|\gamma|} \frac{dx\, d\eta\, dt}{t} \le A_1\,.$$

This finishes the proof. $\square$

We next formulate a sufficient criterion ensuring that the kernel $\mathcal{G}_N$ operates continuously on the function spaces $L^s_{p,q}(\mathbb{X})$ and $\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$. As in Proposition 4.5.13, it can be expected that a sufficiently large $N$ provides such a guarantee, since, intuitively, the larger $N$ the closer $\mathcal{G}_N$ is to the identity operator.

**Proposition 4.5.16.** *Let $\alpha \in [0,1]$, $0 < p,q < \infty$, $r = \min\{1,p,q\}$, and $s \in \mathbb{R}$. If $N > \frac{2}{r} + |s|$ then $\mathcal{G}_N$ operates continuously on $\mathbb{L}^{\alpha,s}_{p,q}(\mathbb{X})$. In the Banach case, i.e., when $r = 1$, the kernel $\mathcal{G}_N$ also operates continuously on $L^s_{p,q}(\mathbb{X})$.*

*Proof.* For every $N > 0$ and all $\mathbf{x}, \mathbf{y} \in \mathbb{X}$ we have by Corollary 2.2.22 and Lemma 4.5.10

$$\sup_{(\mathbf{a},\mathbf{b}) \in U^\alpha_1(\mathbf{x}) \times U^\alpha_1(\mathbf{y})} \mathcal{G}_N(\mathbf{a},\mathbf{b}) \lesssim \mathcal{G}_N(\mathbf{x},\mathbf{y}) \lesssim \mathcal{H}_N(\mathbf{x},\mathbf{y}),$$

and $\mathcal{H}_N(\mathbf{x},\mathbf{y}) = \mathcal{H}_N((x,\eta,t),(y,\theta,u))$ has the representation

$$\mathcal{H}_N(\mathbf{x},\mathbf{y}) = H_N(A^{-1}_{\alpha,t} R_\eta (y-x), t^{-(1-\alpha)}\{\theta - \eta\}, u/t)$$

with the function $H_N$ from (4.59). Hence $\mathcal{G}_N$ is of the required form to apply Lemma 4.5.17 or Lemma 4.5.18.

It only remains to show that the quantities $B_i$, or $\tilde{B}_i$ respectively, are finite for each $i \in \{1,\dots,4\}$. This task is simplified by the following observation. If $q \geq 1$ we have $\ell^1 \hookrightarrow \ell^q$ and thus – as a consequence of Lemma 4.5.14 and Lemma 4.5.12 – using $J := [\frac{1}{2}, 2]$ as a window

$$\begin{aligned} B_2 &\leq \Big( \int_1^\infty \Big( \operatorname*{ess\,sup}_{v \in uJ} \int_{v^{1-\alpha}\mathbf{T}} \int_{\mathbb{R}^2} v^s H_N(y,\theta,v) \frac{dy\,d\theta}{v^2} \Big)^q \frac{du}{u} \Big)^{1/q} \\ &\lesssim \int_{\mathbb{R}_+} \int_{\mathbb{R}} \int_{\mathbb{R}^2} \operatorname*{ess\,sup}_{v \in uJ} v^s H_N(y,\theta,v) \frac{dy\,d\theta}{v^2} \frac{du}{u} \lesssim B_1. \end{aligned}$$

Similarly, one can prove $B_3 \lesssim B_1$, and analogously also $\tilde{B}_2 \lesssim \tilde{B}_1$ and $\tilde{B}_3 \lesssim \tilde{B}_1$.

Finally, since $N > 2/r + |s|$, according to Lemma 4.5.11, we have

$$\begin{aligned} B_1, \tilde{B}_1 &\leq \int_{\mathbb{R}_+} \int_{\mathbb{R}} \int_{\mathbb{R}^2} H_N(y,\theta,u)^r \max\{u, u^{-1}\}^{|s|r} (u+1)^2 \frac{dy\,d\theta\,du}{u^3} < \infty\,, \\ B_4, \tilde{B}_4 &\leq \int_{\mathbb{R}} \int_{\mathbb{R}^2} H_N(y,\theta,1)^r \,dy\,d\theta < \infty\,. \end{aligned}$$

As a consequence, all quantities $B_1, B_4, \tilde{B}_1, \tilde{B}_4$ are finite if $N > 2/r + |s|$, and the statement is proven. □

The proof of the previous proposition builds upon the following two lemmas.

**Lemma 4.5.17.** *Let $\alpha \in [0,1]$, $1 \leq p,q < \infty$, and $s \in \mathbb{R}$ be fixed. Further, put $\mathbf{T} := [-\pi/2, \pi/2)$ and assume that $K : \mathbb{X} \times \mathbb{X} \to \mathbb{C}$ is a kernel with the property*

$$|K(\mathbf{x},\mathbf{y})| \leq H(A^{-1}_{\alpha,t} R_\eta (y-x), t^{-(1-\alpha)}\{\theta - \eta\}, u/t)$$

*for some measurable function $H : \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+ \to [0,\infty)$. Then $K$ is a well-defined bounded operator*

$$K : L^s_{p,q}(\mathbb{X}) \to L^s_{p,q}(\mathbb{X}) \quad \text{with} \quad \|K|L^s_{p,q} \to L^s_{p,q}\| \lesssim B_1 + B_2 + B_3 + B_4,$$

*provided that, with $q' := q/(q-1)$ denoting the dual exponent of $q$,*

$$B_1 := \int_0^\infty \int_{\mathbb{R}} \int_{\mathbb{R}^2} u^s H(y,\theta,u) \frac{dy\,d\theta\,du}{u^3} < \infty\,,$$

$$B_2 := \Big( \int_1^\infty \Big( \int_{u^{1-\alpha}\mathbf{T}} \int_{\mathbb{R}^2} u^s H(y,\theta,u) \frac{dy\,d\theta}{u^2} \Big)^q \frac{du}{u} \Big)^{1/q} < \infty,$$

$$B_3 := \Big( \int_0^1 \Big( \int_{\mathbf{T}} \int_{\mathbb{R}^2} u^s H(y,\theta,u) \frac{dy\,d\theta}{u^2} \Big)^{q'} \frac{du}{u} \Big)^{1/q'} < \infty\,,$$

$$B_4 := \int_{\mathbf{T}} \int_{\mathbb{R}^2} H(y,\theta,1)\,dy\,d\theta < \infty\,.$$

*Proof.* Due to Lemma 4.5.3 and the solidity of $L^s_{p,q}$, we can without loss of generality assume that the kernel $K$ is non-negative. In this special case, for any measurable non-negative function $F : \mathbb{X} \to [0,\infty)$, the integral

$$KF(\mathbf{x}) := \int_{\mathbb{X}} K(\mathbf{x},\mathbf{y}) F(\mathbf{y})\,d\mu(\mathbf{y})\,, \quad \mathbf{x} \in \mathbb{X},$$

has a well-defined value in the extended range $[0,\infty]$ at almost all points $\mathbf{x} \in \mathbb{X}$. One thus obtains a measurable function $KF$ on $\mathbb{X}$ with the target set $[0,\infty]$. The investigation below will further show that the additional assumption $F \in L^s_{p,q}$ ensures $KF \in L^s_{p,q}$ with the estimate $\|KF|L^s_{p,q}\| \lesssim (B_1+B_2+B_3+B_4)\|F|L^s_{p,q}\|$.

For an arbitrary, not necessarily non-negative, function $F \in L^s_{p,q}$ we can then argue as follows. Since $|F| \in L^s_{p,q}$ is non-negative, by the above, we have $K|F| \in L^s_{p,q}$, which in particular entails $K|F|(\mathbf{x}) < \infty$ for almost all $\mathbf{x} \in \mathbb{X}$. At those points, $KF(\mathbf{x})$ is contained in $\mathbb{C}$, giving rise to a measurable function $KF : \mathbb{X} \to \mathbb{C}$. Further, since $|KF| \le K|F|$ almost everywhere, $KF \in L^s_{p,q}$ holds true by solidity and

$$\|KF|L^s_{p,q}\| \le \|K|F||L^s_{p,q}\| \le (B_1+B_2+B_3+B_4)\||F||L^s_{p,q}\| = (B_1+B_2+B_3+B_4)\|F|L^s_{p,q}\|.$$

This proves the assertion of the lemma. All, that remains to be shown, is that $KF \in L^s_{p,q}$ with $\|KF|L^s_{p,q}\| \le (B_1+B_2+B_3+B_4)\|F|L^s_{p,q}\|$ holds true for every non-negative function $F \in L^s_{p,q}$. The proof of this claim is split into several steps.

**Step 1:** First we estimate the functions $K[F\mathcal{X}_{\mathbb{X}_0}]$. Thereby we transfer the integration domain from $\mathbb{P}$ to $\mathbf{P}$ via the canonical projection $\mathfrak{p} : \mathbb{P} \to \mathbf{P}$ defined in (2.12). For this, it is useful to associate to $F : \mathbb{P} \to [0,\infty)$ the auxiliary function $\breve{F} : \mathbf{P} \to [0,\infty)$ by

$$\breve{F}(y,\{\theta\},u) := F(y,\theta,u) + F(y,\theta+\pi,u)\,, \quad \theta \in [0,\pi),$$

where $\{\cdot\} = \{\cdot\}_{\mathbf{T}}$ denotes the projective bracket from (2.9).

Let now $\mathbf{x} = (x, \eta, t) \in \mathbb{X}$ be fixed. We then have

$$\begin{aligned}
K[F\mathcal{X}_{\mathbb{X}_0}](\mathbf{x}) &= \int\limits_{\mathbb{X}_0} K\big((x,\eta,t),(y,\theta,u)\big) F(y,\theta,u)\, d\mu_0(y,\theta,u) \\
&\leq \int\limits_{\mathbb{X}_0} H(A_{\alpha,t}^{-1} R_\eta (y-x), t^{-(1-\alpha)}\{\theta-\eta\}, u/t) F(y,\theta,u)\, d\mu_0(y,\theta,u) \\
&= \int\limits_0^1 \int\limits_{\mathbf{T}} \int\limits_{\mathbb{R}^2} H(A_{\alpha,t}^{-1} R_\eta (y-x), t^{-(1-\alpha)}\{\theta-\eta\}, u/t) \breve{F}(y,\theta,u)\, \frac{dy\, d\theta\, du}{u^3}.
\end{aligned}$$

Note $\{\{a+b\}+c\} = \{a+b+c\}$. With the substitutions $y \mapsto R_\eta^{-1} A_{\alpha,t} y + x$, $\theta \mapsto \{t^{1-\alpha}\theta + \eta\}$, $u \mapsto tu$, we hence arrive at

$$K[F\mathcal{X}_{\mathbb{X}_0}](\mathbf{x}) \leq \int\limits_0^{1/t} \int\limits_{t^{-(1-\alpha)}\mathbf{T}} \int\limits_{\mathbb{R}^2} H(y,\theta,u) \breve{F}(R_\eta^{-1} A_{\alpha,t} y + x, \{t^{1-\alpha}\theta+\eta\}, tu)\, \frac{dy\, d\theta\, du}{u^3}.$$

Analogously, one shows with the substitutions $y \mapsto R_\eta^{-1} y + x$, $\theta \mapsto \{\theta + \eta\}$,

$$\begin{aligned}
K[F\mathcal{X}_{\mathbb{X}_1}](\mathbf{x}) &= \int\limits_{\mathbb{X}_1} K\big((x,\eta,t),(y,\theta,1)\big) F(y,\theta,1)\, d\mu_1(y,\theta,1) \\
&\leq \int\limits_{\mathbb{X}_1} H(A_{\alpha,t}^{-1} R_\eta (y-x), t^{-(1-\alpha)}\{\theta-\eta\}, 1/t) F(y,\theta,1)\, d\mu_1(y,\theta,1) \\
&= \int\limits_{\mathbf{T}} \int\limits_{\mathbb{R}^2} H(A_{\alpha,t}^{-1} R_\eta (y-x), t^{-(1-\alpha)}\{\theta-\eta\}, 1/t) \breve{F}(y,\theta,1)\, dy\, d\theta \\
&= \int\limits_{\mathbf{T}} \int\limits_{\mathbb{R}^2} H(A_{\alpha,t}^{-1} y, t^{-(1-\alpha)}\theta, 1/t) \breve{F}(R_\eta^{-1} y + x, \{\theta+\eta\}, 1)\, dy\, d\theta.
\end{aligned}$$

**Step 2:** For each fixed $t \in (0,1]$, we now decompose

$$\begin{aligned}
&\Big( \int\limits_{\mathbb{T}} \|KF(\cdot,\eta,t)|L_p\|^q\, d\eta \Big)^{1/q} \\
&\asymp \Big( \int\limits_{\mathbb{T}} \|K[F\mathcal{X}_{\mathbb{X}_0}](\cdot,\eta,t)|L_p\|^q\, d\eta \Big)^{1/q} + \Big( \int\limits_{\mathbb{T}} \|K[F\mathcal{X}_{\mathbb{X}_1}](\cdot,\eta,t)|L_p\|^q\, d\eta \Big)^{1/q} =: T_0(t) + T_1(t).
\end{aligned}$$

Then we estimate $T_0(t)$ and $T_1(t)$, using Step 1 and the continuous Minkowski inequality,

$$\begin{aligned}
T_0(t) &\leq \Big( \int\limits_{\mathbb{T}} \Big\| \int\limits_0^{1/t} \int\limits_{t^{-(1-\alpha)}\mathbf{T}} \int\limits_{\mathbb{R}^2} H(y,\theta,u) \breve{F}(R_\eta^{-1} A_{\alpha,t} y + \cdot, \{t^{1-\alpha}\theta+\eta\}, tu)\, \frac{dy\, d\theta\, du}{u^3} \Big| L_p \Big\|^q d\eta \Big)^{1/q} \\
&\leq \int\limits_0^{1/t} \int\limits_{t^{-(1-\alpha)}\mathbf{T}} \int\limits_{\mathbb{R}^2} H(y,\theta,u) \Big( \int\limits_{\mathbb{T}} \Big\| \breve{F}(R_\eta^{-1} A_{\alpha,t} y + \cdot, \{t^{1-\alpha}\theta+\eta\}, tu) \Big| L_p \Big\|^q d\eta \Big)^{1/q} \frac{dy\, d\theta\, du}{u^3} \\
&= \int\limits_0^{1/t} \int\limits_{t^{-(1-\alpha)}\mathbf{T}} \int\limits_{\mathbb{R}^2} H(y,\theta,u) \Big( \int\limits_{\mathbf{T}} \|\breve{F}(\cdot,\eta,tu)|L_p\|^q\, d\eta \Big)^{1/q} \frac{dy\, d\theta\, du}{u^3},
\end{aligned}$$

and analogously

$$T_1(t) \le \Big( \int\limits_{\mathbb{T}} \Big\| \int\limits_{\mathbf{T}} \int\limits_{\mathbb{R}^2} H(A_{\alpha,t}^{-1}y, t^{-(1-\alpha)}\theta, 1/t) \breve{F}(R_\eta^{-1}y + \cdot, \{\theta+\eta\}, 1)\, dy\, d\theta \Big| L_p \Big\|^q d\eta \Big)^{1/q}$$

$$\le \int\limits_{\mathbf{T}} \int\limits_{\mathbb{R}^2} H(A_{\alpha,t}^{-1}y, t^{-(1-\alpha)}\theta, 1/t) \Big( \int\limits_{\mathbb{T}} \Big\| \breve{F}(R_\eta^{-1}y + \cdot, \{\theta+\eta\}, 1) \Big| L_p \Big\|^q d\eta \Big)^{1/q} dy\, d\theta$$

$$= \int\limits_{\mathbf{T}} \int\limits_{\mathbb{R}^2} H(A_{\alpha,t}^{-1}y, t^{-(1-\alpha)}\theta, 1/t)\, dy\, d\theta \Big( \int\limits_{\mathbf{T}} \|\breve{F}(\cdot, \eta, 1)|L_p\|^q\, d\eta \Big)^{1/q}.$$

**Step 3:** Finally, we decompose the kernel $K = \sum_{i=0}^1 \sum_{j=0}^1 K_{i,j}$ with

$$K_{i,j}(\mathbf{x}, \mathbf{y}) := K(\mathbf{x}, \mathbf{y}) \mathcal{X}_{\mathbb{X}_i \times \mathbb{X}_j}(\mathbf{x}, \mathbf{y}) \quad \text{for } (i,j) \in \{0,1\}^2.$$

To finish the proof, it then suffices to check that for every $(i,j) \in \{0,1\}^2$

$$S_{i,j} := \|K_{i,j} F | L_{p,q}^s\| \lesssim \|F \mathcal{X}_{\mathbb{X}_j} | L_{p,q}^s\|.$$

We start with $S_{0,1}$ and $S_{1,1}$ and observe

$$S_{1,1} = T_1(1) \quad \text{and} \quad S_{0,1} = \Big( \int\limits_0^1 t^{-sq} T_1(t)^q \frac{dt}{t} \Big)^{1/q}.$$

Plugging in the estimates from Step 2, we get

$$S_{1,1} \le \int\limits_{\mathbf{T}} \int\limits_{\mathbb{R}^2} H(y, \theta, 1)\, dy\, d\theta \Big( \int\limits_{\mathbf{T}} \|\breve{F}(\cdot, \eta, 1)|L_p\|^q\, d\eta \Big)^{1/q}.$$

Further, taking into account

$$\|F \mathcal{X}_{\mathbb{X}_1} | L_{p,q}^s\| \asymp \Big( \int\limits_{\mathbf{T}} \|\breve{F}(\cdot, \eta, 1)|L_p\|^q\, d\eta \Big)^{1/q}, \tag{4.60}$$

a relation proved in Step 4 below, this yields

$$S_{1,1} \lesssim B_4 \cdot \|F \mathcal{X}_{\mathbb{X}_1} | L_{p,q}^s\| \quad \text{with} \quad B_4 = \int\limits_{\mathbf{T}} \int\limits_{\mathbb{R}^2} H(y, \theta, 1)\, dy\, d\theta.$$

For $S_{0,1}$ we derive from Step 2

$$S_{0,1} \le \Big( \int\limits_0^1 t^{-sq} \Big( \int\limits_{\mathbf{T}} \int\limits_{\mathbb{R}^2} H(A_{\alpha,t}^{-1}y, t^{-(1-\alpha)}\theta, 1/t)\, dy\, d\theta \Big( \int\limits_{\mathbf{T}} \|\breve{F}(\cdot, \eta, 1)|L_p\|^q\, d\eta \Big)^{1/q} \Big)^q \frac{dt}{t} \Big)^{1/q}.$$

We deduce, again with (4.60),

$$S_{0,1} \lesssim B_2 \cdot \|F \mathcal{X}_{\mathbb{X}_1} | L_{p,q}^s\|$$

with

$$B_2 = \Big( \int\limits_0^1 t^{-sq} \Big( \int\limits_{\mathbf{T}} \int\limits_{\mathbb{R}^2} H\big(A_{\alpha,t}^{-1} y, t^{-(1-\alpha)}\theta, 1/t\big)\, dy\, d\theta \Big)^q \frac{dt}{t} \Big)^{1/q}$$
$$= \Big( \int\limits_1^\infty \Big( \int\limits_{t^{1-\alpha}\mathbf{T}} \int\limits_{\mathbb{R}^2} t^s H(y,\theta,t)\, \frac{dy\, d\theta}{t^2} \Big)^q \frac{dt}{t} \Big)^{1/q}.$$

Next, we turn to $S_{1,0}$ and $S_{0,0}$ given by

$$S_{1,0} = T_0(1) \quad \text{and} \quad S_{0,0} = \Big( \int\limits_0^1 t^{-sq} T_0(t)^q \frac{dt}{t} \Big)^{1/q}.$$

To estimate $S_{1,0}$, we use the results of Step 2 and Hölder's inequality, where $q'$ shall denote the dual exponent of $q$ satisfying $1/q + 1/q' = 1$. We obtain

$$S_{1,0} \le \int\limits_0^1 \int\limits_{\mathbf{T}} \int\limits_{\mathbb{R}^2} H(y,\theta,u) \Big( \int\limits_{\mathbf{T}} \|\breve{F}(\cdot,\eta,u)|L_p\|^q\, d\eta \Big)^{1/q} \frac{dy\, d\theta\, du}{u^3}$$
$$= \int\limits_0^1 \Big( \int\limits_{\mathbf{T}} \int\limits_{\mathbb{R}^2} H(y,\theta,u)\, \frac{dy\, d\theta}{u^2} \Big) \cdot \Big( \int\limits_{\mathbf{T}} \|\breve{F}(\cdot,\eta,u)|L_p\|^q\, d\eta \Big)^{1/q} \frac{du}{u}$$
$$\le \Big( \int\limits_0^1 u^{q's} \Big( \int\limits_{\mathbf{T}} \int\limits_{\mathbb{R}^2} H(y,\theta,u)\, \frac{dy\, d\theta}{u^2} \Big)^{q'} \frac{du}{u} \Big)^{1/q'} \cdot \Big( \int_0^1 u^{-sq} \int\limits_{\mathbf{T}} \|\breve{F}(\cdot,\eta,u)|L_p\|^q\, \frac{d\eta\, du}{u} \Big)^{1/q}.$$

Using the relation

$$\|F\mathcal{X}_{\mathbb{X}_0}|L_{p,q}^s\| \asymp \Big( \int\limits_0^1 \int\limits_{\mathbf{T}} t^{-sq} \|\breve{F}(\cdot,\eta,t)|L_p\|^q\, \frac{d\eta\, dt}{t} \Big)^{1/q}, \tag{4.61}$$

whose proof is outsourced to Step 4, we arrive at

$$S_{1,0} \lesssim B_3 \cdot \|F\mathcal{X}_{\mathbb{X}_0}|L_{p,q}^s\| \quad \text{with} \quad B_3 = \Big( \int\limits_0^1 \Big( \int\limits_{\mathbf{T}} \int\limits_{\mathbb{R}^2} u^s H(y,\theta,u) \frac{dy\, d\theta}{u^2} \Big)^{q'} \frac{du}{u} \Big)^{1/q'}.$$

Last but not least, we estimate $S_{0,0}$. Note that $\mathcal{X}_{(0,1/t)}(u) = \mathcal{X}_{(0,1/u)}(t)$ for $t, u \in \mathbb{R}_+$.

$$
\begin{aligned}
S_{0,0} &\le \Big( \int_0^1 t^{-sq} \Big( \int_0^{1/t} \int_{t^{-(1-\alpha)}\mathbf{T}} \int_{\mathbb{R}^2} H(y,\theta,u) \Big( \int_{\mathbf{T}} \|\breve{F}(\cdot,\eta,tu)|L_p\|^q \, d\eta \Big)^{1/q} \frac{dy\, d\theta\, du}{u^3} \Big)^q \frac{dt}{t} \Big)^{1/q} \\
&\le \Big( \int_0^1 t^{-sq} \Big( \int_0^{\infty} \int_{\mathbb{R}} \int_{\mathbb{R}^2} \mathcal{X}_{(0,1/t)}(u) H(y,\theta,u) \Big( \int_{\mathbf{T}} \|\breve{F}(\cdot,\eta,tu)|L_p\|^q \, d\eta \Big)^{1/q} \frac{dy\, d\theta\, du}{u^3} \Big)^q \frac{dt}{t} \Big)^{1/q} \\
&\le \int_0^{\infty} \int_{\mathbb{R}} \int_{\mathbb{R}^2} H(y,\theta,u) \Big( \int_0^1 \int_{\mathbf{T}} t^{-sq} \mathcal{X}_{(0,1/u)}(t) \|\breve{F}(\cdot,\eta,tu)|L_p\|^q \, d\eta \frac{dt}{t} \Big)^{1/q} \frac{dy\, d\theta\, du}{u^3} \\
&= \int_0^{\infty} \int_{\mathbb{R}} \int_{\mathbb{R}^2} H(y,\theta,u) \Big( \int_0^{1/u} \int_{\mathbf{T}} t^{-sq} \|\breve{F}(\cdot,\eta,tu)|L_p\|^q \, \frac{d\eta\, dt}{t} \Big)^{1/q} \frac{dy\, d\theta\, du}{u^3} \\
&= \int_0^{\infty} \int_{\mathbb{R}} \int_{\mathbb{R}^2} H(y,\theta,u) u^s \Big( \int_0^{1} \int_{\mathbf{T}} t^{-sq} \|\breve{F}(\cdot,\eta,t)|L_p\|^q \, \frac{d\eta\, dt}{t} \Big)^{1/q} \frac{dy\, d\theta\, du}{u^3}.
\end{aligned}
$$

With (4.61), this leads to

$$
S_{0,0} \lesssim B_1 \cdot \|F\mathcal{X}_{\mathbb{X}_0}|L^s_{p,q}\| \quad \text{with} \quad B_1 = \int_0^{\infty} \int_{\mathbb{R}} \int_{\mathbb{R}^2} u^s H(y,\theta,u) \frac{dy\, d\theta\, du}{u^3}.
$$

**Step 4:** It remains to show (4.60) and (4.61). First, observe that due to the non-negativity of $F$ for every $\eta \in [0,\pi)$ and $t \in (0,1]$,

$$
\|\breve{F}(\cdot,\{\eta\},t)|L_p\| \asymp \|F(\cdot,\eta,t)|L_p\| + \|F(\cdot,\eta+\pi,t)|L_p\|.
$$

As a consequence, for every fixed $t \in (0,1]$

$$
\begin{aligned}
\Big( \int_{\mathbf{T}} \|\breve{F}(\cdot,\eta,t)|L_p\|^q \, d\eta \Big)^{1/q} &= \Big( \int_0^{\pi} \|\breve{F}(\cdot,\{\eta\},t)|L_p\|^q \, d\eta \Big)^{1/q} \\
&\asymp \Big( \int_0^{\pi} \|F(\cdot,\eta,t)|L_p\|^q + \|F(\cdot,\eta+\pi,t)|L_p\|^q \, d\eta \Big)^{1/q} \\
&\asymp \Big( \int_0^{2\pi} \|F(\cdot,\eta,t)|L_p\|^q \, d\eta \Big)^{1/q} =: I(t).
\end{aligned}
$$

Plugging in $t = 1$ yields (4.60) since $I(1) = \|F\mathcal{X}_{\mathbb{X}_1}|L^s_{p,q}\|$. For $t \in (0,1)$ we get (4.61), since

$$
\|F\mathcal{X}_{\mathbb{X}_0}|L^s_{p,q}\| = \Big( \int_0^1 t^{-sq} I(t)^q \, \frac{dt}{t} \Big)^{1/q} \asymp \mathrm{rhs}(4.61).
$$

□

Under slightly stronger assumptions on the kernel $K$, we can formulate the following companion result to Lemma 4.5.17 which is also valid in the quasi-Banach range.

**Lemma 4.5.18.** *Let $\alpha \in [0,1]$, and assume that $K : \mathbb{X} \times \mathbb{X} \to \mathbb{C}$ is a kernel with the property*

$$\operatorname*{ess\,sup}_{(\mathbf{a},\mathbf{b})\in U_1^\alpha(\mathbf{x})\times U_1^\alpha(\mathbf{y})} |K(\mathbf{a},\mathbf{b})| \le H\big(A_{\alpha,t}^{-1}R_\eta(y-x), t^{-(1-\alpha)}\{\theta-\eta\}, u/t\big)$$

*for some measurable function $H : \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+ \to [0,\infty)$. Assuming $s \in \mathbb{R}$, $0 < p, q < \infty$, the kernel $K$ is a well-defined bounded operator*

$$K : \mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X}) \to \mathbb{L}_{p,q}^{\alpha,s}(\mathbb{X}) \quad \textit{with} \quad \|K|\mathbb{L}_{p,q}^{\alpha,s} \to \mathbb{L}_{p,q}^{\alpha,s}\| \lesssim \tilde{B}_1 + \tilde{B}_2 + \tilde{B}_3 + \tilde{B}_4,$$

*provided that*

$$\begin{aligned}
\tilde{B}_1 &:= \int_0^\infty \int_{\mathbb{R}} \int_{\mathbb{R}^2} u^{sr} H(y,\theta,u)^r \frac{dy\,d\theta\,du}{u^3} < \infty\,,\\
\tilde{B}_2 &:= \bigg( \int_1^\infty \bigg( \int_{u^{1-\alpha}\mathbf{T}} \int_{\mathbb{R}^2} u^{sr} H(y,\theta,u)^r \frac{dy\,d\theta}{u^2} \bigg)^{\tilde{q}} \frac{du}{u} \bigg)^{1/\tilde{q}} < \infty,\\
\tilde{B}_3 &:= \bigg( \int_0^1 \bigg( \int_{\mathbf{T}} \int_{\mathbb{R}^2} u^{sr} H(y,\theta,u)^r \,\frac{dy\,d\theta}{u^2} \bigg)^{\tilde{q}'} \frac{du}{u} \bigg)^{1/\tilde{q}'} < \infty\,,\\
\tilde{B}_4 &:= \int_{\mathbf{T}} \int_{\mathbb{R}^2} H(y,\theta,1)^r \,dy\,d\theta < \infty\,,
\end{aligned}$$

*where $r := \min\{1,p,q\}$, $\mathbf{T} := [-\pi/2,\pi/2)$, $\tilde{q} := q/r$, and $\tilde{q}' := \tilde{q}/(\tilde{q}-1)$.*

*Proof.* Recall the Wiener maximal operator $\tilde{\mathbf{W}}^\alpha = \tilde{\mathbf{W}}_1^\alpha$ defined in (4.17) for any measurable function $F : \mathbb{X} \to \mathbb{C}$. We subsequently use the abbreviation $\mathbf{F} := \tilde{\mathbf{W}}^\alpha F$. For convenience, we also introduce the kernel

$$\mathbf{K}(\mathbf{x},\mathbf{y}) := \operatorname*{ess\,sup}_{(\mathbf{a},\mathbf{b})\in U_1^\alpha(\mathbf{x})\times U_1^\alpha(\mathbf{y})} |K(\mathbf{a},\mathbf{b})|\,, \quad (\mathbf{x},\mathbf{y}) \in \mathbb{X} \times \mathbb{X}.$$

We want to show that for each $F \in \mathbb{L}_{p,q}^{\alpha,s}$ the function $KF$ is well-defined, an element of $\mathbb{L}_{p,q}^{\alpha,s}$, and satisfies

$$\|KF|\mathbb{L}_{p,q}^{\alpha,s}\| = \|\tilde{\mathbf{W}}^\alpha[KF]|L_{p,q}^s\| \lesssim (\tilde{B}_1 + \tilde{B}_2 + \tilde{B}_3 + \tilde{B}_4)\|F|\mathbb{L}_{p,q}^{\alpha,s}\|.$$

With a trick, we can utilize the previous lemma for the proof, even in the quasi-Banach range. For this, in a first step which is only relevant in the quasi-Banach setting, we 'elevate' the parameters of $\mathbb{L}_{p,q}^{\alpha,s}$ into the Banach range: We introduce $\tilde{p} := p/r$, $\tilde{q} := q/r$, $\tilde{s} = sr$, where $r = \min\{1,p,q\}$, and observe that for any measurable function $F : \mathbb{X} \to \mathbb{C}$ we have the equivalence

$$\|F|L_{p,q}^s\|^r = \bigg[ \bigg( \int_{\mathbb{T}} \|F(\cdot,\eta,1)|L_p\|^q \,d\eta \bigg)^{1/q} + \bigg( \int_0^1 \int_{\mathbb{T}} t^{-sq} \|F(\cdot,\eta,t)|L_p\|^q \frac{d\eta\,dt}{t} \bigg)^{1/q} \bigg]^r \tag{4.62}$$

$$\asymp \bigg( \int_{\mathbb{T}} \||F(\cdot,\eta,1)|^r|L_{\tilde{p}}\|^{\tilde{q}} \,d\eta \bigg)^{1/\tilde{q}} + \bigg( \int_0^1 \int_{\mathbb{T}} t^{-\tilde{s}\tilde{q}} \||F(\cdot,\eta,t)|^r|L_{\tilde{p}}\|^{\tilde{q}} \frac{d\eta\,dt}{t} \bigg)^{1/\tilde{q}} = \||F|^r|L_{\tilde{p},\tilde{q}}^{\tilde{s}}\|.$$

As a consequence, for $F \in \mathbb{L}^{\alpha,s}_{p,q}$ we have $\mathbf{F} \in L^s_{p,q}$ and thus $\mathbf{F}^r \in L^{\tilde{s}}_{\tilde{p},\tilde{q}}$. Note that, with $\tilde{p} \geq 1$ and $\tilde{q} \geq 1$, the space $L^{\tilde{s}}_{\tilde{p},\tilde{q}}$ is a Banach space.

As in Lemma 4.5.17, we now assume that $K$ and $F$ are non-negative, such that $KF$ is a well-defined measurable function. From the previous considerations, we obtain

$$\|KF|\mathbb{L}^{\alpha,s}_{p,q}\|^r = \|\tilde{\mathbf{W}}^\alpha[KF]|L^s_{p,q}\|^r \asymp \||\tilde{\mathbf{W}}^\alpha[KF]|^r|L^{\tilde{s}}_{\tilde{p},\tilde{q}}\|. \tag{4.63}$$

Next, using Lemma 4.5.14 with a Wiener amalgam embedding of the type $\ell^r \hookrightarrow \ell^1$, we obtain for almost every $\mathbf{x} \in \mathbb{X}$

$$\begin{aligned}
|\tilde{\mathbf{W}}^\alpha[KF](\mathbf{x})|^r &\leq \operatorname*{ess\,sup}_{\mathbf{a}\in U^\alpha_1(\mathbf{x})} \Big( \int_{\mathbb{X}} \operatorname*{ess\,sup}_{\mathbf{b}\in U^\alpha_1(\mathbf{y})} |K(\mathbf{a},\mathbf{b})F(\mathbf{b})| \, d\mu(\mathbf{y}) \Big)^r \\
&\lesssim \operatorname*{ess\,sup}_{\mathbf{a}\in U^\alpha_1(\mathbf{x})} \int_{\mathbb{X}} \Big( \operatorname*{ess\,sup}_{\mathbf{b}\in U^\alpha_1(\mathbf{y})} |K(\mathbf{a},\mathbf{b})F(\mathbf{b})| \Big)^r \, d\mu(\mathbf{y}) \\
&\leq \int_{\mathbb{X}} (\mathbf{K}(\mathbf{x},\mathbf{y})\mathbf{F}(\mathbf{y}))^r \, d\mu(\mathbf{y}) = \mathbf{K}^r[\mathbf{F}^r](\mathbf{x}).
\end{aligned}$$

Hereby, the implicit constant is independent of $\mathbf{x} \in \mathbb{X}$, since $\mu(U^\alpha_1(\mathbf{x})) \asymp \mu(U'^{,\alpha}_1(\mathbf{x})) \asymp 1$ for all $\mathbf{x} \in \mathbb{X}$ according to Corollary 4.2.12.

Together with (4.63), the last estimate yields

$$\|KF|\mathbb{L}^{\alpha,s}_{p,q}\|^r \lesssim \|\mathbf{K}^r[\mathbf{F}^r]|L^{\tilde{s}}_{\tilde{p},\tilde{q}}\|.$$

Since $L^{\tilde{s}}_{\tilde{p},\tilde{q}}$ is a Banach space and

$$\mathbf{K}^r(\mathbf{x},\mathbf{y}) \leq H^r\big(A^{-1}_{\alpha,t}R_\eta(y-x), t^{-(1-\alpha)}\{\theta-\eta\}, u/t\big)$$

we can now apply Lemma 4.5.17. Indeed, we see that the kernel $\mathbf{K}^r$ is bounded by $H^r$ as required. Further, the function $H^r$ satisfies the prerequisites of Lemma 4.5.17 for the parameters $\tilde{s}$, $\tilde{p}$, and $\tilde{q}$. Note in particular the equality $\tilde{B}_i = B_i$ for $i \in \{1,\dots,4\}$ in case $r = 1$. We can thus conclude that $\mathbf{K}^r$ is a bounded linear operator from $L^{\tilde{s}}_{\tilde{p},\tilde{q}}$ to $L^{\tilde{s}}_{\tilde{p},\tilde{q}}$.

Altogether, this yields for $F \in \mathbb{L}^{\alpha,s}_{p,q}$

$$\|KF|\mathbb{L}^{\alpha,s}_{p,q}\|^r \lesssim \|\mathbf{K}^r[\mathbf{F}^r]|L^{\tilde{s}}_{\tilde{p},\tilde{q}}\| \lesssim (\tilde{B}_1 + \tilde{B}_2 + \tilde{B}_3 + \tilde{B}_4)\|\mathbf{F}^r|L^{\tilde{s}}_{\tilde{p},\tilde{q}}\|,$$

and by (4.62) we have $\|\mathbf{F}^r|L^{\tilde{s}}_{\tilde{p},\tilde{q}}\| \asymp \|\mathbf{F}|L^s_{p,q}\|^r = \|F|\mathbb{L}^{\alpha,s}_{p,q}\|^r$. This settles the proof for non-negative $K$ and $F$.

Finally, with an argument as in the proof of Lemma 4.5.17, the assertion can be shown to be valid for general $K$ and $F$. □

# Chapter 5

# Cartoon Approximation with $\alpha$-Molecules: Bounds

Many applications require efficient encoding of multivariate data in the sense of optimal sparse approximation. This is typically phrased as a problem of best $N$-term approximation with respect to a suitable representation system as explained in Subsection 2.3.1.

In the following two chapters we are concretely interested in the performance of $\alpha$-molecule systems for the sparse approximation of image data. As a model for the data, we will use the model of cartoon-like functions which very well captures the occurrence of discontinuities such as edges in an image. Some guarantees for actually achievable $N$-term approximation rates will be derived in the next chapter. In this chapter the focus is on theoretical bounds limiting the maximally achievable approximation rates by $\alpha$-molecules.

The presented research was conducted in [60] and [102]. At first, we study the approximability of cartoon-like data by arbitrary dictionaries under the assumption of a polynomial search depth. The main result, Theorem 5.3.3 ([60, Thm. 2.8]), will provide an upper bound for the achievable rates in this general setup. In Theorems 5.4.2 and 5.4.4 ([102, Thm. 3.9 & 3.11]), we then prove more specific bounds for the $\alpha$-curvelet frame from Subsection 3.2.3. Those also have implications for more general $\alpha$-molecule frames, as derived in Theorem 5.4.6 ([102, Thm. 5.3]).

## 5.1 Sparse Approximation Bounds

As in Subsection 2.3.1, we again begin with some abstract considerations in a separable Hilbert space $\mathcal{H}$. Assume that $\Phi = \{\varphi_\lambda\}_{\lambda\in\Lambda}$ is a dictionary in $\mathcal{H}$ and let $\Sigma_N := \Sigma_N[\Phi]$ be the associated space of *$N$-term expansions* introduced in (2.29). Further recall that the $N$-term approximation error (2.30) for a signal $f \in \mathcal{H}$ with respect to $\Phi$ is given by

$$\sigma_N(f) := \inf_{g\in\Sigma_N} \|f - g\|.$$

To measure the approximation performance of $\Phi$ with respect of $f$, we will subsequently use the *asymptotic approximation rate*, i.e., the decay of the approximation error $\sigma_N(f)$ as $N \to \infty$. For a subclass $\mathfrak{F} \subseteq \mathcal{H}$, the approximation performance shall be judged by the worst-case approximations, i.e., the worst decay rate of $\sigma_N(f)$ for $f \in \mathfrak{F}$. In this sense, a dictionary $\Phi$ is considered optimal for sparse approximation of $\mathfrak{F}$, if its worst-case approximation rates for signals $f \in \mathfrak{F}$ are the best among all systems.

Without reasonable restrictions, however, the investigation of the best $N$-term approximation error with respect to a given dictionary can be meaningless for practical applications. For example, if $\Phi$ is chosen as a countable dense subset of $\mathcal{H}$ one would obtain arbitrarily good 1-term approximations for every signal $f \in \mathcal{H}$. This would translate to arbitrarily good approximation rates, which clearly cannot be realized in practice.

For a proper assessment of the approximation performance of a dictionary, the actual *approximation scheme* needs to be taken into account, i.e., the utilized process of selecting suitable dictionary elements. A realistic assumption for such a selection procedure is a so-called *polynomial depth search* constraint, which requires that the terms of the $N$-term approximations have to be selected from the first $\pi(N)$ elements of the dictionary, where $\pi$ is some fixed polynomial [38].

### 5.1.1 Polynomial Depth Search in a Dictionary

For the subsequent investigation, let us concretize our considered scenario. Assume that we have a countable dictionary $\Phi = (\varphi_n)_{n\in\mathbb{N}}$ which without loss of generality is indexed by the natural numbers. A non-linear $N$-term approximation scheme can then be described by a set-valued selection function $\mathcal{S}$, which determines for given $f \in \mathcal{H}$ and $N \in \mathbb{N}$ the selected dictionary elements, i.e., $\mathcal{S}(f,N) \subset \Phi$ with $\#\mathcal{S}(f,N) = N$. Note that, due to the allowed dependence of $\mathcal{S}$ on $f$, by this general procedure even adaptive approximation schemes can be implemented. The obtained approximants are the elements $f_N \in \operatorname{span}\mathcal{S}(f,N)$ minimizing the error $\|f - f_N\|$.

A polynomial depth search constraint for the selection rule $\mathcal{S}$ is described by a fixed polynomial $\pi$ and the condition $\mathcal{S}(f,N) \subseteq \{\varphi_1,\dots,\varphi_{\pi(N)}\}$ for all $f \in \mathcal{H}$ and $N \in \mathbb{N}$. Under this condition, an optimal selection $\mathcal{S}$ thus yields *best $N$-term approximations* $f_N \in \Sigma_N$ in the following modified sense,

$$f_N = \operatorname*{arg\,min}_{g=\sum_{\lambda\in\Lambda_N} c_\lambda\varphi_\lambda} \|f-g\| \quad \text{s.t.} \quad \Lambda_N \subseteq \{1,\dots,\pi(N)\}, \quad \#\Lambda_N \le N. \tag{5.1}$$

This definition of $f_N$ should be compared with (2.31). Whereas $f_N$ in the sense of (2.31) might not exist, the existence of $f_N$ as in (5.1) is always guaranteed.

We now recall a benchmark derived in [38] concerning the optimal approximation rate of a dictionary when polynomial depth search is used. Beforehand, we have to recall what it means for a subclass $\mathfrak{F} \subseteq \mathcal{H}$ to contain a copy of $\ell_0^p$ (see also [38, Def. 1&2]).

**Definition 5.1.1** ([60, Def. 2.2]). (i) A subclass $\mathfrak{F} \subseteq \mathcal{H}$ is said to *contain an embedded orthogonal hypercube of dimension $m$ and sidelength $\delta$* if there exist $f_0 \in \mathfrak{F}$ and orthogonal elements $\psi_i \in \mathcal{H}$ for $i = 1,...,m$ with $\|\psi_i\| = \delta$ such that the collection of hypercube vertices

$$\mathfrak{H}(m; f_0, (\psi_i)_i) = \Big\{ h = f_0 + \sum_{i=1}^{m} \epsilon_i \psi_i \ : \ \epsilon_i \in \{0,1\} \Big\}$$

is contained in $\mathfrak{F}$. It should be noted that $\mathfrak{H}$ just consists of its vertices.

(ii) A subclass $\mathfrak{F} \subseteq \mathcal{H}$ is said to *contain a copy of $\ell_0^p$, $p > 0$*, if there exists a sequence of orthogonal hypercubes $(\mathfrak{H}_k)_{k\in\mathbb{N}}$, embedded in $\mathfrak{F}$, which have dimensions $m_k$ and sidelengths $\delta_k$, such that $\delta_k \to 0$ and for some constant $C > 0$

$$m_k \ge C\delta_k^{-p} \quad \text{for all } k \in \mathbb{N}. \tag{5.2}$$

Note, that if $\mathfrak{F}$ contains a copy of $\ell_0^p$, then it also contains a copy of $\ell_0^q$ for all $0 < q < p$. It was shown in [38, Thm. 2] that if a subclass $\mathfrak{F}$ contains a copy of $\ell_0^p$ there exists an upper bound on the maximal achievable approximation rate via reconstruction in a fixed dictionary.

We state a reformulation of this landmark result, which in its original form [38, Thm. 2] is stated in terms of the coefficient decay. The original proof can be adapted to lead to the following formulation from [60, Thm. 2.2], which is in terms of the best $N$-term approximation and more appropriate for our needs.

**Theorem 5.1.2** ([60, Thm. 2.2])**.** *Suppose, that a class $\mathfrak{F} \subseteq \mathcal{H}$ is uniformly bounded and contains a copy of $\ell_0^p$ for $p \in (0, 2]$. Then, allowing only polynomial depth search in a given dictionary, there is a constant $C > 0$ such that for every $N_0 \in \mathbb{N}$ there is a vector $f \in \mathfrak{F}$ and an $N \in \mathbb{N}$, $N \geq N_0$ such that*

$$\|f - f_N\|^2 \geq C\big(N \log_2(N)\big)^{-(2-p)/p},$$

*where $f_N$ denotes the best $N$-term approximation under the polynomial depth search constraint.*

*Proof.* Let $\Phi = (\varphi_n)_{n\in\mathbb{N}}$ be a given dictionary and $\pi$ the polynomial specifying the search depth. The best $N$-term approximation of $f \in \mathcal{H}$ obtained in this setting, i.e., (5.1), shall be denoted by $f_N$, the corresponding optimal selection rule, as described above, by $\mathcal{S}$.

Each system $\mathcal{S}(f, N)$ can be orthonormalized by the Gram-Schmidt procedure (starting from lower indices to higher indices), giving rise to an orthonormal basis of $\operatorname{span} \mathcal{S}(f, N)$ (with the exception of some possible zero vectors). Therefore we can represent each $f_N$ by the unique set of coefficients obtained from an expansion in this basis. (If a basis element is zero, the corresponding coefficient is chosen to be zero.)

In order to apply information theoretic arguments, we consider the following coding procedure. For $f \in \mathfrak{F}$ we select the dictionary elements $\mathcal{S}(f, N)$ and quantize the coefficients of $f_N$ obtained as above by rounding to multiples of the quantity $\mathfrak{q} = N^{-2/p}$.

We need $N \log_2(\pi(N))$ bits of information to encode the locations of the selected elements $\mathcal{S}(f, N)$ and $N \log_2(2T/\mathfrak{q})$ bits for the coefficients themselves, where $T$ is the uniform norm bound for the elements of $\mathfrak{F}$. Hence, in this procedure we are encoding with at most

$$R(N) = N\big(C_1 + C_2 \log_2(N)\big), \qquad C_1, C_2 > 0,$$

bits, and for $N \geq 2$ we have $R(N) \leq C_3 N \log_2(N)$ for some constant $C_3 > 0$. To decode, we simply reconstruct the rounded values of the coefficients and then synthesize using the selected dictionary elements.

Let $\mathfrak{H}$ be a hypercube in $\mathfrak{F}$ of dimension $m$ and sidelength $\delta$. Starting with a vertex $h \in \mathfrak{H}$ the coding-decoding procedure (for some fixed $N \in \mathbb{N}$) yields some $\tilde{h} \in \mathcal{H}$. By passing to the closest vertex $\hat{h}$, we again obtain an element of the hypercube $\mathfrak{H}$.

Every vertex $h \in \mathfrak{H}$ can be represented as a word of $m$ bits, each bit corresponding to one side of the cube. Thus the above coding procedure gives a map of the $m$ bits, which specify the vertex $h \in \mathfrak{H}$, to $R = R(N)$ bits. The decoding then reconstructs the $m$ bits specifying the vertex $\hat{h} \in \mathfrak{H}$. Since at the intermediate step we just have $R$ bits of information we unavoidably loose information if $R < m$.

Now we can apply an information theoretic argument. By rate-distortion theory [38, 5] there must be at least one vertex $h \in \mathfrak{H}$, where the number of false reconstructed bits is larger than $D_m(R)$. Here $D_m(R)$ is the so-called $m$-letter distortion-rate function. Since each bit determines a side of the cube, the error we make for this vertex $h$ obeys

$$\|h - \hat{h}\|^2 \geq \delta^2 \cdot D_m(R).$$

Since by construction $\|\tilde{h} - h\| \geq \|\tilde{h} - \hat{h}\|$ we have $\|\tilde{h} - h\| \geq \frac{1}{2}\|\hat{h} - h\|$. It follows

$$\|\tilde{h} - h\|^2 \geq \frac{1}{4}\delta^2 \cdot D_m(R).$$

By assumption, $\mathfrak{F}$ contains a copy of $\ell_0^p$. Therefore we can find a sequence of hypercubes $\mathfrak{H}_k$ with sidelengths $\delta_k \to 0$ as $k \to \infty$ and dimensions $m_k = m_k(\delta_k) \geq C\delta_k^{-p}$. For large $k$ we then pick $N_k \in \mathbb{N}$ such that $N_k \log_2(N_k) \asymp m_k$ subject to the condition $C_3 N_k \log_2(N_k) \leq \frac{1}{3}m_k$. This ensures that $N_k$ obeys the inequality $R(N_k) \leq \frac{1}{3}m_k$.

Here we can apply another result from rate-distortion theory. If $\frac{R}{m} \leq \rho$ for some $\rho < \frac{1}{2}$ it holds $D_m(R)/m \geq D_1(\rho)$, where $D_1$ is the so-called single-letter distortion-rate function. Hence, if $\frac{R}{m} \leq \frac{1}{3}$, we have

$$\|\tilde{h} - h\|^2 \geq \frac{1}{4}D_1(\frac{1}{3})\delta^2 m.$$

Let $h_k$ denote the vertices with maximal reconstruction error $\|h_k - \tilde{h}_k\|$ at each hypercube $\mathfrak{H}_k$. Taking into account $N_k \log_2(N_k) \asymp m_k \gtrsim \delta_k^{-p}$ we can then conclude for large $k$

$$\|\tilde{h}_k - h_k\|^2 \geq \frac{1}{4}D_1(\frac{1}{3})\delta_k^2 m_k \gtrsim \delta_k^2 m_k \gtrsim (N_k \log_2(N_k))^{-(2-p)/p}.$$

Finally we have to take care of the rounding errors. The best $N_k$-term approximation $h_k'$ differs from $\tilde{h}_k$ by at most $\mathfrak{q}\sqrt{N_k}$, i.e.,

$$\|\tilde{h}_k - h_k'\| \leq \mathfrak{q}\sqrt{N_k}\,,$$

since the coefficients belong to an orthonormal basis. It follows, with some constant $C > 0$,

$$\begin{aligned}\|h_k - h_k'\| &\geq \|\tilde{h}_k - h_k\| - \|\tilde{h}_k - h_k'\| \geq C(N_k \log_2(N_k))^{\frac{1}{2}-\frac{1}{p}} - N_k^{1/2-2/p} \\ &\gtrsim (N_k \log_2(N_k))^{-(2-p)/(2p)}.\end{aligned}$$

This finishes the proof. □

We will apply Theorem 5.1.2 to obtain an upper bound on the achievable approximation rates for cartoon-like functions. This model class for natural images is introduced in the next section.

## 5.2 Cartoon-like Functions

In order to theoretically analyse the approximability of images, a suitable data model for the images under consideration is required. For bivariate data in general, a standard continuum model is given by the Hilbert space $L_2(\mathbb{R}^2)$. Our concrete objects of interest are natural images, however, such as pictures or photographs of real-world motives. Due to their specific structure, the model space $L_2(\mathbb{R}^2)$ can be significantly reduced.

Suitable models for natural images are provided for example by subclasses of $L_2(\mathbb{R}^2)$, consisting of so-called cartoon-like functions. These are functions which consist of smooth regions separated from one another by piecewise-smooth discontinuity curves. Their structure imitates the fact that edges, a typical feature of natural images, are characterized by abrupt changes of color and brightness, whereas changes in the regions in between occur smoothly.

Mathematically, models of cartoon-like functions can be concretised in different ways. The classic model [15] postulates a compact image domain separated into two $C^2$ regions by a closed $C^2$ discontinuity curve. This model was generalized in various directions, e.g., to take into account piecewise-smooth edges or to allow more general $C^\beta$ regularity with $\beta \in [0,\infty)$. Cartoon classes of this kind have been studied extensively, especially in the range $\beta \in (1,2]$, e.g., in [83, 73, 60]. Another variant are the closely related horizon classes, where the discontinuity is not a closed curve in the image domain but a (possibly curved) horizontal or vertical line stretching across. Such classes have been investigated e.g. in [35, 18, 87]. Let us also mention that there exist extensions to multi-dimensions, see e.g. [83]. In particular, the corresponding 3D models have been applied in the investigation of video data. We will have a closer look at the 3D setting in Section 7.5 of Chapter 7. In this chapter, our attention is restricted to 2 dimensions.

The following definition is a template for different classes of bivariate cartoons, comprising many of those mentioned above. It provides the flexibility to taylor the model to the particular needs of specific applications.

**Definition 5.2.1** ([102, Def. 3.1])**.** Let $\beta \in [0,\infty)$ and $\nu > 0$. Given a domain $\Omega \subseteq \mathbb{R}^2$ and a set $\mathcal{A}$ of admissible subsets of $\mathbb{R}^2$, the class $\mathcal{E}^\beta(\Omega;\mathcal{A},\nu)$ consists of all functions $f \in L_2(\mathbb{R}^2)$ of the form

$$f = f_1 + f_2\mathcal{X}_{\mathcal{D}},$$

where $\mathcal{D} \in \mathcal{A}$ and $f_1,\, f_2 \in C^\beta(\mathbb{R}^2)$ with $\operatorname{supp} f_1, f_2 \subseteq \Omega$ and $\|f_1\|_{C^\beta}, \|f_2\|_{C^\beta} \le \nu$. The class $\mathcal{E}^\beta_{\mathrm{bin}}(\Omega;\mathcal{A})$ shall be the collection of all 'binary functions' $\mathcal{X}_{\mathcal{D}}$, where $\mathcal{D} \in \mathcal{A}$ and $\mathcal{D} \subseteq \Omega$.

For particular choices of $\mathcal{A}$ many of the classes appearing in the literature can be retrieved, including classes of horizon-type. In this section we focus on the class $\mathcal{E}^\beta(\Omega;\mathcal{A},\nu)$ with fixed image domain $\Omega = [-1,1]^2$ and certain $C^\beta$ domains as admissible sets $\mathcal{A}$. Similar to [38, 15, 78, 83], we restrict our investigation to star-shaped domains, since those allow a simple parametrization of the boundary curve. The results obtained however also hold true for more general domains.

Let us introduce the collection of admissible sets $\textsc{Star}^\beta(\nu)$, $\nu > 0$, as all translates of sets $B \subseteq \mathbb{R}^2$, whose boundary $\partial B$ possesses a parametrization $b : \mathbb{T} \to \mathbb{R}^2$ of the form

$$b(\varphi) = \rho(\varphi)\begin{pmatrix}\cos(\varphi)\\ \sin(\varphi)\end{pmatrix}, \quad \varphi \in \mathbb{T} = [0,2\pi]\,,$$

where the radius function $\rho : \mathbb{T} \to \mathbb{R}$ is a $C^\beta$ function with

$$|\partial^{\lfloor\beta\rfloor}\rho(\varphi) - \partial^{\lfloor\beta\rfloor}\rho(\varphi')| \le \nu\rho_0|\varphi-\varphi'|^{\beta-\lfloor\beta\rfloor} \quad \text{for all } \varphi,\varphi' \in \mathbb{T}, \tag{5.3}$$

where we set $\rho_0 := \min_{\varphi\in\mathbb{T}} \rho(\varphi) \ge \nu^{-1}$. The condition (5.3) implies that with $C = C(\beta) = (2\pi)^\beta \ge 1$ we have $\|\rho^{(k)}\|_{C^0(\mathbb{T})} \le C\rho_0\nu$ for every $k \in \{1,\dots,\lfloor\beta\rfloor\}$ if $\beta \ge 1$, and $|\rho(\varphi) - \rho(\varphi')| \le C\rho_0\nu$ for $\varphi,\varphi' \in \mathbb{T}$. In particular $\rho_0 \le \rho(\varphi) \le \rho_0(1 + C\nu)$ for all $\varphi \in \mathbb{T}$.

Note, that the set $\textsc{Star}^\beta(\nu)$ differs from the set of star-shaped domains used in [38, 15, 78, 83]. The domains in $\textsc{Star}^\beta(\nu)$ are not restricted to subsets of $[-1,1]^2$. In fact, every star-shaped $C^\beta$ domain with center 0 and $\rho_0 > 0$ is contained in $\textsc{Star}^\beta(\nu)$ for suitably large $\nu$. Moreover, the collection $\textsc{Star}^\beta(\nu)$ is scaling invariant in the sense that for $B \in \textsc{Star}^\beta(\nu)$ and $\lambda > 0$ also $\lambda B \in \textsc{Star}^\beta(\nu)$, provided $\lambda\rho_0 \geq \nu^{-1}$. In addition, with $B \in \textsc{Star}^\beta(\nu)$ also the complement $B^c = \mathbb{R}^2 \backslash B$ is contained in $\textsc{Star}^\beta(\nu)$.

Building upon Definition 5.2.1, we now define the class of functions which we want to study.

**Definition 5.2.2** (see (28) in [102])**.** Assume $\beta \in [0,\infty)$ and $\nu > 0$. We define

$$\mathcal{E}^\beta([-1,1]^2;\nu) := \mathcal{E}^\beta([-1,1]^2; \textsc{Star}^\beta(\nu), \nu)$$

as the class of cartoon-like functions obtained from Definition 5.2.1 by choosing $\Omega = [-1,1]^2$ and $\mathcal{A} = \textsc{Star}^\beta(\nu)$. The associated binary class shall be denoted by $\mathcal{E}^\beta_{\text{bin}}([-1,1]^2;\nu) := \mathcal{E}^\beta_{\text{bin}}([-1,1]^2; \textsc{Star}^\beta(\nu))$.

In the sequel, we will be interested in the approximation performance of $\alpha$-molecule systems with respect to the class $\mathcal{E}^\beta([-1,1]^2;\nu)$. Let us at first assume, however, that we can freely choose the utilized dictionary, and let us aim for a benchmark for the best possible $N$-term approximation rate achievable under a polynomial depth search constraint.

## 5.3 Entropy Bounds for Cartoon-like Functions

In this section we establish an upper bound on the maximal achievable approximation rate for $\mathcal{E}^\beta([-1,1]^2;\nu)$ when polynomial depth search in an arbitrary dictionary is used. A result like this was first derived by Donoho [38, Thm. 1] for binary $C^\beta$ cartoons in the range $\beta \in (1,2]$. Later similar results were proved for more general cartoon classes [83, 73, 60].

In principle, our statement, Theorem 5.3.3, is a known result (see e.g. [83]). However, for reasons of completeness, we outline a short proof based on the technique used in [38]. It relies on Theorem 5.1.2 and the fact that the class $\mathcal{E}^\beta([-1,1]^2;\nu)$ contains a copy of $\ell_0^p$ for $p = 2/(\beta+1)$. To show this, let us introduce the following subclass of smooth functions for $\beta \in [0,\infty)$ and $\nu > 0$,

$$C_0^\beta([-1,1]^2;\nu) := \{ f \in C_0^\beta([-1,1]^2) \ : \ \|f\|_{C^\beta} \leq \nu \}.$$

Note, that the choice $\Omega = [-1,1]^2$ and $\mathcal{A} = \{\emptyset\}$ in Definition 5.2.1 yields this class. As a consequence, we have

$$C_0^\beta([-1,1]^2;\nu) \subset \mathcal{E}^\beta([-1,1]^2;\nu). \tag{5.4}$$

Before turning to $\mathcal{E}^\beta([-1,1]^2;\nu)$, we now first analyze for which $p > 0$ the classes $C_0^\beta([-1,1]^2;\nu)$ and $\mathcal{E}^\beta_{\text{bin}}([-1,1]^2;\nu)$ contain a copy of $\ell_0^p$.

**Lemma 5.3.1.** *Let $\nu > 0$, $\beta \in [0,\infty)$, and $p = 2/(\beta+1)$. Then the following holds true.*

*(i) The function class $C_0^\beta([-1,1]^2;\nu)$ contains a copy of $\ell_0^p$.*

*(ii) The class of binary cartoons $\mathcal{E}^\beta_{\text{bin}}([-1,1]^2;\nu)$ contains a copy of $\ell_0^p$ if $\nu \geq 1$, otherwise it only contains the zero-function.*

*Proof.* The proof is a 2D-adaption of the proof of [83, Thm. 3.2].
Part (i): Let $\phi \in C_0^\infty(\mathbb{R})$ with $\operatorname{supp}\phi \subseteq [0,1]$ and $\phi \geq 0$ and put $\psi(t) = \phi(t_1)\phi(t_2)$ for $t = (t_1, t_2) \in \mathbb{R}^2$. Then $\psi \in C^\infty(\mathbb{R}^2)$ with $\operatorname{supp}\psi \subseteq [0,1]^2$. We choose $\phi \neq 0$ such that $\|\psi\|_{C^\beta(\mathbb{R}^2)} \leq \nu$. Next, we define for $k \in \mathbb{N}$ and $\ell = (\ell_1, \ell_2) \in \{0, ..., k-1\}^2$ the functions

$$\psi_{k,\ell}(t) = k^{-\beta}\phi(kt_1 - \ell_1)\phi(kt_2 - \ell_2).$$

These functions $\psi_{k,\ell} \in C^\infty(\mathbb{R}^2)$ are dilated and translated versions of $\psi$ with $\operatorname{supp}\psi_{k,\ell} \subseteq [\frac{\ell_1}{k}, \frac{\ell_1+1}{k}] \times [\frac{\ell_2}{k}, \frac{\ell_2+1}{k}]$ and $\|\psi_{k,\ell}\|_{C^\beta(\mathbb{R}^2)} \leq \|\psi\|_{C^\beta(\mathbb{R}^2)} \leq \nu$. In particular, $\psi_{k,\ell}$ and $\psi_{k,\ell'}$ are orthogonal in $L_2(\mathbb{R}^2)$ if $\ell \neq \ell'$. The functions in the set

$$\mathfrak{H}_k := \Big\{ \sum_{\ell \in \{0,...,k-1\}^2} \epsilon_\ell \psi_{k,\ell} \ : \ \epsilon_\ell \in \{0,1\} \text{ for every } \ell \in \{0, \dots, k-1\}^2 \Big\}$$

constitute the vertices of an orthogonal hypercube of dimension $m_k = k^2$ and side-length $\delta_k = \|\psi_{k,\ell}\|_2 = k^{-\beta-1}\|\psi\|_2$, which is embedded in the class $C_0^\beta([-1,1]^2; \nu)$. The sequence $(\delta_k)_{k\in\mathbb{N}}$ obeys $\delta_k \to 0$ as $k \to \infty$. Further (5.2) is fulfilled with $p = 2/(\beta+1)$ since

$$m_k = k^2 = (\delta_k/\|\psi\|_2)^{-\frac{2}{\beta+1}} = \|\psi\|_2^{\frac{2}{\beta+1}} \cdot (\delta_k)^{-\frac{2}{\beta+1}}.$$

Part (ii): We start with a function $\phi_0 \in C^\infty(\mathbb{R})$ and assume $\operatorname{supp}\phi_0 \subseteq [0, \pi/4]$, $0 \leq \phi_0 \leq \cos(\pi/8)^{-1} - 1$, and $\|\phi_0\|_{C^\beta(\mathbb{R})} = 1$. Then we define for $k \in \mathbb{N}$ and $\ell \in \{0, ..., k-1\}$ the functions

$$\phi_{k,\ell}(t) = k^{-\beta}\phi_0(kt - \ell\pi/4),$$

which clearly satisfy $\|\phi_{k,\ell}\|_{C^\beta(\mathbb{R})} \leq \|\phi_0\|_{C^\beta(\mathbb{R})} = 1$. Moreover, they have the property $\|\phi_{k,\ell}\|_{C^0(\mathbb{R})} \leq k^{-\beta}(\cos(\pi/8)^{-1} - 1)$ and $\|\phi_{k,\ell}\|_1 = k^{-\beta-1}\|\phi_0\|_1$.

Next, we define the functions $\rho_{k,\ell} \in C^\infty(\mathbb{T})$ on the torus $\mathbb{T} = [0, 2\pi)$ via $\rho_{k,\ell}(t) = 1 + \phi_{k,\ell}(t - \pi/8)$. They satisfy $1 \leq \rho_{k,\ell} \leq \cos(\pi/8)^{-1}$, such that $\mathfrak{D}_{k,\ell} \subset [-1,1]^2$ for the sets

$$\mathfrak{D}_{k,\ell} := \Big\{ x \in \mathbb{R}^2 \ : \ x = (r, \varphi) \text{ in polar coordinates with } 1 < r \leq \rho_{k,\ell}(\varphi),\ \varphi \in \mathbb{T} \Big\}.$$

For fixed $k$, the characteristic functions $\psi_{k,\ell} := \mathcal{X}_{\mathfrak{D}_{k,\ell}}$ are mutually orthogonal in $L_2(\mathbb{R}^2)$ due to their disjoint support. Let $B_2(0,1)$ denote the unit ball in $\mathbb{R}^2$ and consider the orthogonal hypercubes $\mathfrak{H}_k$ of dimension $m_k = k$ and side-length $\delta_k = \|\psi_{k,\ell}\|_2$ given by

$$\mathfrak{H}_k := \Big\{ \mathcal{X}_{B_2(0,1)} + \sum_{\ell \in \{0,...,k-1\}} \epsilon_\ell \psi_{k,\ell} \ : \ \epsilon = (\epsilon_1, \dots, \epsilon_{k-1}) \in \{0,1\}^{k-1} \Big\}.$$

If $\nu \geq 1$ those are contained in $\mathcal{E}^\beta_{\text{bin}}([-1,1]^2; \nu)$. Moreover, since $0 \leq \|\phi_{k,\ell}\|_2^2 \leq \|\phi_{k,\ell}\|_1$, it holds

$$\|\psi_{k,\ell}\|_2^2 = \int_0^{2\pi}\int_0^\infty |\psi_{k,\ell}(r,\varphi)| r\, dr d\varphi = \int_0^{2\pi}\int_1^{1+\phi_{k,\ell}(\varphi)} r\, dr d\varphi = \|\phi_{k,\ell}\|_1 + \frac{1}{2}\|\phi_{k,\ell}\|_2^2 \asymp \|\phi_{k,\ell}\|_1.$$

This implies $\delta_k \asymp k^{-(\beta+1)/2} \to 0$ for $k \to \infty$ and (5.2) with $p = 2/(\beta+1)$ since $m_k = k \asymp (\delta_k)^{-\frac{2}{\beta+1}}$. ☐

As a consequence of (5.4), we can deduce from Lemma 5.3.1(i) that also $\mathcal{E}^\beta([-1,1]^2;\nu)$ contains a copy of $\ell_0^{2/(\beta+1)}$.

**Corollary 5.3.2.** *The function class $\mathcal{E}^\beta([-1,1]^2;\nu)$ contains a copy of $\ell_0^p$ for $p = 2/(\beta+1)$.*

An application of Theorem 5.1.2 thus yields Theorem 5.3.3 below.

**Theorem 5.3.3** (compare [83])**.** *Let $\beta,\gamma \in [0,\infty)$ and $\nu > 0$. Assume that there is a constant $C > 0$ such that*

$$\sup_{f\in\mathcal{E}^\beta([-1,1]^2;\nu)} \|f - f_N\|_2^2 \leq CN^{-\gamma} \quad \textit{for all } N \in \mathbb{N},$$

*where $f_N$ denotes the best $N$-term approximation of $f$ obtained by polynomial depth search in a fixed dictionary. Then necessarily $\gamma \leq \beta$.*

The optimality benchmark $N^{-\beta}$ is also valid for $C_0^\beta([-1,1]^2;\nu)$ and $\mathcal{E}_{\mathrm{bin}}^\beta([-1,1]^2;\nu)$ with $\nu \geq 1$. We end this section with this observation.

**Remark 5.3.4.** According to Lemma 5.3.1(i), the bound of Theorem 5.3.3 even holds true for the class $C_0^\beta([-1,1]^2;\nu)$. This is a stronger statement due to the inclusion (5.4). Further, due to Lemma 5.3.1(ii), a statement analogous to Theorem 5.3.3 holds true for the binary class $\mathcal{E}_{\mathrm{bin}}^\beta([-1,1]^2;\nu)$ if $\nu \geq 1$.

## 5.4 Approximation Bounds for $\alpha$-Molecule Systems

Whereas the subject of the previous section was the approximability of the set $\mathcal{E}^\beta([-1,1]^2;\nu)$ by general dictionaries, we are now more specifically interested in the approximation $\alpha$-molecule systems can provide. In the subsequent investigation, which was first conducted in [102], we will analyze the approximation performance of the discrete Parseval frame of $\alpha$-curvelets $\mathfrak{C}_\alpha^\bullet$ from Subsection 3.2.3. As shown by Proposition 3.2.8, this frame is a system of $\alpha$-molecules of order $(\infty,\infty,\infty,\infty)$, which makes it a suitable anchor system for the application of the transfer principle, Theorem 2.3.6. Hence, the bounds we will obtain in Theorem 5.4.2 ([102, Thm. 3.9]) and Theorem 5.4.4 ([102, Thm. 3.11]) for the achievable rates of $\mathfrak{C}_\alpha^\bullet$ also have consequences for other $\alpha$-molecule systems. These will be stated in Theorem 5.4.6 ([102, Thm. 5.3]).

### 5.4.1 The Anchor System: $\alpha$-Curvelets

Before our investigation of approximation properties, let us shortly revisit the construction of the frame $\mathfrak{C}_\alpha^\bullet = \{\psi_\mu\}_{\mu\in M}$ from Subsection 3.2.3. First, recall that its index set $M$ is of the form $M = \mathbb{J}\times\mathbb{Z}^2$ with

$$\mathbb{J} := \Big\{ J = (j,\ell) \;:\; j \in \mathbb{N}_0,\, \ell \in \{0,\dots,L_j - 1\} \Big\}$$

and

$$L_j = 2^{\lfloor j(1-\alpha)\rfloor}, \quad j \in \mathbb{N}_0. \tag{5.5}$$

Next, recall that the functions $\psi_{j,\ell,k} \in \mathfrak{C}_\alpha^\bullet$ are simply rotations and translations of the $\alpha$-curvelets $\psi_{j,0,0}$. Let $R_\eta$ denote the rotation matrix (2.3) and $A_{\alpha,t}$ the $\alpha$-scaling matrix (2.4), i.e.,

$$R_\eta = \begin{pmatrix} \cos(\eta) & -\sin(\eta) \\ \sin(\eta) & \cos(\eta) \end{pmatrix} \quad \text{and} \quad A_{\alpha,t} = \begin{pmatrix} t & 0 \\ 0 & t^\alpha \end{pmatrix}, \quad \eta \in \mathbb{R},\, t \in \mathbb{R}_+.$$

Then, according to (3.23), for every $(j,\ell,k) \in M$ we have the relation

$$\psi_{j,\ell,k}(\cdot) = \psi_{j,0,0}(R_{\ell\omega_j} \cdot - A_{\alpha,2^j}^{-1} k),$$

where $\omega_j$ is the angle

$$\omega_j = \pi L_j^{-1} = \pi 2^{-\lfloor j(1-\alpha)\rfloor}, \quad j \in \mathbb{N}_0.$$

Finally, recall that the Fourier representation of $\psi_{j,\ell,k} \in \mathfrak{C}_\alpha^\bullet$ is given by

$$\hat{\psi}_{j,\ell,k}(\xi) = W_{j,\ell}(\xi) u_{j,\ell,k}(\xi), \quad \xi \in \mathbb{R}^2,$$

where $W_{j,\ell} : \mathbb{R}^2 \to [0,1]$ is a wedge function as in (3.18) and $u_{j,\ell,k} = u_{j,0,k}(R_{\ell\omega_j} \cdot)$ is obtained by rotating the exponential

$$u_{j,0,k}(\xi) = 2^{-j(1+\alpha)/2} \exp\left(2\pi i (2^{-j} k_1 \xi_1 + 2^{-j\alpha} k_2 \xi_2)\right), \quad \xi = (\xi_1, \xi_2) \in \mathbb{R}^2.$$

We will now elaborate a bit on the geometric aspects of the frequency tiling induced by $\mathfrak{C}_\alpha^\bullet$. Clearly, it is determined by the support of the functions $W_J \in C_0^\infty(\mathbb{R}^2)$ which, according to their definition (3.18), are polar tensor products of respective radial and angular components. For each $J = (j,\ell) \in \mathbb{J}$ we have

$$W_{j,\ell}(\xi) = U_j(|\xi|_2) V_{j,\ell}(\xi/|\xi|_2), \quad \xi \in \mathbb{R}^2, \tag{5.6}$$

with a radial function $U_j \in C^\infty(\mathbb{R}_0^+, [0,1])$ and an angular function $V_{j,\ell} \in C^\infty(\mathbb{S}^1, [0,1])$.

The functions $U_j \in C^\infty(\mathbb{R}_0^+)$, $j \in \mathbb{N}_0$, satisfy the support condition $\operatorname{supp} U_j \subseteq \mathcal{I}_j$, whereby

$$\mathcal{I}_0 := \frac{1}{6\pi} \cdot [0,2] \quad \text{and} \quad \mathcal{I}_j := \frac{1}{6\pi} \cdot [2^{j-1}, 2^{j+1}], \quad j \geq 1. \tag{5.7}$$

Further, due to (3.15) and (3.16), they equal 1 on the respective intervals

$$\mathcal{I}_0^- := \frac{1}{6\pi} \cdot [0, \tau_1] \quad \text{and} \quad \mathcal{I}_j^- := \frac{1}{6\pi} \cdot [2^{j-1}\tau_2, 2^j \tau_1], \quad j \geq 1. \tag{5.8}$$

As a consequence, all functions $W_J$ belonging to a fixed scale $j \in \mathbb{N}_0$ have support in a corona $\mathcal{C}_j$ defined by $\mathcal{C}_0 := \{\xi \in \mathbb{R}^2 \ : \ 6\pi|\xi|_2 \leq 2\}$ for $j = 0$ and

$$\mathcal{C}_j := \left\{\xi \in \mathbb{R}^2 \ : \ 2^{j-1} \leq 6\pi|\xi|_2 \leq 2^{j+1}\right\}, \quad \text{for } j \geq 1.$$

More concretely, taking into account the support of the functions $V_{j,\ell}$, the approximate support of $W_{j,\ell}$ corresponds to a pair of opposite wedges $\mathcal{W}_{j,\ell} := R_{\ell\omega_j}^{-1} \mathcal{W}_{j,0}$, which is obtained as a rotation of the set

$$\mathcal{W}_{j,0} := \left\{\xi = (\xi_1, \xi_2) \in \mathcal{C}_j \ : \ |\xi_1| \geq \cos(\varphi_j/2)|\xi|_2\right\}.$$

To analyze the support of $W_{j,\ell}$ in more detail, note that the angular function $\widetilde{V}_{j,0}$ covers an angle range of $\frac{3}{2}\omega_j$ on $\mathbb{S}^1$. Moreover, $\widetilde{V}_{j,0} \equiv 1$ on a range of size $\frac{1}{2}\omega_j$. Hence, $\operatorname{supp} V_{j,\ell} \subseteq \mathcal{A}_{j,\ell}$ and $V_{j,\ell} \equiv 1$ on $\mathcal{A}^-_{j,\ell}$ for the angular intervals

$$\mathcal{A}_{j,\ell} := R^{-1}_{j,\ell}\mathcal{A}_{j,0} \quad \text{with} \quad \mathcal{A}_{j,0} := \Big\{\xi = (\xi_1, \xi_2) \in \mathbb{S}^1 \;:\; |\xi_1| \geq \cos(3\omega_j/4)\Big\},$$
$$\mathcal{A}^-_{j,\ell} := R^{-1}_{j,\ell}\mathcal{A}^-_{j,0} \quad \text{with} \quad \mathcal{A}^-_{j,0} := \Big\{\xi = (\xi_1, \xi_2) \in \mathbb{S}^1 \;:\; |\xi_1| \geq \cos(\omega_j/4)\Big\}.$$

Introducing the wedge pairs

$$\mathcal{W}^+_J := \Big\{\xi \in \mathbb{R}^2 \;:\; |\xi|_2 \in \mathcal{I}_j,\, \varphi(\xi) \in \mathcal{A}_J\Big\} \quad \text{and} \quad \mathcal{W}^-_J := \Big\{\xi \in \mathbb{R}^2 \;:\; |\xi|_2 \in \mathcal{I}^-_j,\, \varphi(\xi) \in \mathcal{A}^-_J\Big\}, \tag{5.9}$$

we can thus formulate the following support properties, which will be of essential importance later,

$$\operatorname{supp} W_J \subseteq \mathcal{W}^+_J \qquad \text{and} \qquad W_J \equiv 1 \text{ on } \mathcal{W}^-_J. \tag{5.10}$$

A geometric illustration is displayed in Figure 5.1.

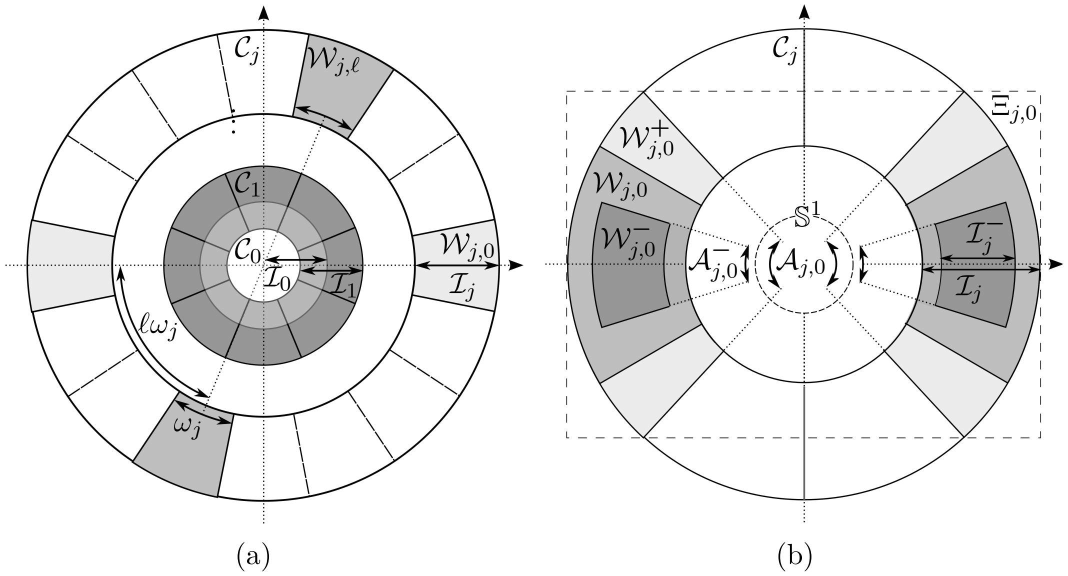



Figure 5.1: (a): Tiling of Fourier domain into coronae $\mathcal{C}_j$ and wedges $\mathcal{W}_{j,\ell}$. (b): Schematic display of the frequency support of a wedge function $W_{j,0}$.

We finally note that the sets $\mathcal{W}^+_J$ are contained in respective rectangles $\Xi_J$ of size $2^j \times 2^{j\alpha}$. Those were defined in (3.20) and are given by

$$\Xi_J = R^{-1}_J \Xi_{j,0}\,, \quad \text{where} \quad \Xi_{j,0} = [-2^{j-1}, 2^{j-1}] \times [-2^{j\alpha-1}, 2^{j\alpha-1}].$$

### 5.4.2 Approximation Bounds for $\alpha$-Curvelets

The main results of this subsection, Theorems 5.4.2 and 5.4.4, will establish bounds on the achievable $N$-term approximation rate for the class $\mathcal{E}^\beta([-1,1]^2;\nu)$, $\beta \in [0,\infty)$, when using the $\alpha$-curvelet frame $\mathfrak{C}_\alpha^\bullet$ for approximation. Unlike the more general bounds in Theorem 5.3.3, these bounds are tied to the particular approximation system $\mathfrak{C}_\alpha^\bullet$. They are established by studying the approximability of certain example cartoons.

We choose the characteristic function of the ball $B_2(0,\frac{1}{2}) \subset \mathbb{R}^2$ of radius $\frac{1}{2}$, for which we subsequently use the symbol

$$\Theta(x) := \mathcal{X}_{B_2(0,\frac{1}{2})}(x_1,x_2)\,, \quad x \in \mathbb{R}^2. \tag{5.11}$$

This function embodies an exceptionally regular binary cartoon. Its boundary is a closed $C^\infty$-curve. It is radial symmetric and contained in $\mathcal{E}^\beta_{\text{bin}}([-1,1]^2,\nu)$ for arbitrary $\beta \in [0,\infty)$ and $\nu \ge 2$. Furthermore, for every $\beta \in [0,\infty)$ and $\nu \ge 2$ there is $\gamma > 0$ such that $\gamma\Theta \in \mathcal{E}^\beta([-1,1]^2;\nu)$, wherefore the approximability of $\Theta$ has implications for the approximability of these cartoon classes.

The Fourier transform of $\Theta$ is explicitly computable. Let $\mathcal{J}_1$ denote the Bessel function of order 1, then according to (5.21)

$$\widehat{\Theta}(\xi) = \frac{\mathcal{J}_1(\pi|\xi|)}{2|\xi|}\,, \quad \xi \in \mathbb{R}^2. \tag{5.12}$$

Some properties of $\mathcal{J}_1$ and Bessel functions in general are collected in the appendix, Section 5.5.

At the center of the following investigation is the lemma below, which estimates the energy of $\widehat{\Theta}$ contained in the wedges $\mathcal{W}_J$, $J \in \mathbb{J}$. Let $\{W_J\}_{J\in\mathbb{J}}$ be a family of wedge functions of the kind (3.18) with property (3.17). Further, let

$$W_J^- := \mathcal{X}_{\mathcal{W}_J^-} \quad \text{and} \quad W_J^+ := \mathcal{X}_{\mathcal{W}_J^+}$$

be the characteristic functions of the sets $\mathcal{W}_J^-$ and $\mathcal{W}_J^+$ defined in (5.9).

**Lemma 5.4.1** ([102, Lem. 3.8])**.** *There are constants $0 < C_1 \le C_2 < \infty$, independent of scale $j \ge j_0$, where $j_0 \in \mathbb{N}_0$ is a suitable base scale, such that for all $J \in \mathbb{J}$ with $|J| \ge j_0$*

$$C_1 2^{-j(2-\alpha)} \le \|\widehat{\Theta}W_J^-\|_2^2 \le \|\widehat{\Theta}W_J\|_2^2 \le \|\widehat{\Theta}W_J^+\|_2^2 \le C_2 2^{-j(2-\alpha)},$$

*whereby $|J| = j$ for $J = (j,\ell) \in \mathbb{J}$.*

*Proof.* Let us recall the Bessel function $\mathcal{J}_1$ of order 1 and its asymptotic behavior. According to (5.23) there is a constant $C > 0$ and a function $R_1$ on $[1,\infty)$ satisfying $|R_1(r)| \le Cr^{-3/2}$ such that

$$\mathcal{J}_1(r) = \sqrt{\frac{2}{\pi r}}\cos(r - \frac{3\pi}{4}) + R_1(r) \quad \text{ for } r \ge 1.$$

This allows to separate terms of higher order from $\mathcal{J}_1^2$. We decompose

$$\mathcal{J}_1^2(r) = \Big[\frac{2}{\pi}\cos^2(r - \frac{3\pi}{4})r^{-1}\Big] + \Big[\sqrt{\frac{8}{\pi}}\cos(r - \frac{3\pi}{4})r^{-1/2}R_1(r) + R_1(r)^2\Big] =: T_1(r) + T_2(r).$$

For the following argumentation we need the square wave function $\sqcap : \mathbb{R} \to \{0,1\}$ defined by

$$\sqcap(r) := \begin{cases} 1 & , r \in \bigcup_{k\in\mathbb{Z}} k\pi + [-\frac{\pi}{2}, 0], \\ 0 & , r \in \bigcup_{k\in\mathbb{Z}} k\pi + (0, \frac{\pi}{2}). \end{cases}$$

For all $r \in \mathbb{R}$ it has the property $2\cos^2(r - 3\pi/4) \geq \sqcap(r)$. Therefore we can deduce for $1 \leq a \leq b$

$$\int_a^b T_1(r) r^{-1}\, dr = \frac{1}{\pi}\int_a^b 2\cos^2(r - \frac{3\pi}{4}) r^{-2}\, dr \geq \frac{1}{\pi}\int_a^b \sqcap(r) r^{-2}\, dr \geq \frac{1}{2}\sum_{k\in I_{a,b}} (k\pi)^{-2}$$

with $I_{a,b} := \{k \in \mathbb{Z} \,:\, k\pi \in [a+\pi, b]\}$. To proceed, we use the relation

$$\sum_{k=m}^{n} (k\pi)^{-2} \geq \frac{1}{\pi}\int_{m\pi}^{(n+1)\pi} k^{-2}\, dk,$$

which is valid for all $m, n \in \mathbb{N}$ and $m \leq n$. We obtain

$$\frac{1}{2}\sum_{k\in I_{a,b}} (k\pi)^{-2} \geq \frac{1}{2\pi}\int_{a+2\pi}^{b} k^{-2}\, dk = \frac{1}{2\pi}\Big(\int_a^b k^{-2}\, dk - \int_a^{a+2\pi} k^{-2}\, dk\Big) \geq \frac{1}{2\pi}(a^{-1} - b^{-1}) - a^{-2}.$$

Next, we see that with a constant $C > 0$ independent of $1 \leq a \leq b$

$$\int_a^b |T_2(r)| r^{-1}\, dr \leq C\int_a^b r^{-3}\, dr \leq C\int_a^\infty r^{-3}\, dr \leq Ca^{-2}.$$

Altogether, we conclude that

$$\int_a^b \frac{\mathcal{J}_1^2(r)}{r}\, dr \geq \frac{1}{2\pi}(1 - ab^{-1})a^{-1} - (1+C)a^{-2}.$$

If $c = ab^{-1} \leq 1$ is fixed, we can deduce for $a \geq 4\pi\frac{1+C}{1-c}$ the estimate

$$\int_a^{a/c} \frac{\mathcal{J}_1^2(r)}{r}\, dr \geq \frac{1}{4\pi}(1-c)a^{-1}. \tag{5.13}$$

After this preparation, we can now turn to the actual proof of the assertion. The relation

$$\|\widehat{\Theta} W_J^-\|_2^2 \leq \|\widehat{\Theta} W_J\|_2^2 \leq \|\widehat{\Theta} W_J^+\|_2^2$$

is a direct consequence of (5.10) and $\|W_J\|_\infty \leq 1$. Let $\mathcal{I}_j$ be the intervals defined in (5.7). Further, recall the intervals $\mathcal{I}_j^- \subset \mathcal{I}_j$ defined in (5.8). Using (5.12) and the definition (5.9) of $\mathcal{W}_J^-$ we calculate

$$\|\widehat{\Theta} W_J^-\|_2^2 = \int_{\mathcal{W}_J^-} \frac{\mathcal{J}_1^2(\pi|\xi|)}{4|\xi|^2}\, d\xi = \int_{\mathcal{I}_j^-}\int_{\mathcal{A}_J^-} \frac{\mathcal{J}_1^2(\pi r)}{4r}\, d\varphi dr \asymp 2^{-j(1-\alpha)}\int_{\pi\mathcal{I}_j^-} \frac{\mathcal{J}_1^2(r)}{r}\, dr.$$

The intervals $\mathcal{I}_j^-$ scale like $\sim 2^j$. Hence, if $j \in \mathbb{N}$ is chosen large enough by (5.13)

$$\|\widehat{\Theta} W_J^-\|_2^2 \asymp 2^{-j(1-\alpha)}\int_{\pi\mathcal{I}_j^-} \mathcal{J}_1^2(r) r^{-1}\, dr \gtrsim 2^{-j(1-\alpha)}2^{-j} = 2^{-j(2-\alpha)}.$$

The estimate from above is much easier to establish. If $j \in \mathbb{N}$ such that $\pi\mathcal{I}_j \subset [1, \infty)$ we have

$$\begin{aligned}\|\widehat{\Theta} W_J^+\|_2^2 &= \int_{\mathcal{W}_J^+} \frac{\mathcal{J}_1^2(\pi|\xi|)}{4|\xi|^2}\, d\xi = \int_{\mathcal{I}_j}\int_{\mathcal{A}_J} \frac{\mathcal{J}_1^2(\pi r)}{4r}\, d\varphi dr \asymp 2^{-j(1-\alpha)}\int_{\pi\mathcal{I}_j} \frac{\mathcal{J}_1^2(r)}{r}\, dr \\ &\lesssim 2^{-j(1-\alpha)}\int_{\mathcal{I}_j} r^{-2}\, dr \lesssim 2^{-j(2-\alpha)}.\end{aligned}$$

□

Based on Lemma 5.4.1 we can prove the first main result of this subsection.

**Theorem 5.4.2** ([102, Thm. 3.9]). *Let $\mathfrak{C}_\alpha^\bullet$ be the $\alpha$-curvelet frame constructed in Subsection 3.2.3 for fixed $\alpha \in [0,1)$. There exists a constant $C > 0$ such that for any given $N \in \mathbb{N}$ every $N$-term approximation $f_N$ of $\Theta$ with respect to $\mathfrak{C}_\alpha^\bullet$ (not even subject to a polynomial depth search constraint) satisfies*

$$\|\Theta - f_N\|_2^2 \geq C N^{-\frac{1}{1-\alpha}}.$$

*Proof.* Let $N \in \mathbb{N}$ be fixed and assume that

$$f_N = \sum_{r=1}^{N} \theta_{J_r,k_r} \psi_{J_r,k_r}$$

is a linear combination of $\alpha$-curvelets $\psi_{J_r,k_r}$ with coefficients $\theta_{J_r,k_r} \in \mathbb{R}$. The curvelets $\psi_{J_r,k_r} \in \mathfrak{C}_\alpha^\bullet$ satisfy $\operatorname{supp} \widehat{\psi}_{J_r,k_r} \subseteq \mathcal{W}_{J_r}^+$ as recorded in (5.10). It follows $\operatorname{supp} \widehat{f}_N \subseteq \mathcal{W}_N$ where $\mathcal{W}_N := \bigcup_{J \in \mathbb{J}_N} \mathcal{W}_J^+$ for $\mathbb{J}_N := \{J_1, \ldots, J_N\} \subset \mathbb{J}$. Using the notation $\mathbb{J}_N^c := \mathbb{J} \backslash \mathbb{J}_N$ and $\mathcal{W}_N^c := \mathbb{R}^2 \backslash \mathcal{W}_N$ we get with Lemma 5.4.1

$$\|\Theta - f_N\|_2^2 = \|\widehat{\Theta} - \widehat{f}_N\|_2^2 \geq \|\widehat{\Theta}\|_{L_2(\mathcal{W}_N^c)}^2 \geq \sum_{J \in \mathbb{J}_N^c} \|\widehat{\Theta} W_J^-\|_2^2 \gtrsim \sum_{J \in \mathbb{J}_N^c} 2^{-j(2-\alpha)}.$$

We want to bound the right-hand side from below. By (5.5), the number of tiles in each corona $\mathcal{C}_j$, $j \in \mathbb{N}_0$, is given by $L_j$, where $L_j = 2^{\lfloor j(1-\alpha) \rfloor}$ for $j \in \mathbb{N}_0$. Let $j(N) \in \mathbb{N}$ denote the unique number such that

$$\sum_{j=0}^{j(N)-1} L_j < N \leq \sum_{j=0}^{j(N)} L_j.$$

Since $2^{-j(2-\alpha)}$ decreases with rising scale we obtain

$$\sum_{J \in \mathbb{J}_N^c} 2^{-j(2-\alpha)} \geq \sum_{j=j(N)+1}^{\infty} L_j 2^{-j(2-\alpha)} \geq \frac{1}{2} \sum_{j=j(N)+1}^{\infty} 2^{-j} \gtrsim 2^{-j(N)}.$$

Here we used $L_j \geq 2^{j(1-\alpha)-1}$. Since $N \gtrsim \sum_{j=0}^{j(N)-1} 2^{j(1-\alpha)} \gtrsim 2^{j(N)(1-\alpha)}$ we can finally deduce

$$\|\Theta - f_N\|_2^2 \gtrsim 2^{-j(N)} = \left(2^{j(N)(1-\alpha)}\right)^{-\frac{1}{1-\alpha}} \gtrsim N^{-\frac{1}{1-\alpha}}. \qquad \square$$

This result can be strengthened if we restrict to greedy $N$-term approximations obtained by thresholding the coefficients. Essential is the following observation, which has also been used in [60]. Due to its importance we give a rigorous proof here.

**Lemma 5.4.3** ([102, Lem. 3.10]). *There is a constant $C > 0$ such that all curvelets $\psi_\mu \in \mathfrak{C}_\alpha^\bullet$, $\mu \in M$, satisfy*

$$\|\psi_\mu\|_1 \leq C 2^{-j(1+\alpha)/2}.$$

*Proof.* Let $a_j$ be the functions from (3.25) and recall that according to (3.27) the support of $\widehat{a}_j$ is contained in the unit square $\Xi_{0,0}$ for every $j \in \mathbb{N}_0$. Let $Id$ denote the identity operator. We have the estimate

$$\left\| \mathcal{F}^{-1}\Big((Id+\partial_1^2)(Id+\partial_2^2)\widehat{a}_j\Big)\right\|_\infty \le \|(Id+\partial_1^2)(Id+\partial_2^2)\widehat{a}_j\|_1 \le \|(Id+\partial_1^2)(Id+\partial_2^2)\widehat{a}_j\|_\infty.$$

According to Proposition 3.2.8(ii) the right-hand side is bounded uniformly over all scales. We conclude that there is a constant $C > 0$, independent of $j \in \mathbb{N}_0$, such that

$$\sup_{x\in\mathbb{R}^2} |(1+x_1^2)(1+x_2^2)a_j(x)| \le C.$$

In other words $|a_j(x)| \le C(1+x_1^2)^{-1}(1+x_2^2)^{-1}$. Using the representation (3.26) we obtain

$$|\psi_{j,0,0}(x)| = 2^{j(1+\alpha)/2}|a_j(A_j x)| \le C 2^{j(1+\alpha)/2}(1+2^{2j}x_1^2)^{-1}(1+2^{2j\alpha}x_2^2)^{-1}$$

and hence

$$\begin{aligned}\int_{\mathbb{R}^2} |\psi_{j,0,0}(x)|\,dx &\lesssim 2^{j(1+\alpha)/2}\int_{\mathbb{R}^2}(1+2^{2j}x_1^2)^{-1}(1+2^{2j\alpha}x_2^2)^{-1}\,dx \\ &= 2^{-j(1+\alpha)/2}\int_{\mathbb{R}^2}(1+x_1^2)^{-1}(1+x_2^2)^{-1}\,dx \lesssim 2^{-j(1+\alpha)/2}.\end{aligned}$$

Since $\|\psi_{j,\ell,k}\|_1 = \|\psi_{j,0,0}\|_1$ the proof is finished. □

Lemma 5.4.3 allows to deduce a simple a-priori estimate of the curvelet coefficient size, namely

$$|\theta_\mu| = |\langle f,\psi_\mu\rangle| \le \|f\|_\infty \|\psi_\mu\|_1 \le C\|f\|_\infty 2^{-j(1+\alpha)/2} \quad \text{for } \mu=(j,\ell,k)\in M. \tag{5.14}$$

Note, that the constant $C > 0$ is fully determined by $\mathfrak{C}_\alpha^\bullet$. Using (5.14) we now prove a stronger statement than Theorem 5.4.2 for greedy approximations.

**Theorem 5.4.4** ([102, Thm. 3.11])**.** *Let $\alpha \in [0,1]$ be fixed. Further, let $f_N$ denote the $N$-term approximation of $\Theta$ with respect to the $\alpha$-curvelet frame $\mathfrak{C}_\alpha^\bullet$ obtained by thresholding the coefficients. There is a constant $C > 0$ such that for every $N \in \mathbb{N}$*

$$\|\Theta - f_N\|_2^2 \ge C N^{-\frac{1}{\max\{\alpha,1-\alpha\}}}.$$

*Proof.* If $\alpha \le \frac{1}{2}$ the assertion is true by Theorem 5.4.2. It remains to handle the range $1 \ge \alpha > \frac{1}{2}$. Let $\theta_{J_r,k_r} = \langle \Theta, \psi_{J_r,k_r}\rangle$, $r \in \{1,\dots,N\}$, be the $N$ largest curvelet coefficients which determine the approximant $f_N := \sum_{r=1}^N \theta_{J_r,k_r}\psi_{J_r,k_r}$. On the Fourier side the curvelet $\psi_{J,k} \in \mathfrak{C}_\alpha^\bullet$ is the product of the functions $W_J$ and $u_{J,k}$ defined in (3.18) and (3.21), respectively. Using condition (3.17) we first estimate

$$\|\Theta - f_N\|_2^2 = \|\widehat{\Theta} - \widehat{f}_N\|_2^2 \ge \sum_{J\in\mathbb{J}} \|\widehat{\Theta}W_J - \widehat{f}_N W_J\|_2^2 \ge \sum_{J\in\mathbb{J}} \|\widehat{\Theta}W_J^- - \widehat{f}_N W_J^-\|_2^2,$$

where $W_J^-$ is the characteristic function of the set $\mathcal{W}_J^-$ defined in (5.9). The triangle inequality yields

$$\frac{1}{2}\|\widehat{\Theta}W_J^-\|_2^2 \le \|\widehat{\Theta}W_J^- - \widehat{f}_N W_J^-\|_2^2 + \|\widehat{f}_N W_J^-\|_2^2 \quad \text{for every } J \in \mathbb{J}. \tag{5.15}$$

Observe the relation $W_J^- = W_J^- W_J$ and $W_J^- W_{J'} = 0$ for $J \neq J'$. Therefore, it holds

$$\widehat{f}_N W_J^- = \sum_{r=1}^{N} \theta_{J_r,k_r} \widehat{\psi}_{J_r,k_r} W_J^- = \sum_{r=1}^{N} \theta_{J_r,k_r} u_{J_r,k_r} W_{J_r} W_J^- = \sum_{k \in K_J} \theta_{J,k} u_{J,k} W_J^-$$

with $K_J = \{k_r \in \mathbb{Z}^2 \ : \ r \in \{1,\dots,N\},\ J_r = J\}$. Next, we use that $\{u_{J,k}\}_{k\in\mathbb{Z}^2}$ is an orthonormal basis for $L_2(\Xi_J)$, where $\Xi_J \supset \mathcal{W}_J^-$ is the set defined in (3.20). We estimate

$$\Big\| \sum_{k\in K_J} \theta_{J,k} u_{J,k} W_J^- \Big\|_2^2 \le \Big\| \sum_{k\in K_J} \theta_{J,k} u_{J,k} \Big\|_{L_2(\Xi_J)}^2 = \sum_{k\in K_J} |\theta_{J,k}|^2.$$

The frame coefficients satisfy the a-priori estimate $|\theta_{J,k}|^2 \lesssim 2^{-j(1+\alpha)}$ according to (5.14). Thus we obtain

$$\|\widehat{f}_N W_J^-\|_2^2 = \Big\| \sum_{k\in K_J} \theta_{J,k} u_{J,k} W_J^- \Big\|_2^2 \lesssim (\#K_J) 2^{-j(1+\alpha)}.$$

By Lemma 5.4.1 we have $\|\widehat{\Theta} W_J^-\|_2^2 \gtrsim 2^{-j(2-\alpha)}$. We deduce from (5.15)

$$\|\widehat{\Theta} W_J^- - \widehat{f}_N W_J^-\|_2^2 \ge \frac{1}{2} \|\widehat{\Theta} W_J^-\|_2^2 - \|\widehat{f}_N W_J^-\|_2^2 \gtrsim 2^{-j(2-\alpha)} - (\#K_J) 2^{-j(1+\alpha)}.$$

Altogether, we conclude

$$\|\Theta - f_N\|_2^2 \ge \sum_{J\in\mathbb{J}} \|\widehat{\Theta} W_J^- - \widehat{f}_N W_J^-\|_2^2 \gtrsim \sum_{J\in\mathbb{J}} \max\big\{0, 2^{-j(2-\alpha)} - (\#K_J) 2^{-j(1+\alpha)}\big\}.$$

Note that $\sum_J (\#K_J) \le N$. To derive a lower bound let us consider the following minimization problem:

$$\underset{\{N_J\}_{J\in\mathbb{J}}}{\textsc{Minimize}} \quad \sum_{J\in\mathbb{J}} \max\{0, 2^{-j(2-\alpha)} - N_J 2^{-j(1+\alpha)}\} \quad \text{s.t.} \quad \sum_{J\in\mathbb{J}} N_J \le N,\ N_J \in [0,\infty)\ (J \in \mathbb{J}).$$

The condition $N_J \in [0,\infty)$, which simplifies the subsequent argumentation, is possible since we are only interested in a bound. For the optimal choice $\{N_J\}_J$, it necessarily holds $\sum_J N_J = N$ and

$$N_J \le 2^{-j(2-\alpha)} 2^{j(1+\alpha)} = 2^{j(2\alpha-1)}.$$

Hence, the minimization problem can be reformulated as minimizing the term

$$\sum_{J\in\mathbb{J}} \big( 2^{-j(2-\alpha)} - N_J 2^{-j(1+\alpha)} \big)$$

under the constraints $\sum_J N_J = N$ and $N_J \le 2^{j(2\alpha-1)}$. Assume that the family $\{N_J\}_J$ fulfills these constraints. Further, let $j(N) \in \mathbb{N}$ denote the number determined by the property

$$\sum_{j=0}^{j(N)-1} 2^{j(2\alpha-1)} L_j < N \le \sum_{j=0}^{j(N)} 2^{j(2\alpha-1)} L_j, \tag{5.16}$$

where $L_j$ from (5.5) counts the wedges in the corona $\mathcal{C}_j$. Then the following estimate holds true

$$\sum_{J\in\mathbb{J}}\Big(2^{-j(2-\alpha)}-N_J 2^{-j(1+\alpha)}\Big)\geq \sum_{j=j(N)+1}^{\infty}\Big(\sum_{|J|=j}2^{-j(2-\alpha)}\Big)\geq \sum_{j=j(N)+1}^{\infty}2^{-j}\gtrsim 2^{-j(N)}.$$

To see this, note that $2^{-j(1+\alpha)}$ is decreasing with rising scale and that $L_j\geq 2^{j(1-\alpha)-1}$. Since $N\asymp 2^{j(N)\alpha}$, which follows from (5.16), we have proven

$$\|\Theta-f_N\|_2^2\gtrsim \sum_{J\in\mathbb{J}}\max\left\{0,2^{-j(2-\alpha)}-(\#K_J)2^{-j(1+\alpha)}\right\}\gtrsim 2^{-j(N)}\asymp N^{-\frac{1}{\alpha}}$$

and the proof is finished. □

The approximation results for $\Theta$ have direct implications for the class-wise approximation of cartoon-like functions. If $\nu\geq 2$, then $\Theta\in\mathcal{E}^{\beta}_{\text{bin}}([-1,1]^2;\nu)$ for arbitrary $\beta\in[0,\infty)$. Moreover, we can always find $\gamma>0$ such that $\gamma\Theta\in\mathcal{E}^{\beta}([-1,1]^2;\nu)$. This allows to draw the following conclusion.

**Corollary 5.4.5** ([102, Cor. 3.12]). *Let $\beta\in[0,\infty)$ and $\nu\geq 2$. The uniform decay of the $N$-term approximation error for $\mathcal{E}^{\beta}_{\text{bin}}([-1,1]^2;\nu)$ and $\mathcal{E}^{\beta}([-1,1]^2;\nu)$ provided by $\mathfrak{C}^{\bullet}_{\alpha}$ cannot exceed $N^{-\frac{1}{1-\alpha}}$. Futhermore, thresholding of coefficients cannot yield rates better than $N^{-\frac{1}{\max\{\alpha,1-\alpha\}}}$.*

If $\beta>2$ it is thus impossible for $\mathfrak{C}^{\bullet}_{\alpha}$ to reach the theoretically possible approximation order of $N^{-\beta}$ for the class $\mathcal{E}^{\beta}([-1,1]^2;\nu)$. In this case, the best performance is still achieved for the classic choice $\alpha=\frac{1}{2}$, with an associated approximation rate of order $N^{-2}$. A smaller $\alpha$ leads to a deterioration of the rate.

To be more precise, this behavior applies to cartoons with curved edges exemplified by the function $\Theta=\mathcal{X}_{B_2(0,\frac{1}{2})}$ from (5.11). For such cartoons the rate inevitably deteriorates as $\alpha$ tends to 0. This can be explained by the distribution of the Fourier energy of such functions which is spread more or less uniformly across all directions of the Fourier plane.

For cartoons with straight edges, on the other hand, a smaller $\alpha$ improves the approximation rate [102]. In a certain sense, such cartoons are the opposite extreme of the isotropic function $\Theta$. They are highly anisotropic and their Fourier energy is concentrated in only one distinguished direction.

### 5.4.3 Limitations for $\alpha$-Molecule Systems

As shown by Proposition 3.2.8(ii), the Parseval frame $\mathfrak{C}^{\bullet}_{\alpha}$ is a system of $\alpha$-molecules of order $(\infty,\infty,\infty,\infty)$ with respect to the $\alpha$-curvelet parametrization $(M,\Phi_M)$ where

$$\Phi_M: M\to\mathbb{P}\,,\quad (j,\ell,k)\mapsto(x_{j,\ell,k},\ell\omega_j,2^{-j})=(R^{-1}_{\ell\omega_j}A^{-1}_j k,\ell\omega_j,2^{-j}). \tag{5.17}$$

By the transfer principle, Theorem 2.3.6, we can deduce the following result from Theorem 5.4.4.

**Theorem 5.4.6** ([102, Thm. 5.3]). *Let $\alpha \in [0,1]$ and let $\mathfrak{M}_\alpha^\bullet := \{m_\lambda\}_{\lambda\in\Lambda}$ be a discrete frame of $\alpha$-molecules whose parametrization, for some $k > 0$, is $(\alpha, k)$-consistent with the $\alpha$-curvelet parametrization* (5.17). *Further, assume that for some $\gamma > \tilde{\gamma} := \max\{\alpha, 1-\alpha\}^{-1}$ the order $(L, M, N_1, N_2)$ of $\mathfrak{M}_\alpha^\bullet$ satisfies*

$$L \geq k(1+\gamma), \quad M \geq \frac{3k}{2}(1+\gamma) + \frac{\alpha-3}{2}, \quad N_1 \geq \frac{k}{2}(1+\gamma) + \frac{1+\alpha}{2}, \quad N_2 \geq k(1+\gamma). \tag{5.18}$$

*Then the coefficients $\{c_\lambda\}_{\lambda\in\Lambda}$ of each representation $\Theta = \sum_{\lambda\in\Lambda} c_\lambda m_\lambda$ of the function $\Theta$ from* (5.11) *with respect to $\mathfrak{M}_\alpha^\bullet$ satisfy $\{c_\lambda\}_{\lambda\in\Lambda} \notin \ell^p(\Lambda)$ for $p < \frac{2}{1+\tilde{\gamma}}$.*

*Proof.* Let $p < \frac{2}{1+\tilde{\gamma}}$ and assume that $\{c_\lambda\}_\lambda \in \ell^p(\Lambda)$. According to Theorem 2.3.6 condition (5.18) ensures that the systems $\mathfrak{M}_\alpha^\bullet$ and $\mathfrak{C}_\alpha^\bullet$ are sparsity equivalent in $\ell^p$, which means $\|\{\langle m_\lambda, \psi_\mu\rangle\}_{\lambda,\mu}\|_{\ell^p\to\ell^p} < \infty$. (see [59, Def. 5.3]). Hence, by sparsity equivalence, $\{\langle\Theta, \psi_\mu\rangle\}_\mu \in \ell^p(M)$. Using $\Theta = \sum_\mu \langle\Theta, \psi_\mu\rangle\psi_\mu$ and Lemma 2.3.1, this then implies an $N$-term approximation rate of order $N^{-\gamma}$, in contradiction to Theorem 5.4.4. □

## 5.5 Appendix: *Bessel Functions*

In this appendix we collect some useful facts about Bessel functions mainly taken from [71] and [51]. We are only interested in Bessel functions $J_\nu$ of integer and half-integer order in the range $\nu \in \{-\frac{1}{2}, 0, \frac{1}{2}, 1, \ldots\}$. Bessel functions of this kind occur naturally in the Fourier analysis of radial functions. For $t \in \mathbb{R}_+$ the value $J_\nu(t)$ is conveniently defined by either of the two series (see [71] and [51, Appendix B.3])

$$J_\nu(t) = \Big(\frac{t}{2}\Big)^\nu \sum_{k=0}^\infty \frac{(-1)^k}{\Gamma(k+1)\Gamma(k+\nu+1)}\Big(\frac{t}{2}\Big)^{2k} = \frac{1}{\sqrt{\pi}}\Big(\frac{t}{2}\Big)^\nu \sum_{k=0}^\infty \frac{(-1)^k\Gamma(k+\frac{1}{2})}{\Gamma(k+\nu+1)}\frac{t^{2k}}{(2k)!}\,, \tag{5.19}$$

where the Gamma function $\Gamma$ extends the factorial $z!$ to the complex numbers with $\Gamma(z) = (z-1)!$. To verify the equivalence of both representations, it is useful to note that $\Gamma(k+\frac{1}{2}) = \frac{(2k)!}{k!4^k}\sqrt{\pi}$ for $k \in \mathbb{N}_0$. We explicitly remark, that definition (5.19) is also valid for $\nu = -\frac{1}{2}$, although this case is not included in the exposition of [51]. As is obvious from the second representation, the functions $J_\nu$ of half-integer order can be expressed in closed form in terms of trigonometric functions. For integer orders such closed form representations do not exist.

If $f(x) = f_0(|x|)$ is a radial function on $\mathbb{R}^d$, $d \in \mathbb{N}$, with a suitable function $f_0$ defined on $\mathbb{R}_0^+ = [0,\infty)$, the Fourier transform of $f$ is given by the formula

$$\widehat{f}(\xi) = \frac{2\pi}{|\xi|^{(d-2)/2}} \int_0^\infty f_0(r) J_{d/2-1}(2\pi r|\xi|) r^{d/2}\, dr\,, \quad \xi \in \mathbb{R}^d.$$

Applying this formula to the characteristic function $\mathcal{X}_{B_d(0,1)}$ of the $d$-dimensional unit ball $B_d(0,1)$ centered at the origin of $\mathbb{R}^d$ yields

$$(\mathcal{X}_{B_d(0,1)})^\wedge(\xi) = \frac{2\pi}{|\xi|^{(d-2)/2}} \int_0^1 J_{d/2-1}(2\pi|\xi|r) r^{d/2}\, dr = \frac{J_{d/2}(2\pi|\xi|)}{|\xi|^{d/2}}\,, \quad \xi \in \mathbb{R}^d. \tag{5.20}$$

Here, for the integration, we used the second of the following recurrence relations [51, Appendix B.2], which are valid for $\nu \in \frac{1}{2}\mathbb{N}$ and all $t \in \mathbb{R}_+$,

$$t^{-\nu+1} J_\nu(t) = -\frac{d}{dt}(t^{-\nu+1} J_{\nu-1}(t)) \quad \text{and} \quad t^\nu J_{\nu-1}(t) = \frac{d}{dt}(t^\nu J_\nu(t)) \,.$$

The case $\nu = \frac{1}{2}$ is not treated in [51], yet it can be easily confirmed by a direct calculation.

By scaling, we can further deduce from (5.20) the following Fourier representation of the bivariate function $\Theta(x) = \mathcal{X}_{B_2(0,1)}(2x)$, $x \in \mathbb{R}^2$, from (5.11),

$$\widehat{\Theta}(\xi) = \frac{1}{4}(\mathcal{X}_{B_2(0,1)})^\wedge(\xi/2) = \frac{J_1(\pi|\xi|)}{2|\xi|}, \quad \xi \in \mathbb{R}^2. \tag{5.21}$$

Important for our investigation in Section 5.4 is the asymptotic behavior of $J_\nu(r)$ as $r \to \infty$. We cite the following result from [51, Appendix B.8], which states for $\nu \in \frac{1}{2}\mathbb{N}_0$ the identity

$$J_\nu(r) = \sqrt{\frac{2}{\pi r}} \cos(r - \frac{\pi\nu}{2} - \frac{\pi}{4}) + R_\nu(r) \,, \quad r \in \mathbb{R}_+, \tag{5.22}$$

with a function $R_\nu$ given on $\mathbb{R}_+$ by

$$\begin{aligned} R_\nu(r) = {} & \frac{(2\pi)^{-1/2} r^\nu}{\Gamma(\nu+1/2)} e^{i(r-\pi\nu/2-\pi/4)} \int_0^\infty e^{-rt} t^{\nu+1/2} [(1+it/2)^{\nu-1/2} - 1] \, \frac{dt}{t} \\ & + \frac{(2\pi)^{-1/2} r^\nu}{\Gamma(\nu+1/2)} e^{-i(r-\pi\nu/2-\pi/4)} \int_0^\infty e^{-rt} t^{\nu+1/2} [(1-it/2)^{\nu-1/2} - 1] \, \frac{dt}{t}. \end{aligned}$$

Further, for each $\nu \in \frac{1}{2}\mathbb{N}_0$ there is a constant $C_\nu > 0$ such that $R_\nu$ satisfies the estimate

$$|R_\nu(r)| \le C_\nu r^{-3/2} \quad \text{whenever } r \ge 1. \tag{5.23}$$

The representation (5.22) and the estimate (5.23) play an important role in the proof of Lemma 5.4.1. For completeness, let us finally note that the identity (5.22) especially holds true in case $\nu = -\frac{1}{2}$, with vanishing $R_{-\frac{1}{2}} \equiv 0$. This is a direct consequence of the definition (5.19) and the Taylor series of the cosine.

# Chapter 6

# Cartoon Approximation with $\alpha$-Molecules: Guarantees

In this chapter, which is a follow-up of Chapter 5, we continue with the investigation of the approximation performance of $\alpha$-molecule systems with respect to cartoon-like data. The model setting is the same, i.e., we still consider the cartoon classes $\mathcal{E}^{\beta}_{\text{bin}}([-1,1]^2;\nu)$ and $\mathcal{E}^{\beta}([-1,1]^2;\nu)$ specified in Definition 5.2.2. The main results, Theorem 6.0.1 ([60, Thm. 4.1]) and Theorem 6.0.2 ([59, Thm. 5.12]) which will be proved below, are approximation guarantees which nicely complement the bounds established in the previous chapter.

Recall that the order of the $N$-term approximation rate achievable for the classes $\mathcal{E}^{\beta}_{\text{bin}}([-1,1]^2;\nu)$, $\nu \geq 1$, and $\mathcal{E}^{\beta}([-1,1]^2;\nu)$, $\nu > 0$, cannot exceed $N^{-\beta}$. This bound, established in Theorem 5.3.3 (see also Remark 5.3.4), is valid for arbitrary dictionaries and independent of the approximation scheme employed, as long as a polynomial depth search condition is fulfilled. Even adaptive approximation schemes cannot perform better.

Schemes, where these rates are provably achieved, at least up to order, have been developed for binary cartoons based on wedgelets [35] and surflets [19], for general cartoons utilizing bandelets [86, 87], to give a few examples. These results show that the optimality benchmark $N^{-\beta}$ can indeed be realized in practice, at least up to order. However, the utilized schemes are mostly adaptive, only for certain cartoon classes nonadaptive methods with quasi-optimal performance are known.

A breakthrough concerning the nonadaptive approximation of cartoon-like functions was the introduction of the classic curvelets by Candès and Donoho [14, 15]. By a simple thresholding scheme, they achieve an approximation rate for the class $\mathcal{E}^2([-1,1]^2;\nu)$ matching the class bound $N^{-2}$ up to a log-factor. The reason for this quasi-optimal performance is the parabolic scaling law employed. The following argument shall heuristically explain, why parabolic scaling is ideal for the representation of $C^2$ edges.

In local Cartesian coordinates, a $C^2$ curve can be represented as the graph $(E(x), x)$ of a function $E \in C^2(\mathbb{R})$ and one can choose a coordinate system such that $E'(0) = E(0) = 0$. A Taylor expansion then yields approximately $E(x) \approx \frac{1}{2}E''(0)x^2$, which matches the essential support $\textit{width} \approx \textit{length}^2$ of parabolically scaled functions. Hence, those can provide optimal resolution of the curve across all scales.

The quasi-optimal performance of curvelets for the class $\mathcal{E}^{\beta}([-1,1]^2;\nu)$ with $\beta = 2$ raised the question if similar quasi-optimal results can be obtained for other cartoon classes with a regularity $\beta \neq 2$. At least in the range $\beta \in (1,2)$, a heuristic, similar to the one given above for $C^2$ curves, applies to $C^{\beta}$ curves. Generally, if $\beta \in (1,2]$, a Taylor expansion of $E \in C^{\beta}(\mathbb{R})$ yields $|E(x)| \lesssim x^{\beta}$, and thus the boundary curve is contained in a rectangle of size $\textit{width} \approx \textit{length}^{\beta}$. This suggests $\alpha$-scaling with $\alpha = \beta^{-1}$ for optimal approximation.

And indeed, quasi-optimal approximation could be shown in [73, 83] for $\alpha$-shearlet frames if $\alpha = \beta^{-1}$ and $\beta \in (1, 2]$. For $\alpha$-curvelets, the classic approximation result by Candès and Donoho was extended in [60, Thm. 4.1]. This extension is stated below, slightly modified to fit into our model setting, since the class $\mathcal{E}^\beta([-1,1]^2;\nu)$ used here is not fully identical to the class used in [60].

**Theorem 6.0.1** ([60, Thm. 4.1])**.** *Let $\beta \in (1,2]$, $\nu > 0$. For the choice $\alpha = \beta^{-1}$, the Parseval frame of $\alpha$-curvelets $\mathfrak{C}^\bullet_\alpha$ constructed in Subsection 3.2.3 provides almost optimal sparse approximations for the class of cartoon-like functions $\mathcal{E}^\beta([-1,1]^2;\nu)$. More precisely, there exists a constant $C > 0$ such that for every $f \in \mathcal{E}^\beta([-1,1]^2;\nu)$ and $N \in \mathbb{N}$*

$$\|f - f_N\|_2^2 \le C N^{-\beta} \log_2(1+N)^{1+\beta} \,,$$

*where $f_N$ denotes the $N$-term approximation of $f$ obtained by choosing the $N$ largest coefficients.*

When we compare this theorem with the benchmark Theorem 5.3.3, we see that the frame $\mathfrak{C}^\bullet_\alpha = \{\psi_\mu\}_{\mu\in M}$ attains the maximal achievable approximation rate up to a log-factor. Moreover, as for the classical curvelets, this rate is achieved by simply thresholding the frame coefficients, leading to an intrinsically non-adaptive approximation scheme.

Unfortunately, Theorem 5.4.2 proved in the previous chapter shows that $\alpha$-curvelets are not able to provide approximation rates beyond $N^{-2}$, which is suboptimal for $\mathcal{E}^\beta([-1,1]^2;\nu)$ if $\beta > 2$. This is due to the fact that $\alpha$-scaling is not able to take advantage of smoothness beyond $C^2$. Further, Theorem 5.4.6 is an indicator that also more general $\alpha$-molecule systems are not able to overcome this $N^{-2}$ barrier. In fact, up to now no frame construction is known where a nonadaptive thresholding scheme yields approximation rates better than $N^{-2}$ for $\mathcal{E}^\beta([-1,1]^2;\nu)$. Therefore further research is required and new ideas need to be considered.

Let us now turn to the proof of Theorem 6.0.1. It is based on an analysis of the decay of the curvelet coefficients $\{\theta_\mu\}_{\mu\in M}$ given by $\theta_\mu = \langle f, \psi_\mu\rangle$ for a signal $f \in \mathcal{E}^\beta([-1,1]^2;\nu)$. This analysis will be conducted below, beginning in Section 6.1, with the main result being Theorem 6.1.1. It shows that $\{\theta_\mu\}_\mu \in \omega\ell^p(M)$ with $p = 2/(1+\beta) + \varepsilon$ and $\varepsilon > 0$ arbitrarily small. This decay rate is in fact sufficient for a proof of Theorem 6.0.1 via Lemma 2.3.1.

Before turning to Section 6.1, let us state another consequence of Theorem 6.1.1. By applying the transfer principle, Theorem 2.3.6, it is possible to deduce approximation rates for more general $\alpha$-molecule systems. Let $(M, \Phi_M)$ denote the parametrization (3.28) of the Parseval frame of $\alpha$-curvelets $\mathfrak{C}^\bullet_\alpha$. Then we can formulate the following theorem, which is the other main result of this chapter.

**Theorem 6.0.2** ([59, Thm. 5.12])**.** *Let $\beta \in (1,2]$, $\nu > 0$, and $\alpha = \beta^{-1}$. Assume that, for some $k > 0$, a discrete frame $\{m_\lambda\}_{\lambda\in\Lambda}$ of $\alpha$-molecules satisfies the following two conditions:*

*(i) its parametrization $(\Lambda, \Phi_\Lambda)$ and $(M, \Phi_M)$ are $(\alpha, k)$-consistent,*

*(ii) its order $(L, M, N_1, N_2)$ satisfies*

$$L \ge k(1+\beta), \quad M \ge \frac{3k}{2}(1+\beta) + \frac{\alpha - 3}{2}, \quad N_1 \ge \frac{k}{2}(1+\beta) + \frac{1+\alpha}{2}, \quad \text{and} \quad N_2 \ge k(1+\beta).$$

*Then each dual frame $\{\tilde{m}_\lambda\}_{\lambda\in\Lambda}$ possesses an almost optimal $N$-term approximation rate for the class of cartoon-like functions $\mathcal{E}^\beta([-1,1]^2;\nu)$, i.e., for all $f \in \mathcal{E}^\beta([-1,1]^2;\nu)$,*

$$\|f - f_N\|_2^2 \lesssim N^{-\beta+\varepsilon}, \quad \varepsilon > 0 \text{ arbitrary},$$

*where $f_N$ denotes the $N$-term approximation obtained from the $N$ largest frame coefficients.*

Note that, by Corollary 3.4.4, the required condition (i) holds in particular for the $\alpha$-curvelet and $\alpha$-shearlet parametrizations, for $k > 2$. Thus, this result allows a simple and systematic derivation of approximation results for $\alpha$-curvelets and $\alpha$-shearlets. For example, we obtain the statement of [59, Thm. 5.13] on cartoon approximation with band-limited $\alpha$-shearlet systems, where the optimal approximation rate is reached up to an arbitrarily small deviation $\varepsilon > 0$.

## 6.1 Sparsity of Curvelet Coefficients

The main statements of this chapter, Theorem 6.0.1 and Theorem 6.0.2, are both consequences of the following result on the coefficient decay of curvelet coefficients.

**Theorem 6.1.1** ([60, Thm. 4.2])**.** *Let $\theta_N^*$ denote the (in modulus) $N$th largest curvelet coefficient. Then there exists some universal constant $C$ such that*

$$\sup_{f\in\mathcal{E}^\beta([-1,1]^2;\nu)} |\theta_N^*| \le C \cdot N^{-(1+\beta)/2} \cdot (\log_2 N)^{(1+\beta)/2}.$$

This theorem is the centerpiece for proving both Theorem 6.0.1 and Theorem 6.0.2. In fact, Theorem 6.1.1 together with Lemma 2.3.1 directly leads to Theorem 6.0.1.

*Proof of Theorem 6.0.1.* Applying Lemma 2.3.1 and Theorem 6.1.1 we can estimate

$$\|f - f_N\|^2 \lesssim \sum_{m>N} |\theta_m^*|^2 \lesssim \sum_{m>N} m^{-(1+\beta)} \cdot (\log_2 m)^{(1+\beta)} \lesssim \int_N^\infty t^{-(1+\beta)} \cdot (\log_2 t)^{(1+\beta)}\, dt.$$

Using partial integration we obtain

$$\begin{aligned}
\int_N^\infty t^{-(1+\beta)} \cdot (\log_2 t)^{(1+\beta)}\, dt &\lesssim [-t^{-\beta}\,(\log_2 t)^{(1+\beta)}]_N^\infty + \int_N^\infty t^{-(1+\beta)} \cdot (\log_2 t)^{\beta}\, dt \\
&\lesssim N^{-\beta}\,(\log_2 N)^{(1+\beta)} + \int_N^\infty t^{-(1+\beta)} \cdot (\log_2 t)^{\lceil\beta\rceil}\, dt \\
&\lesssim \ldots \lesssim N^{-\beta}\,(\log_2 N)^{(1+\beta)} + \int_N^\infty t^{-(1+\beta)}\, dt \\
&\lesssim N^{-\beta}\,(\log_2 N)^{(1+\beta)}.
\end{aligned}$$

□

Using the transfer principle, Theorem 2.3.6, it is further possible to deduce Theorem 6.0.2.

*Proof of Theorem 6.0.2.* Let $\mathfrak{C}_\alpha^\bullet = \{\psi_\mu\}_{\mu\in M}$ be the $\alpha$-curvelet frame from Definition 3.2.6, and let $f \in \mathcal{E}^\beta([-1,1]^2;\nu)$. By Theorem 6.1.1, the sequence of curvelet coefficients $\{\theta_\mu\}_\mu$ given by $\theta_\mu = \langle f, \psi_\mu\rangle$ belongs to $\omega\ell^p(M)$ for every $p > \frac{2}{1+\beta}$. Since $\omega\ell^p \hookrightarrow \ell^{p+\varepsilon}$ for arbitrary $\varepsilon > 0$, this further implies $\{\theta_\mu\}_\mu \in \ell^p(M)$ for every $p > \frac{2}{1+\beta}$.

By Theorem 2.3.6, conditions (i) and (ii) guarantee that the frame $\{m_\lambda\}_{\lambda\in\Lambda}$ is sparsity equivalent to $\{\psi_\mu\}_{\mu\in M}$ in $\ell^p$ for every $p > \frac{2}{1+\beta}$. This implies that the cross-Gramian $\{\langle\psi_\mu, m_\lambda\rangle\}_{\mu,\lambda}$ is a bounded operator $\ell^p(M) \to \ell^p(\Lambda)$. It maps $\{\theta_\mu\}_\mu$ to the sequence $\{c_\lambda\}_\lambda$ given by

$$c_\lambda = \sum_\mu \langle\psi_\mu, m_\lambda\rangle\theta_\mu = \langle f, m_\lambda\rangle.$$

As a consequence, $\{c_\lambda\}_\lambda \in \ell^p(\Lambda)$ for every $p > \frac{2}{1+\beta}$. The embedding $\ell^p \hookrightarrow \omega\ell^p$ further proves $\{c_\lambda\}_\lambda \in \omega\ell^p(\Lambda)$ for every $p > \frac{2}{1+\beta}$.

Finally observe that we can expand $f$ with respect to the dual frame $\{\tilde{m}_\lambda\}_\lambda$ in the following way,

$$f = \sum_{\lambda\in\Lambda} c_\lambda \tilde{m}_\lambda.$$

Hence, for arbitrary $\varepsilon > 0$, an application of Lemma 2.3.1 yields

$$\|f - f_N\|_2^2 \lesssim N^{-\beta+\varepsilon},$$

where $f_N$ denotes the $N$-term approximation with respect to the dual frame $\{\tilde{m}_\lambda\}_\lambda$ obtained by choosing the $N$ largest coefficients. □

It remains to prove Theorem 6.1.1. For this, we first recall the a-priori estimate (5.14) for the size of the curvelet coefficients $\theta_\mu = \langle f, \psi_\mu\rangle$ at scale $j$, namely

$$|\theta_\mu| = |\langle f, \psi_\mu\rangle| \le \|f\|_\infty \|\psi_\mu\|_1 \le B\|f\|_\infty 2^{-(1+\alpha)j/2} \tag{6.1}$$

with a constant $B > 0$ independent of the index $\mu \in M$.

Using this estimate together with Theorem 6.1.2, which is stated and proved below, we can give a proof of Theorem 6.1.1.

*Proof of Theorem 6.1.1.* Let $M_j \subset M$ denote the indices corresponding to curvelets at scale $j$, and for $\varepsilon > 0$ put

$$M_{j,\varepsilon} = \Big\{\mu \in M_j, |\theta_\mu| > \varepsilon\Big\}.$$

By Theorem 6.1.2, which is stated and proved below, we have for $\varepsilon > 0$

$$\#M_{j,\varepsilon} = \#\Big\{\mu \in M_j, |\theta_\mu| > \varepsilon\Big\} \lesssim \varepsilon^{-2/(1+\beta)}. \tag{6.2}$$

On the other hand, (6.1) shows that there is a constant $B$, independent of scale, such that

$$|\theta_\mu| \le B\|f\|_\infty 2^{-(1+\alpha)j/2}.$$

It follows that for each $\varepsilon > 0$ there is $j_\varepsilon$ such that at scales $j \ge j_\varepsilon$ the coefficients satisfy $\theta_\mu < \varepsilon$. Hence, for $j \ge j_\varepsilon$

$$\#M_{j,\varepsilon} = \#\Big\{\mu \in M_j, |\theta_\mu| > \varepsilon\Big\} = 0.$$

The number of scales at which $M_{j,\varepsilon}$ is nonempty is therefore bounded by

$$\frac{2}{1+\alpha}\Big(\log_2(B)+\log_2(\|f\|_\infty)+\log_2(\varepsilon^{-1})\Big)\lesssim \log_2(\varepsilon^{-1}). \tag{6.3}$$

It follows from (6.2) and (6.3) that there is a constant $\widetilde{C}\geq 1$ such that

$$\#\Big\{\mu\in M, |\theta_\mu|>\varepsilon\Big\}=\sum_j \#\Big\{\mu\in M_j, |\theta_\mu|>\varepsilon\Big\}\leq \widetilde{C}\varepsilon^{-2/(1+\beta)}\log_2(\varepsilon^{-1}).$$

Let $\theta_N^*$ be the $N$th largest coefficient. Then for $\varepsilon_N>\delta_N$, where $\delta_N$ satisfies $N=\widetilde{C}\delta_N^{-2/(1+\beta)}\log_2(\delta_N^{-1})$, we have $|\theta_N^*|\leq\varepsilon_N$. If $N\geq 2$ it holds $\widetilde{C}N^{4/(1+\beta)}\log_2(N^2)\geq N$, because $1\leq\beta\leq 2$ and $\widetilde{C}\geq 1$. For $N\geq 2$ therefore

$$N^{4/(1+\beta)}\log_2(N^2)\geq \delta_N^{-2/(1+\beta)}\log_2(\delta_N^{-1}).$$

This implies $\delta_N\geq N^{-2}$, and we can conclude that $\varepsilon_N>\delta_N$ if we choose $\varepsilon_N$ as the solution of

$$N=\widetilde{C}\varepsilon_N^{-2/(1+\beta)}\log_2(N^2).$$

This choice leads to

$$\varepsilon_N=(2\widetilde{C})^{(1+\beta)/2}\cdot N^{-(1+\beta)/2}(\log_2 N)^{(1+\beta)/2},$$

which proves our claim with constant $C=(2\widetilde{C})^{(1+\beta)/2}$. □

The last missing piece is now Theorem 6.1.2.

**Theorem 6.1.2** ([60, Thm. 4.3]). *Let $M_j$ denote the curvelet indices at scale $j$. The sequence $\{\theta_\mu\}_{\mu\in M_j}$ obeys*

$$\|\{\theta_\mu\}_{\mu\in M_j}\|_{w\ell^{2/(1+\beta)}}\leq C,$$

*for some constant $C>0$ independent of scale $j$.*

The proof of this theorem is rather involved. In fact, the remainder of this whole chapter is solely devoted to this task. Thereby, we follow the exposition in [60]. The techniques used are very similar to those in [15, Sec. 5]. Due to the presence of fractional smoothness, however, some new tools involving divided differences have to be applied.

In a first step, we smoothly decompose $f\in\mathcal{E}^\beta([-1,1]^2;\nu)$ into so-called *fragments*, which can then be analyzed separately. For that we cover $\mathbb{R}^2$ at each scale $j\in\mathbb{N}_0$ with cubes

$$Q=[(k_1-1)2^{-j\alpha},(k_1+1)2^{-j\alpha}]\times[(k_2-1)2^{-j\alpha},(k_2+1)2^{-j\alpha}]\,,\quad (k_1,k_2)\in\mathbb{Z}^2,$$

which we collect in the sets $\mathcal{Q}_j$. Further, we put $\mathcal{Q}:=\bigcup_{j\in\mathbb{N}_0}\mathcal{Q}_j$. Note how the size of the squares depends upon the scale $2^{-j}$: The 'width' of the curvelets at scale $j$ obeys $\sim 2^{-j}$ and the 'length' of the curvelets is approximately $\sim 2^{-\alpha j}$. Thus, the size of the squares is about the length of the curvelets.

Next, we take a smooth partition of unity $\{\omega_Q\}_{Q\in\mathcal{Q}_j}$, where these squares are used as the index set and the functions $\omega_Q$ are supported in the corresponding squares $Q:=(2^{-j\alpha}k_1,2^{-j\alpha}k_2)+[-2^{-j\alpha},2^{-j\alpha}]^2$. More precisely, for some fixed nonnegative $C^\infty$-function

$\omega$ vanishing outside the square $[-1,1]^2$, we put $\omega_Q = \omega(2^{j\alpha}x_1 - k_1, 2^{j\alpha}x_2 - k_2)$ and assume that $\sum_{Q\in\mathcal{Q}_j} \omega_Q(x) \equiv 1$. The cartoon $f = f^0 + f^1\mathcal{X}_\mathcal{B} \in \mathcal{E}^\beta([-1,1]^2;\nu)$ can then at each scale $j \in \mathbb{N}_0$ be smoothly localized into the fragments

$$f_Q := f\omega_Q, \quad Q \in \mathcal{Q}_j.$$

For $Q \in \mathcal{Q}_j$ let $\theta_Q$ denote the curvelet coefficient sequence of $f_Q$ at scale $j$, i.e.,

$$\theta_Q = \Big\{\langle f_Q, \psi_\mu\rangle\Big\}_{\mu\in M_j}. \tag{6.4}$$

The strategy laid out in [15] is to analyze the sparsity of the sequences $\theta_Q$ and combine these results to obtain Theorem 6.1.2. In this investigation we have to distinguish between two cases: Either the square $Q \in \mathcal{Q}_j$ meets the edge curve $\Gamma = \partial\mathcal{B}$ of the cartoon or not. Accordingly, we let $\mathcal{Q}_j^1$ be the subset of $\mathcal{Q}_j$ containing those cubes, which intersect the edge curve $\Gamma$. Among the remaining cubes of $\mathcal{Q}_j$ we collect those, which intersect $\operatorname{supp} f$, in $\mathcal{Q}_j^0$. The others can be neglected, because they lead to trivial sequences $\theta_Q$.

The following two propositions ([60, Thm. 4.4] and [60, Thm. 4.5]) directly lead to Theorem 6.1.2.

**Theorem 6.1.3** (Analysis of a Smooth Fragment)**.** *Let $Q$ be a square such that $Q \in \mathcal{Q}_j^0$. The curvelet coefficient sequence $\theta_Q$ defined in* (6.4) *obeys*

$$\|\theta_Q\|_{w\ell^{2/(1+\beta)}} \le C \cdot 2^{-(1+\alpha)j},$$

*for some constant $C > 0$ independent of $Q$ and $j$.*

**Theorem 6.1.4** (Analysis of an Edge Fragment)**.** *Let $Q$ be a square such that $Q \in \mathcal{Q}_j^1$. The curvelet coefficient sequence $\theta_Q$ defined in* (6.4) *obeys*

$$\|\theta_Q\|_{w\ell^{2/(1+\beta)}} \le C \cdot 2^{-(1+\alpha)j/2},$$

*for some constant $C > 0$ independent of $Q$ and $j$.*

Theorem 6.1.2 is an easy consequence of these two results and the observation, that there are constants $A_0$ and $A_1$, independent of scale, such that

$$\#\mathcal{Q}_j^0 \le A_0 2^{2\alpha j} \quad \text{and} \quad \#\mathcal{Q}_j^1 \le A_1 2^{\alpha j}. \tag{6.5}$$

The estimates (6.5) hold true since $f$ is supported in $[-1,1]^2$.

*Proof of Theorem 6.1.2.* For $0 < p \le 1$ we have the $p$-triangle inequality

$$\|a+b\|_{w\ell^p}^p \le \|a\|_{w\ell^p}^p + \|b\|_{w\ell^p}^p, \qquad a, b \in w\ell^p.$$

Since $\{\theta_\mu\}_{\mu\in M_j} = \sum_{Q\in\mathcal{Q}_j} \theta_Q$, we can conclude

$$\begin{aligned}\|\{\theta_\mu\}_{\mu\in M_j}\|_{w\ell^{2/(1+\beta)}}^{2/(1+\beta)} &\le \sum_{Q\in\mathcal{Q}_j} \|\theta_Q\|_{w\ell^{2/(1+\beta)}}^{2/(1+\beta)} \\ &\le (\#\mathcal{Q}_j^1) \cdot \sup_{\mathcal{Q}_j^1} \|\theta_Q\|_{w\ell^{2/(1+\beta)}}^{2/(1+\beta)} + (\#\mathcal{Q}_j^0) \cdot \sup_{\mathcal{Q}_j^0} \|\theta_Q\|_{w\ell^{2/(1+\beta)}}^{2/(1+\beta)}.\end{aligned}$$

The claim follows now from the above two theorems together with observation (6.5). $\square$

It remains to prove Theorems 6.1.3 and 6.1.4. For that let $Q \in \mathcal{Q}_j$ be a fixed cube at a fixed scale $j \in \mathbb{N}_0$, which nontrivially intersects $\operatorname{supp} f$. We need to analyze the decay of the sequence $\theta_Q = \{\langle f_Q, \psi_\mu \rangle\}_{\mu \in M_j}$. Since the frame elements $\psi_\mu$ are bandlimited, it is advantageous to turn to the Fourier side. The Plancherel identity yields

$$\langle f_Q, \psi_\mu \rangle = \langle \widehat{f}_Q, \widehat{\psi}_\mu \rangle.$$

These scalar products can be estimated, if we have knowledge about the localization of the functions $\widehat{f}_Q$. This investigation is carried out separately in Sections 6.2 and 6.3 for the cases $Q \in \mathcal{Q}_j^0$ and $Q \in \mathcal{Q}_j^1$, respectively.

## 6.2 Analysis of a Smooth Fragment

The goal of this section ([60, Sec. 5]) is a proof of Theorem 6.1.3. Broadly, we follow the arguments in [15, Sec. 8] for $\beta = 2$, but with the important difference that our signal class $\mathcal{E}^\beta([-1,1]^2;\nu)$ generally involves functions with smoothness of fractional order. This forces us to translate several estimates in [15] for derivatives of various functions into estimates for corresponding moduli of smoothness. The same remark applies to the next Section 6.3.

An important tool is the forward difference operator $\Delta_{(h_1,h_2)}$, where $h_1, h_2 \in \mathbb{R}$, which acts on a bivariate function $f : \mathbb{R}^2 \to \mathbb{R}$ as follows,

$$\Delta_{(h_1,h_2)} f(x_1,x_2) := f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2). \tag{6.6}$$

Its one-dimensional analogon takes the simple form $\Delta_h f(t) := f(t+h) - f(t)$ for $h \in \mathbb{R}$. When applied to a bivariate function $f : \mathbb{R}^2 \to \mathbb{R}$, to simplify notation, the operator $\Delta_h$ shall exclusively act on variables denoted $t$ or $\tau \in \mathbb{R}$, e.g., the symbol $\Delta_h f(t,u)$ denotes the function $(t,u) \mapsto f(t+h,u) - f(t,u)$. Note that the symbol $\Delta$ without a subscript denotes the standard Laplacian.

Let us now come back to the proof of Theorem 6.1.3 and recall the notation introduced at the end of the previous section. We treat the case $Q \in \mathcal{Q}_j^0$, where the cube $Q$ does not intersect the edge curve $\Gamma = \partial \mathcal{B}$. In this case we call $f_Q = f\omega_Q$ a *smooth fragment*.

Before we begin, we briefly recall our setting. The parameters $\alpha \in [\frac{1}{2}, 1)$ and $\beta = \alpha^{-1} \in (1, 2]$ are fixed, as is $f \in \mathcal{E}^\beta([-1,1]^2;\nu)$. Since $Q$ does not intersect the edge curve, there is a function $g \in C^\beta(\mathbb{R}^2)$ such that $f_Q = g\omega_Q$. By smoothly cutting $g$ off outside the square $[-1,1]^2$, we can even assume $g \in C_0^\beta(\mathbb{R}^2)$.

We want to analyze $\widehat{f}_Q$ and for simplicity we look at the following model situation. Without loss of generality we assume that the cube $Q$ is centered at the origin, by possibly translating the coordinate system. In this case the smooth fragment takes the simple form

$$f_j(x) := f_Q(x) = g(x)\omega(2^{\alpha j} x), \quad x \in \mathbb{R}^2,$$

where $g \in C_0^\beta(\mathbb{R}^2)$ and $\omega \in C_0^\infty(\mathbb{R}^2)$ with $\operatorname{supp} \omega \subset [-1,1]^2$ is the fixed window generating the partition of unity $(\omega_Q)_Q$ (note that by our simplifications the fragment $f_Q$ only depends on the scale $j$ and therefore the notation $f_j$ is justified).

By rescaling $f_j$ we obtain for each scale $j$ the functions

$$F_j(x) = g(2^{-\alpha j} x)\omega(x), \quad x \in \mathbb{R}^2, \tag{6.7}$$

with $\operatorname{supp} F_j \subset [-1,1]^2$. We put $g_j(x) := g(2^{-\alpha j} x)$, so that we can write $F_j(x) = g_j(x)\omega(x)$. It is important to note, that $g_j$ and $F_j$ depend on the scale, whereas $\omega$ remains fixed.

### 6.2.1 Fourier Analysis of a Smooth Fragment

We first analyze the localization of $\widehat{F}_j$, where $F_j$ is given as in (6.7). The key result in this direction will be Proposition 6.2.2. Its proof relies on Lemma 6.2.1 below. Here we use the forward difference operator (6.6) defined above.

**Lemma 6.2.1.** *Let $\alpha \in [\frac{1}{2}, 1)$ and $\beta = \alpha^{-1} \in (1, 2]$. Assume that $h = C2^{-(1-\alpha)j}$ for some fixed constant $C > 0$, and put $N := 2 + \lceil \frac{\alpha}{1-\alpha} \rceil$. We then have*

$$\|\Delta^N_{(h,0)} \partial_1 F_j\|_2^2 \lesssim h^{2\beta} 2^{-2j\alpha},$$

*where the implicit constant is independent of the scale $j$. Notice that $h$ is not independent and depends on the scale $j$.*

*Proof.* Since $\operatorname{supp} F_j \subset [-1, 1]^2$ it suffices to prove

$$\|\Delta^N_{(h,0)} \partial_1 F_j\|_\infty \lesssim h^\beta 2^{-j\alpha}. \tag{6.8}$$

By the product rule we have $\partial_1 F_j = \partial_1 g_j \cdot \omega + g_j \cdot \partial_1 \omega$ and it holds

$$\begin{aligned}
\Delta^N_{(h,0)}(\partial_1 g_j \cdot \omega) &= \sum_{k=0}^{N} \Delta^k_{(h,0)} \partial_1 g_j \cdot \Delta^{N-k}_{(h,0)} \omega(\cdot + kh, \cdot), \\
\Delta^N_{(h,0)}(g_j \cdot \partial_1 \omega) &= \sum_{k=0}^{N} \Delta^k_{(h,0)} g_j \cdot \Delta^{N-k}_{(h,0)} \partial_1 \omega(\cdot + kh, \cdot).
\end{aligned} \tag{6.9}$$

Clearly, we have $\|\partial_1 g_j\|_\infty \lesssim 2^{-\alpha j}$ and for every $k \in \mathbb{N}_0$ the estimates $\|\Delta^k_{(h,0)} \omega\|_\infty \lesssim h^k$ and $\|\Delta^k_{(h,0)} \partial_1 \omega\|_\infty \lesssim h^k$. According to Lemma 6.4.3, it further holds

$$\|\Delta_{(h,0)} g_j\|_\infty \lesssim h 2^{-\alpha j} \quad \text{and} \quad \|\Delta^k_{(h,0)} g_j\|_\infty \lesssim h^\beta 2^{-j} \quad \text{for } k \geq 2,$$
$$\text{as well as} \quad \|\Delta^k_{(h,0)} \partial_1 g_j\|_\infty \lesssim h^\beta 2^{-\alpha j} \quad \text{for } k \geq 1.$$

Since $N \geq 3$, these estimates suffice to bound the summands in (6.9) for $k \neq 0$. In case $k = 0$, we observe that $h^N \lesssim h^\beta 2^{-j\alpha}$, due to $N \geq \beta + \frac{\alpha}{1-\alpha}$. The assertion (6.8) follows. □

The previous lemma is key to the proof of the following proposition. Here we use the notation $|\xi| \sim 2^{(1-\alpha)j}$ to indicate $|\xi| \in [C_1 2^{(1-\alpha)j}, C_2 2^{(1-\alpha)j}]$ for some arbitrary but fixed constants $0 < C_1 \leq C_2 < \infty$. A typical choice would be $C_1 = 1$ and $C_2 = 2^{1-\alpha}$.

**Proposition 6.2.2** ([60, Prop. 5.2]). *It holds independently of the scale $j$*

$$\int_{|\xi| \sim 2^{(1-\alpha)j}} |\widehat{F}_j(\xi)|^2 \, d\xi \lesssim 2^{-2\beta j}.$$

*Proof.* Let $0 < C_1 \leq C_2 < \infty$ be fixed and choose $C > 0$ such that $C_2 C < 2\pi$. Putting $h := C2^{-(1-\alpha)j}$, there then exists $c > 0$ such that $|e^{i\xi_1 h} - 1|^2 \geq c$ for every $\xi_1$ with $|\xi_1| \in [C_1 2^{(1-\alpha)j}, C_2 2^{(1-\alpha)j}]$. Using Lemma 6.2.1 with $N := 2 + \lceil \frac{\alpha}{1-\alpha} \rceil$ we then estimate the integrals on the vertical strips:

$$\begin{aligned}
2^{2(1-\alpha)j} \int_{|\xi_1|\sim 2^{(1-\alpha)j}} \int_{\xi_2} |\widehat{F}_j(\xi_1,\xi_2)|^2 \, d\xi_2 \, d\xi_1 &\asymp \int_{|\xi_1|\sim 2^{(1-\alpha)j}} \int_{\xi_2} |\xi_1|^2 |\widehat{F}_j(\xi_1,\xi_2)|^2 \, d\xi_2 \, d\xi_1 \\
&\lesssim \int_{|\xi_1|\sim 2^{(1-\alpha)j}} \int_{\xi_2} |e^{i\xi_1 h} - 1|^{2N} |\xi_1|^2 |\widehat{F}_j(\xi_1,\xi_2)|^2 \, d\xi_2 \, d\xi_1 \\
&= \int_{|\xi_1|\sim 2^{(1-\alpha)j}} \int_{\xi_2} |\widehat{\Delta^N_{(h,0)} \partial_1 F_j}(\xi_1,\xi_2)|^2 \, d\xi_2 \, d\xi_1 \\
&\le \int_{\mathbb{R}^2} |\widehat{\Delta^N_{(h,0)} \partial_1 F_j}(\xi)|^2 \, d\xi = \|\Delta^N_{(h,0)} \partial_1 F_j\|_2^2 \lesssim h^{2\beta} 2^{-2j\alpha}.
\end{aligned}$$

Interchanging $\xi_1$ and $\xi_2$ yields analogous estimates for the horizontal strips. Altogether, we obtain

$$\int_{|\xi|\sim 2^{(1-\alpha)j}} |\widehat{F}_j(\xi)|^2 \, d\xi \lesssim 2^{-2j(1-\alpha)} h^{2\beta} 2^{-2j\alpha} \asymp 2^{-2\beta j}.$$

□

As an immediate conclusion, we deduce a corresponding estimate for the original smooth fragment $f_j$.

**Theorem 6.2.3** ([60, Thm. 5.3])**.** *We have independently of scale $j$*

$$\int_{|\xi|\sim 2^j} |\widehat{f}_j(\xi)|^2 \, d\xi \lesssim 2^{-2(\beta+\alpha)j}.$$

*Proof.* The statement follows from the relation $\widehat{f}_j(\xi) = 2^{-2\alpha j} \widehat{F}_j(2^{-\alpha j}\xi)$. □

Finally, we state a refinement of Theorem 6.2.3.

**Corollary 6.2.4** ([60, Cor. 5.4])**.** *Let $m = (m_1, m_2) \in N_0^2$ and $\partial^m = \partial_1^{m_1} \partial_2^{m_2}$. We have*

$$\int_{|\xi|\sim 2^j} |\partial^m \widehat{f}_j(\xi)|^2 \, d\xi \lesssim 2^{-2j\alpha|m|_1} 2^{-2(\beta+\alpha)j}.$$

*Proof.* Recall that $f_j = g\omega(2^{\alpha j}\cdot)$. Let us define the window $\tilde{\omega}(x) := x^m \omega(x)$ and the function $\tilde{f}_j(x) := g(x)\tilde{\omega}(2^{\alpha j}x)$ for $x \in \mathbb{R}^2$. Then because of $\tilde{\omega}(2^{\alpha j}x) = 2^{j\alpha|m|_1} x^m \omega(2^{\alpha j}x)$ for every $x \in \mathbb{R}^2$

$$x^m f_j(x) = g(x) x^m \omega(2^{\alpha j}x) = 2^{-j\alpha|m|_1} g(x)\tilde{\omega}(2^{\alpha j}x) = 2^{-j\alpha|m|_1} \tilde{f}_j(x).$$

We conclude with Theorem 6.2.3

$$\begin{aligned}
\int_{|\xi|\sim 2^j} |\partial^m \widehat{f}_j(\xi)|^2 \, d\xi &= \int_{|\xi|\sim 2^j} |\widehat{x^m f_j(x)}(\xi)|^2 \, d\xi \\
&= 2^{-2j\alpha|m|_1} \int_{|\xi|\sim 2^j} |\mathcal{F}\tilde{f}_j(\xi)|^2 \, d\xi \lesssim 2^{-2j\alpha|m|_1} 2^{-2(\beta+\alpha)j}.
\end{aligned}$$

□

### 6.2.2 Curvelet Analysis of a Smooth Fragment

Let $J = (j, \ell)$ be a scale-angle pair and $W_J$ the wedge function from (5.6), used in the construction of the $\alpha$-curvelet frame $\mathfrak{C}^\bullet_\alpha$. Then $W_J$ is a non-negative real-valued function, supported in the wedge pair $\mathcal{W}_J^+$ given in (5.9) and satisfying $\|W_J\|_\infty \le 1$.

Theorem 6.2.3 directly leads to a central result, namely that it holds

$$\int_{\mathbb{R}^2} \sum_{|J|=j} |(\widehat{f}_j W_J)(\xi)|^2 \, d\xi \lesssim 2^{-2j(\beta+\alpha)}. \tag{6.10}$$

Recall the notation $|J| = j$. Our next goal, is to refine this result. Let us first record a basic fact.

**Lemma 6.2.5** ([60, Lem. 5.5])**.** *Let $m \in \mathbb{N}_0^2$. It holds for all $\xi \in \mathbb{R}^2$*

$$\sum_{|J|=j} |\partial^m W_J(\xi)|^2 \lesssim 2^{-2j\alpha|m|_1}.$$

*Proof.* From the definition it follows that $W_J$ scales with $2^{-\alpha j}$ in one direction and with $2^{-j}$ in the orthogonal direction. No matter what direction, we always do better than $|\partial^m W_J(\xi)|^2 \lesssim 2^{-2j\alpha|m|_1}$. For fixed $\xi$ only a fixed number of summands are not zero, uniformly for all $\xi$. The claim follows. □

Next we prove an auxiliary lemma. Here $\Delta = \partial_1^2 + \partial_2^2$ denotes the standard Laplacian.

**Lemma 6.2.6** ([60, Lem. 5.6])**.** *Let $\Delta = \partial_1^2 + \partial_2^2$ denote the standard Laplacian. It holds for $m \in \mathbb{N}_0$*

$$\int_{\mathbb{R}^2} \sum_{|J|=j} |\Delta^m(\widehat{f}_j W_J)(\xi)|^2 \, d\xi \lesssim 2^{-2j(\beta+\alpha)} \cdot 2^{-4m\alpha j}.$$

*Proof.* For $m = 0$ this is just (6.10), a direct consequence of Theorem 6.2.3. Now let $m > 0$. It holds with $a, b \in \mathbb{N}_0^2$ and certain coefficients $c_{a,b} \in \mathbb{N}_0$

$$\Delta^m(\widehat{f}_j W_J)(\xi) = \sum_{|a|+|b|=2m} c_{a,b} \partial^a \widehat{f}_j(\xi) \partial^b W_J(\xi).$$

Let $a, b \in \mathbb{N}_0^2$ such that $|a|_1 + |b|_1 = 2m$. Then with Lemma 6.2.5 and Corollary 6.2.4

$$\begin{aligned}
\int_{\mathbb{R}^2} \sum_{|J|=j} |\partial^a \widehat{f}_j(\xi)|^2 |\partial^b W_J(\xi)|^2 \, d\xi &\lesssim 2^{-2j\alpha|b|_1} \int_{|\xi|\sim 2^j} |\partial^a \widehat{f}_j(\xi)|^2 \, d\xi \\
&\lesssim 2^{-2j\alpha|b|_1} 2^{-2j\alpha|a|_1} 2^{-2(\beta+\alpha)j} \\
&= 2^{-2j\alpha(|a|_1+|b|_1)} 2^{-2(\beta+\alpha)j} = 2^{-4j\alpha m} 2^{-2(\beta+\alpha)j}.
\end{aligned}$$
□

Now we come to the refinement of (6.10). For that, we need the differential operator

$$\mathcal{L} = \mathcal{I} - 2^{2\alpha j} \Delta, \tag{6.11}$$

where $\mathcal{I}$ is the identity and $\Delta$ the standard Laplacian. The theorem below shows that $\mathcal{L}^2(\widehat{f}_j W_J)$ obeys the same estimate (6.10) as $\widehat{f}_j W_J$.

**Theorem 6.2.7** ([60, Thm. 5.7]). *Let $\mathcal{L}$ be the differential operator defined in* (6.11)*. It holds*

$$\int_{\mathbb{R}^2} \sum_{|J|=j} |\mathcal{L}^2(\widehat{f}_j W_J)(\xi)|^2 \, d\xi \lesssim 2^{-2j(\beta+\alpha)}.$$

*Proof.* It holds

$$\mathcal{L}^2 = \mathcal{I} - 2 \cdot 2^{2\alpha j} \Delta + 2^{4\alpha j} \Delta^2.$$

Applying (6.10) and Lemma 6.2.6 yields the desired result. □

Finally we can give the proof of Theorem 6.1.3.

### 6.2.3 Proof of Theorem 6.1.3

*Proof.* Recall the curvelet frame $\mathfrak{C}^{\bullet}_{\alpha} = \{\psi_\mu\}_{\mu \in M}$. On the Fourier side

$$\widehat{\psi}_{j,\ell,k} = W_J u_{j,k}(R_J \cdot),$$

with rotation matrix $R_J$ given as in (3.24) and functions

$$u_{j,k}(\xi) := 2^{-j(1+\alpha)/2} e^{2\pi i (2^{-j} k_1, 2^{-\alpha j} k_2) \cdot \xi}, \quad \xi \in \mathbb{R}^2.$$

We have to study the decay of the sequence $\theta_Q$ defined in (6.4). Its elements $\tilde{\theta}_{j,\ell,k} := \langle f_j, \psi_{j,\ell,k} \rangle$ are given by the formula

$$\tilde{\theta}_{j,\ell,k} = \int_{\mathbb{R}^2} \widehat{f}_j W_J(\xi) \overline{u_{j,k}(R_J \xi)} \, d\xi.$$

We observe

$$\mathcal{L} u_{j,k} = (1 + 2^{-2j(1-\alpha)} k_1^2 + k_2^2) u_{j,k},$$

which also holds for the rotated versions $u_{j,k}(R_J \cdot)$. Partial integration thus yields

$$\tilde{\theta}_{j,\ell,k} = \int_{\mathbb{R}^2} \widehat{f}_j(\xi) W_J(\xi) \overline{u_{j,k}(R_J \xi)} \, d\xi = (1 + 2^{-2j(1-\alpha)} k_1^2 + k_2^2)^{-2} \int_{\mathbb{R}^2} \mathcal{L}^2(\widehat{f}_j W_J) \overline{u_{j,k}(R_J \xi)} \, d\xi.$$

For $j \in \mathbb{N}$ and $K = (K_1, K_2) \in \mathbb{Z}^2$ we define the set

$$\mathfrak{Z}_{j,K} := \Big\{ (k_1, k_2) \in \mathbb{Z}^2 \ : \ k_1 2^{-j(1-\alpha)} \in [K_1, K_1 + 1), \, k_2 = K_2 \Big\}.$$

Further, we put

$$M_{j,K} := \Big\{ \mu = (j, \ell, k) \in M_j \ : \ k \in \mathfrak{Z}_{j,K} \Big\},$$

where $M_j$ denotes the curvelet indices at scale $j$. It follows from the orthogonality properties of the Fourier system $\{u_{j,k}\}_{k \in \mathbb{Z}^2}$ that

$$\sum_{k \in \mathfrak{Z}_{j,K}} |\tilde{\theta}_{j,\ell,k}|^2 \le (1 + |K|_2^2)^{-4} \int_{\mathbb{R}^2} |\mathcal{L}^2(\widehat{f}_j W_J)(\xi)|^2 \, d\xi,$$

where $|K|_2^2 = K_1^2 + K_2^2$. We further conclude

$$\sum_{\mu \in M_{j,K}} |\tilde{\theta}_\mu|^2 = \sum_{|J|=j} \sum_{k \in \mathfrak{Z}_{j,K}} |\tilde{\theta}_{J,k}|^2 \leq (1+|K|_2^2)^{-4} \int_{\mathbb{R}^2} \sum_{|J|=j} |\mathcal{L}^2(\widehat{f}_j W_J)(\xi)|^2 \, d\xi.$$

Now we apply Theorem 6.2.7 and obtain

$$\sum_{\mu \in M_{j,K}} |\tilde{\theta}_\mu|^2 \lesssim 2^{-2j(\beta+\alpha)}(1+|K|_2^2)^{-4},$$

which directly implies

$$\|\{\tilde{\theta}_\mu\}_{\mu \in M_{j,K}}\|_{\ell^2} \lesssim 2^{-j(\beta+\alpha)}(1+|K|_2^2)^{-2}. \tag{6.12}$$

It holds $\#\mathfrak{Z}_{j,K} \leq 1 + 2^{j(1-\alpha)}$ and therefore, since $L_j = 2^{\lfloor j(1-\alpha) \rfloor}$, the estimate $\#M_{j,K} \leq 2 \cdot 2^{2j(1-\alpha)}$. Now we recall the interpolation inequality $\|\{c_\lambda\}_\lambda\|_{\ell_p} \leq n^{1/p-1/2} \|\{c_\lambda\}_\lambda\|_{\ell_2}$ for a finite sequence $\{c_\lambda\}_\lambda$ with $n$ nonzero entries. Applying this inequality with $p = 2/(1+\beta)$ and $n = 2 \cdot 2^{2j(1-\alpha)}$ the maximal size of $M_{j,K}$, we get from (6.12)

$$\|\{\tilde{\theta}_\mu\}_{\mu \in M_{j,K}}\|_{\ell^{2/(1+\beta)}} \lesssim 2^{j(\beta-1)} \|\{\tilde{\theta}_\mu\}_{\mu \in M_{j,K}}\|_{\ell^2} \leq 2^{-j(1+\alpha)}(1+|K|_2^2)^{-2}.$$

It follows

$$\sum_{\mu \in M_{j,K}} |\tilde{\theta}_\mu|^{2/(1+\beta)} \lesssim 2^{-j(1+\alpha)2/(1+\beta)} \cdot (1+|K|_2^2)^{-4/(1+\beta)} = 2^{-2\alpha j}(1+|K|_2^2)^{-4/(1+\beta)}.$$

Finally, we have

$$\sum_{\mu \in M_j} |\tilde{\theta}_\mu|^{2/(1+\beta)} = \sum_{K \in \mathbb{Z}^2} \sum_{\mu \in M_{j,K}} |\tilde{\theta}_\mu|^{2/(1+\beta)} \leq 2^{-2\alpha j} \sum_{K \in \mathbb{Z}^2} (1+|K|_2^2)^{-4/(1+\beta)} \lesssim 2^{-2\alpha j}.$$

The desired estimate for the sequence $\theta_Q = \{\tilde{\theta}_\mu\}_{\mu \in M_j}$ follows, i.e.,

$$\|\theta_Q\|_{\ell^{2/(1+\beta)}} \lesssim 2^{-j(\alpha+1)}.$$

□

The following section is devoted to the proof of Theorem 6.1.4.

## 6.3 Analysis of an Edge Fragment

Let us turn to the more complicated case $Q \in \mathcal{Q}_j^1$ and the proof of Theorem 6.1.4. In this case the cube $Q$ intersects the edge curve $\Gamma$ and $f_Q = f\omega_Q$ is accordingly called an *edge fragment*. The subsequent exposition is taken from [60, Sec. 6].

Again we follow the broad outline for the case $\beta = 2$ which has been established in [15], and again we will need to adapt the estimates of [15, Sec. 6] to functions of fractional order smoothness via the use of moduli of smoothness. This turns out to cause serious difficulties and forces us to turn to techniques based on the forward difference operator (6.6).

In order to prove Theorem 6.1.4 we need to analyze the decay of the sequence $\theta_Q = \{\langle f_Q, \psi_\mu \rangle\}_{\mu \in M_j} = \{\langle \widehat{f}_Q, \widehat{\psi}_\mu \rangle\}_{\mu \in M_j}$. To estimate these scalar products, we again study the localization of the function $\widehat{f}_Q$. As in the treatment of the smooth fragments, our investigation starts with some simplifying reductions.

First, we note that it suffices to prove Theorem 6.1.4 for an edge fragment $f_Q = f\omega_Q$, where $f \in \mathcal{E}^\beta([-1,1]^2;\nu)$ is a cartoon of the simple form $f = g\mathcal{X}_\mathcal{B}$ with $g \in C_0^\beta(\mathbb{R}^2)$. In fact, the curvelet coefficient sequence of a general edge fragment $f_Q = f^0\omega_Q + f^1\mathcal{X}_\mathcal{B}\omega_Q =: f_Q^0 + f_Q^1$ can be decomposed into $\theta_Q^{(0)} = \{\langle f_Q^0, \psi_\mu\rangle\}_{\mu\in M_j}$ and $\theta_Q^{(1)} = \{\langle f_Q^1, \psi_\mu\rangle\}_{\mu\in M_j}$. From Theorem 6.1.3 we already know

$$\|\theta_Q^{(0)}\|_{w\ell^{2/(1+\beta)}} \lesssim 2^{-(1+\alpha)j} \lesssim 2^{-(1+\alpha)j/2}\,.$$

Therefore it only remains to show $\|\theta_Q^{(1)}\|_{w\ell^{2/(1+\beta)}} \lesssim 2^{-(1+\alpha)j/2}$. Since $\mathcal{B} \subset [-1,1]^2$, we can further smoothly cut off $f^1 \in C^\beta(\mathbb{R}^2)$ outside of $[-1,1]^2$ to obtain a function $g \in C_0^\beta(\mathbb{R}^2)$ such that $f_Q^1 = g\mathcal{X}_\mathcal{B}\omega_Q$.

Second, without loss of generality we restrict ourselves to the following model situation. The cube $Q$ is centered at the origin and the edge curve $\Gamma$ is the graph of a function $E : [-2^{-j\alpha}, 2^{-j\alpha}] \to [-2^{-j\alpha}, 2^{-j\alpha}]$ belonging to $C^\beta(\mathbb{R})$, with $x_1 = E(x_2)$. Further, it shall hold $E(0) = E'(0) = 0$, so that $\Gamma$ approximates a vertical line through the origin. If the scale $j$ is big enough, say bigger than some fixed base scale $j_0 \in \mathbb{N}_0$, it is always possible to arrive at this setting by possibly translating or rotating the coordinate axes. Henceforth, we assume $j \in \mathbb{N}$ and $j \geq j_0$ which clearly poses no loss of generality.

In this simplified model situation the edge fragment $f_Q$ can be written in the form

$$f_j(x) := f_Q(x) = \omega(2^{\alpha j}x)g(x)\mathcal{X}_{\{x_1\geq E(x_2)\}}, \quad x = (x_1,x_2) \in \mathbb{R}^2, \tag{6.13}$$

where $g \in C_0^\beta(\mathbb{R}^2)$, and $\omega \in C_0^\infty(\mathbb{R}^2)$ is the nonnegative window with $\operatorname{supp}\omega \subset [-1,1]^2$, generating the partition of unity $\{\omega_Q\}_Q$.

As in the discussion of the smooth fragments in the previous section, we introduce the notation $f_j := f_Q$ for the *standard edge fragment* (6.13) to indicate the sole dependence on the scale $j$. In addition, it is again more convenient to work with rescaled versions $F_j$ of the edge fragments $f_j$. Therefore, we put $g_j := g(2^{-\alpha j}\cdot)$ and define

$$F_j(x) := \omega(x)g_j(x)\mathcal{X}_{\{x_1\geq E_j(x_2)\}}, \quad x = (x_1,x_2) \in \mathbb{R}^2, \tag{6.14}$$

with the rescaled edge functions

$$E_j : [-1,1] \to [-1,1]\ ,\ E_j(x_2) := 2^{\alpha j}E(2^{-\alpha j}x_2).$$

It holds $E_j \in C^\beta([-1,1])$ with $E_j' = E'(2^{-\alpha j}\cdot)$ and $\textit{Höl}(E_j', \beta-1) \leq \delta_j$, where

$$\delta_j := 2^{-j(1-\alpha)} \cdot \textit{Höl}(E', \beta-1).$$

Observe that $\textit{Höl}(E', \beta-1)$ is a constant independent of the scale $j$. Together with $E_j(0) = E_j'(0) = 0$ this implies for all $u \in [-1,1]$ that

$$|E_j(u)| \leq \delta_j \quad \text{and} \quad |E_j'(u)| \leq \delta_j. \tag{6.15}$$

For convenience, we continuously extend the function $E_j$ to the whole of $\mathbb{R}$ by attaching straight lines on the left and on the right, with constant slopes $E_j'(1)$ and $E_j'(-1)$ respectively. Since this extension occurs outside of the square $[-1,1]^2$, it does not change the representation (6.14) of the edge fragment. Furthermore, it also does not alter the regularity and the Hölder constant.

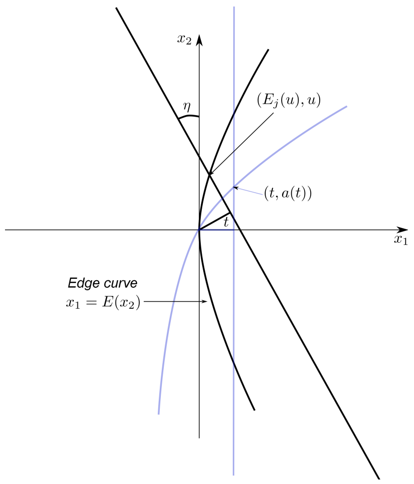



Figure 6.1: Illustration of standard edge fragment.

### 6.3.1 Fourier Analysis of an Edge Fragment

Our first goal is to analyze the Fourier transform $\widehat{F_j}$ (and thus also $\widehat{f_j}$) along radial lines, whose orientations are specified by angles $\eta \in [-\pi/2, \pi/2]$ with respect to the $x_1$-axis. If the angle $\eta$ satisfies $|\sin\eta| > \delta_j$, it is possible because of (6.15) to define a function $u = u_j(\cdot, \eta) : \mathbb{R} \to \mathbb{R}$ implicitly by

$$E_j(u(t))\cos\eta + u(t)\sin\eta = t. \tag{6.16}$$

The value $u(t)$ is the $x_2$-coordinate of the intersection point of the (extended) edge curve $\Gamma$ and the line $\mathfrak{L}_{t,\eta}$ defined by

$$\mathfrak{L}_{t,\eta} := \Big\{ x = (x_1, x_2) \in \mathbb{R}^2 \ : \ x_1\cos\eta + x_2\sin\eta = t \Big\}. \tag{6.17}$$

Further, we can define the function $a = a_j(\cdot, \eta) : \mathbb{R} \to \mathbb{R}$ by

$$a(t) := -E_j(u(t))\sin\eta + u(t)\cos\eta. \tag{6.18}$$

The value $a(t)$ is the $x_2$-coordinate of the point $(E_j(u), u)^T \in \Gamma$ in the coordinate system rotated by the angle $\eta$. For an illustration we refer to Figure 6.1.

The functions $u$ and $a$ are strictly monotone, increasing if $\eta > 0$ and decreasing if $\eta < 0$. Note, that we suppressed the dependence of $u$ and $a$ on $j$ and $\eta$ in the notation. The following lemma studies the regularity of $u$ under the assumption $|\sin\eta| \geq 2\delta_j$.

**Lemma 6.3.1** ([60, Lem. 6.1])**.** *Assume $|\sin\eta| \geq 2\delta_j$. Then the function $u : \mathbb{R} \to \mathbb{R}$ defined implicitly by* (6.16) *belongs to $C^\beta(\mathbb{R})$. Moreover, we have $\|u'\|_\infty \lesssim |\sin\eta|^{-1}$ and*

$$\|\Delta_h u'\|_\infty \lesssim \delta_j h^{\beta-1} |\sin\eta|^{-1-\beta},$$

*where the implicit constants are independent of the scale $j$, the angle $\eta$, and $h \geq 0$.*

*Proof.* First of all it is not difficult to show that $u = u_j(\cdot, \eta) \in C^1(\mathbb{R})$ with

$$u'(t) = \big(\sin\eta + E_j'(u(t))\cos\eta\big)^{-1}.$$

Under the assumption $|\sin\eta| \geq 2\delta_j$ it follows $\|u'\|_\infty \lesssim |\sin\eta|^{-1}$ because of $|E_j'(u)| \leq \delta_j \leq \frac{1}{2}|\sin\eta|$ for all $u \in [-1,1]$. Finally, we examine $\Delta_h u'$. For $t \in \mathbb{R}$

$$\begin{aligned}\Delta_h u'(t) &= u'(t+h) - u'(t) = u'(t+h)u'(t)\big(u'(t)^{-1} - u'(t+h)^{-1}\big)\\ &= u'(t+h)u'(t)\cos\eta\big(E_j'(u(t)) - E_j'(u(t+h))\big).\end{aligned}$$

Using *Höl*$(E_j', \beta-1) \leq \delta_j$ and the mean value theorem leads to

$$\|\Delta_h u'\|_\infty \leq \|u'\|_\infty^2 \delta_j \|\Delta_h u\|_\infty^{\beta-1} \leq \|u'\|_\infty^{\beta+1}\delta_j h^{\beta-1} \lesssim \delta_j h^{\beta-1}|\sin\eta|^{-1-\beta}.$$

□

The following lemma collects some properties of the function $a : \mathbb{R} \to \mathbb{R}$ defined in (6.18).

**Lemma 6.3.2** ([60, Lem. 6.2])**.** *Assume $|\sin\eta| \geq 2\delta_j$. It holds $a \in C^\beta(\mathbb{R})$ with*

$$\|a'\|_\infty \lesssim |\sin\eta|^{-1}, \quad \|\Delta_h a\|_\infty \lesssim h|\sin\eta|^{-1}, \quad \|\Delta_h a'\|_\infty \lesssim \delta_j h^{\beta-1}|\sin\eta|^{-1-\beta},$$

*with implicit constants independent of $j$, $\eta$, and $h \geq 0$.*

*Proof.* This is an easy consequence of the properties of $u$ proved in the previous lemma. □

Next, we introduce the scale-dependent interval

$$I(\eta) := I_j(\eta) := [a_j(\eta), b_j(\eta)],$$

where $a_j(\eta) = E_j(-1)\cos\eta - \sin\eta$ and $b_j(\eta) = E_j(1)\cos\eta + \sin\eta$. The restrictions of $u$ and $a$ to $I(\eta)$ correspond precisely to that part of the edge curve $\Gamma$ lying inside the square $[-1,1]^2$. In particular, we have a bijection $u : I(\eta) \to [-1,1]$. In the sequel it is more convenient to work with an extension of $I(\eta)$, given by

$$\widetilde{I}(\eta) := \widetilde{I}_j(\eta) := [a_j(\eta) - C\delta_j, b_j(\eta) + C\delta_j], \tag{6.19}$$

for some suitable fixed constant $C > 0$.

**Lemma 6.3.3** ([60, Lem. 6.3])**.** *For $|\sin\eta| > \delta_j$ we have*

$$|I(\eta)| \lesssim |\sin\eta| \quad \textit{and} \quad |\widetilde{I}(\eta)| \lesssim |\sin\eta|.$$

*Proof.* In view of $|\sin\eta| > \delta_j$ and (6.15) we can estimate

$$|I(\eta)| \leq |E_j(1) - E_j(-1)||\cos\eta| + 2|\sin\eta| \leq 2\delta_j + 2|\sin\eta| \lesssim |\sin\eta|.$$

The estimate for $\widetilde{I}(\eta)$ then follows directly from $|\sin\eta| > \delta_j$. □

We want to analyze $\widehat{F}_j$ along lines through the origin with orientation $\eta \in [-\pi/2, \pi/2]$. The central tool in this investigation is the Fourier slice theorem. In view of this theorem it makes sense to first study the Radon transform $\mathcal{R}F_j$ (see (1.5) for a definition), in particular its regularity. By Paley-Wiener type arguments we can then later extract information about the decay of $\widehat{F}_j$. Basically, this is the same approach taken in [15]. Due to the lack of regularity in our case, however, we have to use a more refined technique in this investigation. The main idea is to use finite differences instead of derivatives.

The value $\mathcal{R}F_j(t,\eta)$ of the Radon transform is obtained by integrating $F_j$ along the line $\mathfrak{L}_{t,\eta}$ defined in (6.17). Rotating $F_j$ by the angle $\eta$ yields the function $F_j^\eta$ and we can write

$$(\mathcal{R}F_j)(t,\eta) = \int_{\mathbb{R}} F_j^\eta(t,u)\,du.$$

The rescaled edge fragment $F_j$ can be rewritten as the product $F_j = G_j \mathcal{X}_{\{x_1 \geq E_j(x_2)\}}$ with the function

$$G_j := \omega g(2^{-\alpha j}\cdot) = \omega g_j. \tag{6.20}$$

Then we have

$$(\mathcal{R}F_j)(t,\eta) = \int_{-\infty}^{a(t,\eta)} G_j^\eta(t,u)\,du, \tag{6.21}$$

where $G_j^\eta$ is the function obtained by rotating $G_j$ by the angle $\eta$. Using the notation $g_j^\eta$ and $\omega^\eta$ for the rotated versions of $g_j$ and $\omega$, the integrand of (6.21) takes the form

$$G_j^\eta = g_j^\eta \omega^\eta. \tag{6.22}$$

We see, that the component $g_j^\eta = g^\eta(2^{-\alpha j}\cdot) \in C_0^\beta(\mathbb{R}^2)$ of $G_j^\eta$ is scaled and the window $\omega^\eta \in C_0^\infty(\mathbb{R}^2)$ remains fixed.

The central lemma of this subsection is given below. Its proof relies on estimates of the functions $g_j^\eta$ and $\omega^\eta$ and is outsourced to Section 6.4.

**Lemma 6.3.4** ([60, Lem. 6.4]). *Assume that* $|\sin\eta| \geq 2\delta_j$. *For* $h = C2^{-(1-\alpha)j}$, *where* $C > 0$ *is some fixed constant, we then have*

$$\Delta_h \partial_1 \mathcal{R}F_j(t,\eta) = S_{1,j}(t,\eta) + S_{2,j}(t,\eta)$$

*with functions* $S_{1,j}, S_{2,j}$ *such that*

$$\|S_{1,j}(\cdot,\eta)\|_2^2 \lesssim \delta_j^2 h^{2(\beta-1)} |\sin\eta|^{-1-2\beta},$$
$$\|\Delta_{(h,0)} S_{2,j}(\cdot,\eta)\|_2^2 \lesssim h^{2\beta} |\sin\eta|^{-1-2\beta},$$

*where the implicit constants are independent of the scale* $j$ *and the angle* $\eta$.

The previous lemma is the key to the following proposition. The notation $|\lambda| \sim 2^{(1-\alpha)j}$ indicates that $|\lambda| \in [C_1 2^{(1-\alpha)j}, C_2 2^{(1-\alpha)j}]$ for fixed constants $0 < C_1 \leq C_2 < \infty$. A typical choice would be $C_1 = 1$ and $C_2 = 2^{1-\alpha}$.

**Proposition 6.3.5** ([60, Prop. 6.5]). *It holds*

$$\int_{|\lambda|\sim 2^{(1-\alpha)j}} |\widehat{F}_j(\lambda\cos\eta, \lambda\sin\eta)|^2\, d\lambda \lesssim 2^{-(1-\alpha)j}\Big(1 + 2^{(1-\alpha)j}|\sin\eta|\Big)^{-1-2\beta}$$

*with an implicit constant independent of $j$ and $\eta$.*

*Proof.* First we assume $|\sin\eta| \geq 2\delta_j$. The integration domain is $[C_1 2^{(1-\alpha)j}, C_2 2^{(1-\alpha)j}]$ for fixed constants $0 < C_1 \leq C_2 < \infty$. Let us fix $h := C2^{-j(1-\alpha)}$, where $C > 0$ is chosen such that $C_2 C < 2\pi$. For this choice of $h$ there is $c > 0$ such that $|e^{i\lambda h} - 1|^2 \geq c$ for all $|\lambda| \in [C_1 2^{(1-\alpha)j}, C_2 2^{(1-\alpha)j}]$ at all scales. We conclude, where $S_{1,j}$ and $S_{2,j}$ denote the entities from Lemma 6.3.4:

$$\begin{aligned}
\int\limits_{|\lambda|\sim 2^{j(1-\alpha)}} |\lambda|^2 |\widehat{F}_j(\lambda\cos\eta, \lambda\sin\eta)|^2\, d\lambda &\lesssim \int\limits_{|\lambda|\sim 2^{j(1-\alpha)}} |e^{i\lambda h} - 1|^2 |\lambda|^2 |\widehat{F}_j(\lambda\cos\eta, \lambda\sin\eta)|^2\, d\lambda \\
&= \int\limits_{|\lambda|\sim 2^{j(1-\alpha)}} |\mathcal{F}[\Delta_h \partial_1 \mathcal{R} F_j(\cdot,\eta)](\lambda)|^2\, d\lambda \\
&\lesssim \int\limits_{|\lambda|\sim 2^{j(1-\alpha)}} |\widehat{S_{1,j}(\cdot,\eta)}(\lambda)|^2\, d\lambda + \int\limits_{|\lambda|\sim 2^{j(1-\alpha)}} |\widehat{S_{2,j}(\cdot,\eta)}(\lambda)|^2\, d\lambda \\
&\lesssim \int\limits_{|\lambda|\sim 2^{j(1-\alpha)}} |\widehat{S_{1,j}(\cdot,\eta)}(\lambda)|^2\, d\lambda + \int\limits_{|\lambda|\sim 2^{j(1-\alpha)}} |e^{i\lambda h} - 1|^2 |\widehat{S_{2,j}(\cdot,\eta)}(\lambda)|^2\, d\lambda \\
&\leq \int_{\mathbb{R}} |\widehat{S_{1,j}(\cdot,\eta)}(\lambda)|^2\, d\lambda + \int_{\mathbb{R}} |e^{i\lambda h} - 1|^2 |\widehat{S_{2,j}(\cdot,\eta)}(\lambda)|^2\, d\lambda \\
&= \|S_{1,j}(\cdot,\eta)\|_2^2 + \|\Delta_h S_{2,j}(\cdot,\eta)\|_2^2 \lesssim h^{2\beta}|\sin\eta|^{-1-2\beta}.
\end{aligned}$$

It follows

$$\int_{|\lambda|\sim 2^{j(1-\alpha)}} |\widehat{F}_j(\lambda\cos\eta, \lambda\sin\eta)|^2\, d\lambda \lesssim 2^{-j(1-\alpha)}\big(2^{j(1-\alpha)}|\sin\eta|\big)^{-1-2\beta}.$$

Next, we handle the case $|\sin\eta| < 2\delta_j$. We want to show

$$\int_{|\lambda|\sim 2^{j(1-\alpha)}} |\widehat{F}_j(\lambda\cos\eta, \lambda\sin\eta)|^2\, d\lambda \lesssim 2^{-j(1-\alpha)}. \tag{6.23}$$

Altogether we then obtain the desired estimate

$$\int_{|\lambda|\sim 2^{(1-\alpha)j}} |\widehat{F}_j(\lambda\cos\eta, \lambda\sin\eta)|^2\, d\lambda \lesssim 2^{-(1-\alpha)j}\Big(1 + 2^{(1-\alpha)j}|\sin\eta|\Big)^{-1-2\beta},$$

since $2^{(1-\alpha)j}|\sin\eta| \gtrsim 1$ if $|\sin\eta| \geq 2\delta_j$ and $2^{(1-\alpha)j}|\sin\eta| \lesssim 1$ if $|\sin\eta| < 2\delta_j$.

It remains to show (6.23). For this we write the edge fragment as a sum $F_j = F_j^0 + F_j^1$, where

$$F_j^0(x) = g(2^{-j\alpha}x)\omega(x)\mathcal{X}_{\{x_1 \geq \delta_j\}}, \quad x \in \mathbb{R}^2, \tag{6.24}$$

is a fragment with a straight edge and $F_j^1(x) = F_j(x) - F_j^0(x)$ is the deviation.

The function $F_j^1$ is supported in a vertical strip around the $x_2$-axis of width $2\delta_j$. For $\eta$ satisfying $|\sin\eta| < 2\delta_j$ the Radon transform $\mathcal{R}F_j^1(\cdot,\eta)$ is $L_\infty$-bounded and supported in an interval of length

$$2(\delta_j \cos\eta + \sin\eta) \lesssim \delta_j.$$

It follows $\|\mathcal{R}F_j^1(\cdot,\eta)\|_2^2 \lesssim \delta_j \lesssim 2^{-j(1-\alpha)}$, and therefore

$$\int_{|\lambda|\sim 2^{j(1-\alpha)}} |\widehat{F_j^1}(\lambda\cos\eta,\lambda\sin\eta)|^2\, d\lambda \le \int_{\mathbb{R}} |\widehat{\mathcal{R}F_j^1(\cdot,\eta)}(\lambda)|^2\, d\lambda \lesssim 2^{-j(1-\alpha)}.$$

Finally, the estimate

$$\int_{|\lambda|\sim 2^{j(1-\alpha)}} |\widehat{F_j^0}(\lambda\cos\eta,\lambda\sin\eta)|^2\, d\lambda \lesssim 2^{-j(1-\alpha)}$$

follows from the fact, that we have decay $|\widehat{F_j^0}(\lambda,0)| \sim |\lambda|^{-1/2}$ normal to the straight singularity curve, and that the second argument $(\lambda\sin\eta)$ remains bounded due to the condition $|\sin\eta| < 2\delta_j$. This finishes the proof. $\square$

A direct consequence is the following theorem.

**Theorem 6.3.6** ([60, Thm. 6.6]). *We have*

$$\int_{|\lambda|\sim 2^j} |\widehat{f}_j(\lambda\cos\eta,\lambda\sin\eta)|^2\, d\lambda \lesssim 2^{-(1+2\alpha)j}\Big(1+2^{(1-\alpha)j}|\sin\eta|\Big)^{-1-2\beta}.$$

*Proof.* The statement follows directly from the relation $\widehat{f}_j(\xi) = 2^{-2\alpha j}\widehat{F_j}(2^{-\alpha j}\xi)$. $\square$

A refinement of the discussion in this subsection, which can be found in Section 6.5, yields the following theorem.

**Theorem 6.3.7** ([60, Thm. 6.7]). *We have for $m=(m_1,m_2)\in\mathbb{N}_0^2$ the estimate*

$$\begin{aligned}\int\limits_{|\lambda|\sim 2^j} &|\partial^m \widehat{f}_j(\lambda\cos\eta,\lambda\sin\eta)|^2\, d\lambda \\ &\lesssim 2^{-j2\alpha|m|_1}\Big(2^{-j2(1-\alpha)m_1}2^{-(1+2\alpha)j}\Big(1+2^{(1-\alpha)j}|\sin\eta|\Big)^{-1-2\beta} + 2^{(-1-2\beta)j}\Big).\end{aligned}$$

### 6.3.2 Curvelet Analysis of an Edge Fragment

In the following $J=(j,\ell)$ shall denote a scale-angle pair with $j\in\mathbb{N}_0$, $\ell\in\{0,\dots,L_j-1\}$, and $W_J$ shall be the wedge functions from (5.6). Recall also the characteristic angle $\omega_j = \pi 2^{-\lfloor j(1-\alpha)\rfloor}$ at scale $j$ and the corresponding orientations $\omega_J := \ell\omega_j$, ranging between 0 and $\pi$. From Theorem 6.3.6 we can directly conclude the following result.

**Theorem 6.3.8** ([60, Thm. 6.8]). *We have*

$$\int_{\mathbb{R}^2} |\widehat{f}_j W_J(\xi)|^2\, d\xi \lesssim 2^{-(1+\alpha)j}\Big(1+2^{(1-\alpha)j}|\sin\omega_J|\Big)^{-1-2\beta}. \tag{6.25}$$

*Proof.* It holds $\|W_J\|_\infty \leq 1$ and $\operatorname{supp} W_J \subset \mathcal{W}_J^+$, where $\mathcal{W}_J^+$ is the wedge defined in (5.9). Let us define the intervals $\mathcal{K}_j = \frac{1}{6\pi}[2^{j-1}, 2^{j+1}]$ and $\mathcal{A}_J = [\omega_J - 3\omega_j/4, \omega_J + 3\omega_j/4]$. Using Theorem 6.3.6 we calculate

$$\begin{aligned}\int_{\mathbb{R}^2} |\widehat{f}_j W_J(\xi)|^2\, d\xi &\leq \int_{\mathcal{W}_J^+} |\widehat{f}_j(\xi)|^2\, d\xi \\ &= \int_{\mathcal{K}_j} \Big( \int_{\mathcal{A}_J} + \int_{\pi+\mathcal{A}_J} \Big) |\widehat{f}_j(\lambda,\eta)|^2 \lambda \, d\eta\, d\lambda \\ &\lesssim \Big( \int_{\mathcal{A}_J} + \int_{\pi+\mathcal{A}_J} \Big) 2^{-(1+2\alpha)j} \Big( 1 + 2^{(1-\alpha)j} |\sin\eta| \Big)^{-1-2\beta} 2^j \, d\eta \\ &\lesssim 2^{-(1+\alpha)j} \Big( 1 + 2^{(1-\alpha)j} |\sin\omega_J| \Big)^{-1-2\beta}.\end{aligned}$$

□

For a scale-angle pair $J = (j, \ell)$ let us define the quantity

$$\ell_J = 1 + 2^{(1-\alpha)j} |\sin\omega_J| \tag{6.26}$$

and the differential operator

$$\mathcal{L} = (\mathcal{I} - (2^j/\ell_J)^2 \mathcal{D}_1^2)(\mathcal{I} - 2^{2\alpha j}\mathcal{D}_2^2) = \mathcal{I} - 2^{2j}\ell_J^{-2}\mathcal{D}_1^2 - 2^{2\alpha j}\mathcal{D}_2^2 + 2^{2(1+\alpha)j}\ell_J^{-2}\mathcal{D}_1^2\mathcal{D}_2^2, \tag{6.27}$$

with identity $\mathcal{I}$ and partial derivatives

$$\mathcal{D}_1 = \cos\omega_J \cdot \partial_1 + \sin\omega_J \cdot \partial_2 \quad \text{and} \quad \mathcal{D}_2 = -\sin\omega_J \cdot \partial_1 + \cos\omega_J \cdot \partial_2. \tag{6.28}$$

We will show that $\mathcal{L}(\widehat{f}_j W_J)$ obeys the same estimate (6.25) as $\widehat{f}_j W_J$. The key result for this statement is Theorem 6.3.7

**Theorem 6.3.9** ([60, Thm. 6.9])**.** *Let $\mathcal{L}$ be the differential operator defined in* (6.27)*. We have*

$$\int_{\mathbb{R}^2} |\mathcal{L}(\widehat{f}_j W_J)(\xi)|^2\, d\xi \lesssim 2^{-(1+\alpha)j} \Big( 1 + 2^{(1-\alpha)j} |\sin\omega_J| \Big)^{-1-2\beta}.$$

*Proof.* First, observe that for each pair $m = (m_1, m_2) \in N_0^2$ the mixed derivative of $W_J$ obeys

$$\|\mathcal{D}_1^{m_1}\mathcal{D}_2^{m_2} W_J\|_\infty = \mathcal{O}(2^{-jm_1} \cdot 2^{-j\alpha m_2}). \tag{6.29}$$

This follows from the fact, that the functions $W_J$ from (5.6) scale with their support wedges $\mathcal{W}_J^+$, which are of length $\sim 2^j$ and width $\sim 2^{\alpha j}$.

Next, from the definition (6.28) of the operators $\mathcal{D}_1$ and $\mathcal{D}_2$ we deduce for $m_1 \in \mathbb{N}_0$

$$\mathcal{D}_1^{m_1}\widehat{f}_j = \sum_{a+b=m_1} c_{a,b} (\cos\omega_J)^a (\sin\omega_J)^b \partial^{(a,b)} \widehat{f}_j$$

with binomial coefficients $c_{a,b} \in \mathbb{N}$. A similar formula holds for $\mathcal{D}_2^{m_2}\widehat{f}_j$ and $m_2 \in \mathbb{N}_0$. Using $|\sin\omega_J| \leq 2^{-(1-\alpha)j}\ell_J$ with the quantity $\ell_J$ from (6.26) we obtain the estimate

$$\begin{aligned}\|\mathcal{D}_1^{m_1}\widehat{f}_j\|^2_{L_2(\mathcal{W}_J^+)} &\lesssim \sum_{a+b=m_1} |\sin\omega_J|^{2b} \cdot \|\partial^{(a,b)}\widehat{f}_j\|^2_{L_2(\mathcal{W}_J^+)} \\ &\leq \sum_{a+b=m_1} (2^{-(1-\alpha)j}\ell_J)^{2b} \cdot \|\partial^{(a,b)}\widehat{f}_j\|^2_{L_2(\mathcal{W}_J^+)}.\end{aligned}$$

Analogously, we obtain

$$\|\mathcal{D}_2^{m_2}\widehat{f}_j\|^2_{L_2(\mathcal{W}_J^+)} \lesssim \sum_{a+b=m_2} \|\partial^{(a,b)}\widehat{f}_j\|^2_{L_2(\mathcal{W}_J^+)}.$$

Taking into account the width $\sim 2^{\alpha j}$ of the wedges $\mathcal{W}_J^+$, Theorem 6.3.7 gives for $(a,b) \in \mathbb{N}_0^2$ the bound

$$\|\partial^{(a,b)}\widehat{f}_j\|^2_{L_2(\mathcal{W}_J^+)} \le C_{a,b}2^{\alpha j} \cdot 2^{-j2\alpha(a+b)}\big(2^{-2j(1-\alpha)a}2^{-(1+2\alpha)j}\ell_J^{-1-2\beta} + 2^{(-1-2\beta)j}\big),$$

with some constant $C_{a,b} > 0$ independent of scale. Therefore we can estimate for $m_1 \in \mathbb{N}_0$

$$\|\mathcal{D}_1^{m_1}\widehat{f}_j\|^2_{L_2(\mathcal{W}_J^+)} \lesssim 2^{-j2\alpha m_1}\big(2^{-2j(1-\alpha)m_1}2^{-(1+\alpha)j}\ell_J^{2m_1-1-2\beta} + 2^{\alpha j}2^{(-1-2\beta)j}\big).$$

If $m_1 \le 2$ this further simplifies to

$$\|\mathcal{D}_1^{m_1}\widehat{f}_j\|^2_{L_2(\mathcal{W}_J^+)} \lesssim 2^{-2jm_1}2^{-(1+\alpha)j}\ell_J^{2m_1-1-2\beta}, \tag{6.30}$$

since for every $m_1 \le (1+\alpha)/\alpha$ we have

$$2^{\alpha j}2^{(-1-2\beta)j} \lesssim 2^{-2j(1-\alpha)m_1}2^{-(1+\alpha)j}\ell_J^{2m_1-1-2\beta}.$$

Similar calculations lead to

$$\|\mathcal{D}_2^{m_2}\widehat{f}_j\|^2_{L_2(\mathcal{W}_J^+)} \lesssim 2^{-2\alpha j m_2}2^{-(1+\alpha)j}\ell_J^{-1-2\beta}. \tag{6.31}$$

Indeed, if $a+b=m_2$ we have

$$\|\partial_1^a\partial_2^b\widehat{f}_j\|^2_{L_2(\mathcal{W}_J^+)} \le C_{a,b}\cdot 2^{\alpha j}\cdot 2^{-2\alpha j m_2}\big(2^{-(1+2\alpha)j}\ell_J^{-1-2\beta} + 2^{(-1-2\beta)j}\big).$$

Since $1 \le \ell_J \le 2\cdot 2^{(1-\alpha)j}$ it holds

$$2^{j(1+\alpha)}2^{-2\beta j} = 2^{-(1-\alpha)(1+2\beta)j} \lesssim \ell_J^{-1-2\beta}.$$

Therefore, taking into account $1 \le 2^{(1-\alpha)j}$, we can conclude

$$2^{(-1-2\beta)j} \le 2^{j(1+\alpha)}2^{-2\beta j}2^{-(1+2\alpha)j} \lesssim 2^{-(1+2\alpha)j}\ell_J^{-1-2\beta}.$$

Altogether we obtain the desired estimate (6.31).

After this preliminary work we can finally prove the statement of Theorem 6.3.9. We have

$$\mathcal{L}(\widehat{f}_jW_J) = \widehat{f}_jW_J - 2^{2j}\ell_J^{-2}\mathcal{D}_1^2(\widehat{f}_jW_J) - 2^{2\alpha j}\mathcal{D}_2^2(\widehat{f}_jW_J) + 2^{2(1+\alpha)j}\ell_J^{-2}\mathcal{D}_1^2\mathcal{D}_2^2(\widehat{f}_jW_J),$$

which allows us to show the desired estimate for each term separately. For $\widehat{f}_jW_J$ the estimate holds true by Theorem 6.3.8.

Let us turn to the second term. The product rule yields

$$\mathcal{D}_1^2(\widehat{f}_jW_J) = (\mathcal{D}_1^2\widehat{f}_j)W_J + 2(\mathcal{D}_1\widehat{f}_j)(\mathcal{D}_1W_J) + \widehat{f}_j(\mathcal{D}_1^2W_J).$$

The previous estimates together with the Hölder inequality then lead to

$$\|\mathcal{D}_1^2(\widehat{f}_j W_J)\|_2^2 \lesssim 2^{-4j}\ell_J^4 2^{-(1+\alpha)j}\ell_J^{-1-2\beta}.$$

Here (6.30) was used, and that $\|\mathcal{D}_1^{m_1}W_J\|_\infty^2 \le 2^{-2jm_1} \le 2^{-2jm_1}\ell_J^{2m_1}$ by (6.29). This settles the claim for the second term.

Analogously, we can deduce

$$\|\mathcal{D}_2^2(\widehat{f}_j W_J)\|_2^2 \lesssim 2^{-4\alpha j}2^{-(1+\alpha)j}\ell_J^{-1-2\beta},$$

using (6.31) and that $\|\mathcal{D}_2^{m_2}W_J\|_\infty^2 \le 2^{-2\alpha j m_2}$ by (6.29). This gives the estimate for the third term.

Finally, it also holds

$$\|\mathcal{D}_1^2\mathcal{D}_2^2(\widehat{f}_j W_J)\|_2^2 \lesssim 2^{-4j(1+\alpha)}\ell_J^4 2^{-(1+\alpha)j}\ell_J^{-1-2\beta},$$

which establishes the result for the fourth term. □

At last we are ready to give the proof of Theorem 6.1.4. The essential tool is Theorem 6.3.9.

### 6.3.3 Proof of Theorem 6.1.4

*Proof.* Recall the curvelet frame $\mathfrak{C}_\alpha^\bullet = \{\psi_\mu\}_{\mu\in M}$. On the Fourier side we have

$$\widehat{\psi}_{j,\ell,k} = W_J u_{j,k}(R_J\cdot),$$

with rotation matrix $R_J$ given as in (3.24) and functions

$$u_{j,k}(\xi) := 2^{-j(1+\alpha)/2}e^{2\pi i(2^{-j}k_1,2^{-\alpha j}k_2)\cdot\xi}, \quad \xi\in\mathbb{R}^2.$$

The elements $\tilde{\theta}_{j,\ell,k} := \langle f_j, \psi_{j,\ell,k}\rangle$ of the sequence $\theta_Q$ are therefore given by the formula

$$\tilde{\theta}_{j,\ell,k} = \int_{\mathbb{R}^2} \widehat{f}_j W_J(\xi)\overline{u_{j,k}}(R_J\xi)\,d\xi.$$

Since

$$\mathcal{L}(u_{j,k}) = (1+\ell_J^{-2}k_1^2)(1+k_2^2)u_{j,k},$$

integration by parts yields

$$\tilde{\theta}_{j,\ell,k} = (1+\ell_J^{-2}k_1^2)^{-1}(1+k_2^2)^{-1}\int_{\mathbb{R}^2}\mathcal{L}(\widehat{f}_j W_J)(\xi)\overline{u_{j,k}}(R_J\xi)\,d\xi.$$

Let $J=(j,\ell)$ be a scale-angle pair, $K=(K_1,K_2)\in\mathbb{Z}^2$ and define

$$\mathfrak{Z}_{J,K} := \Big\{(k_1,k_2)\in\mathbb{Z}^2 \;:\; \ell_J^{-1}k_1\in[K_1,K_1+1),\, k_2=K_2\Big\}.$$

For fixed $J=(j,\ell)$ the Fourier system $\{u_{j,k}(R_J\cdot)\}_{k\in\mathbb{Z}^2}$ is an orthonormal basis for $L_2(\Xi_J)$, where $\Xi_J$ is the rectangle defined in (3.20) containing the support of $W_J$. Therefore,

$$\sum_{k\in\mathfrak{Z}_{J,K}}|\tilde{\theta}_{j,\ell,k}|^2 \lesssim (1+K_1^2)^{-2}(1+K_2^2)^{-2}\int_{\mathbb{R}^2}|\mathcal{L}(\widehat{f}_j W_J)(\xi)|^2\,d\xi.$$

The integral on the right-hand side is bounded by Theorem 6.3.9, and we thus arrive at

$$\sum_{k\in\mathfrak{Z}_{J,K}} |\tilde{\theta}_{j,\ell,k}|^2 \lesssim (L_K)^{-2} 2^{-(1+\alpha)j} \ell_J^{-1-2\beta} \tag{6.32}$$

with $L_K := (1+K_1^2)(1+K_2^2)$.

Let $M_J$ denote the subset of curvelet coefficients associated with a fixed scale-angle pair $J=(j,\ell)$. Further, let $N_{J,K}(\varepsilon)$ be the number of indices $\mu \in M_J$ such that $k \in \mathfrak{Z}_{J,K}$ and $|\tilde{\theta}_\mu| > \varepsilon$.

Since $\#\mathfrak{Z}_{J,K} \le \ell_J$ and because of (6.32) we can conclude

$$N_{J,K}(\varepsilon) \lesssim \min\Big\{\ell_J, (\varepsilon L_K)^{-2} 2^{-(1+\alpha)j} \ell_J^{-1-2\beta}\Big\}. \tag{6.33}$$

For $\omega_J = \pi\ell 2^{-\lfloor j(1-\alpha)\rfloor} \in [0,\pi)$ let $\langle\omega_J\rangle$ denote the equivalent angle modulo $\pi$ in the interval $(-\pi/2,\pi/2]$. The corresponding indices in the range $\{\lfloor -L_j/2+1\rfloor,\dots,\lfloor L_j/2\rfloor\}$ shall be denoted by $\langle\ell\rangle$. Since it holds $|\sin\eta| \asymp |\eta|$ for $\eta\in[-\pi/2,\pi/2]$, it follows

$$\ell_J = 1 + 2^{(1-\alpha)j}|\sin\omega_J| = 1 + 2^{(1-\alpha)j}|\sin\langle\omega_J\rangle| \asymp 1 + |\langle\ell\rangle|. \tag{6.34}$$

Let $\ell_*$ be the solution of the equation $\ell_* = (\varepsilon L_K)^{-2}2^{-(1+\alpha)j}\ell_*^{-1-2\beta}$ and put $L^* = \lfloor \ell_* \rfloor$. Utilizing (6.33) and (6.34) yields

$$\begin{aligned}
\sum_{|J|=j} N_{J,K}(\varepsilon) &\lesssim \sum_{\substack{\ell\in\{0,\dots,L_j-1\}\\ |\langle\ell\rangle|\le L^*-1}} (1+|\langle\ell\rangle|) + \sum_{\substack{\ell\in\{0,\dots,L_j-1\}\\ |\langle\ell\rangle|\ge L^*}} (\varepsilon L_K)^{-2}2^{-(1+\alpha)j}(1+|\langle\ell\rangle|)^{-1-2\beta} \\
&\lesssim \sum_{\ell=0}^{L^*-1}(1+|\ell|) + \sum_{\ell=L^*}^{\infty} (\varepsilon L_K)^{-2}2^{-(1+\alpha)j}(1+|\ell|)^{-1-2\beta} \\
&\lesssim (L^*)^2 + (\varepsilon L_K)^{-2}2^{-(1+\alpha)j}(L^*)^{-2\beta}.
\end{aligned}$$

This translates to

$$\sum_{|J|=j} N_{J,K}(\varepsilon) \lesssim \varepsilon^{-2/(1+\beta)} \cdot L_K^{-2/(1+\beta)} \cdot 2^{-(1+\alpha)j/(1+\beta)}.$$

Since $\beta < 3$ we have $\sum_{K\in\mathbb{Z}^2} L_K^{-2/(1+\beta)} < \infty$. Hence

$$\#\Big\{\mu\in M_j, |\tilde{\theta}_\mu| > \varepsilon\Big\} = \sum_{K\in\mathbb{Z}^2}\sum_{|J|=j} N_{J,K}(\varepsilon) \lesssim 2^{-(1+\alpha)j/(1+\beta)}\varepsilon^{-2/(1+\beta)}.$$

This finishes the proof. □

## 6.4 Appendix A: *Proof of Lemma 6.3.4*

This section corresponds to [60, Appendix A]. Let us start with a simple result, that shows how scaling affects the Hölder constant.

**Lemma 6.4.1** ([60, Lem. 6.10])**.** *Let $f \in C^\alpha(\mathbb{R})$ and $0<\alpha<1$. Then for $t,s>0$*

$$\textit{Höl}(sf(t\cdot),\alpha) = st^\alpha \cdot \textit{Höl}(f,\alpha).$$

We proceed with some technical estimates of the functions $g_j^\eta$ and $\omega^\eta$, which occur as components of the functions $G_j^\eta$ defined in (6.22). These estimates will provide the basis for the more complex estimates needed in the actual proof of Lemma 6.3.4.

### 6.4.1 Estimates for $g_j^\eta$

The functions $g_j^\eta$ are given for $j \in \mathbb{N}_0$ by $g_j^\eta = g^\eta(2^{-\alpha j}\cdot)$, where $g^\eta$ is a rotated version of the fixed function $g \in C_0^\beta(\mathbb{R}^2)$ with $\beta \in (1,2]$. Thus clearly, $g^\eta \in C_0^\beta(\mathbb{R}^2)$ and also $g_j^\eta \in C_0^\beta(\mathbb{R}^2)$. However, the parameters of the regularity change. Applying Lemma 6.4.1 yields the following result.

**Lemma 6.4.2** ([60, Lem. 6.11])**.** *Let $\beta \in (1,2]$ and $g^\eta \in C_0^\beta(\mathbb{R}^2)$. Then for $g_j^\eta = g^\eta(2^{-\alpha j}\cdot)$*

$$Höl(\partial_1 g_j^\eta, \beta - 1) = 2^{-j} Höl(\partial_1 g^\eta, \beta - 1).$$

*Proof.* In view of Lemma 6.4.1 we have

$$Höl(\partial_1 g_j^\eta, \beta - 1) = Höl(2^{-\alpha j}\partial_1 g^\eta(2^{-\alpha j}\cdot), \beta - 1) = 2^{-j} Höl(\partial_1 g^\eta, \beta - 1).$$

□

It is obvious that $\|g_j^\eta\|_\infty \lesssim 1$. Further, the chain rule yields

$$\|\partial_1 g_j^\eta\|_\infty \lesssim 2^{-\alpha j} \quad \text{and} \quad \|\partial_2 g_j^\eta\|_\infty \lesssim 2^{-\alpha j}.$$

Some more estimates for $g_j^\eta$ are collected in the following two lemmas. Here $\Delta_{(h,0)}$ is a forward difference operator as in (6.6).

**Lemma 6.4.3** ([60, Lem. 6.12])**.** *The following estimates hold true for $g_j^\eta$:*

$$\begin{aligned} \|\Delta_{(h,0)} g_j^\eta\|_\infty &\lesssim 2^{-\alpha j} h, \\ \|\Delta_{(h,0)}\partial_1 g_j^\eta\|_\infty, \|\Delta_{(h,0)}\partial_2 g_j^\eta\|_\infty &\lesssim 2^{-j} h^{\beta-1} = 2^{-\alpha j} h^\beta, \\ \|\Delta_{(h,0)}^2 g_j^\eta\|_\infty &\lesssim 2^{-j} h^\beta, \end{aligned}$$

*with implicit constants, that do not depend on $j \in \mathbb{N}_0$ and $h \geq 0$.*

*Proof.* Applying the mean value theorem yields

$$\|\Delta_{(h,0)} g_j^\eta\|_\infty \leq h\|\partial_1 g_j^\eta\|_\infty \lesssim 2^{-\alpha j} h.$$

Considering Lemma 6.4.2 we obtain

$$\|\Delta_{(h,0)}\partial_1 g_j^\eta\|_\infty \lesssim 2^{-j} h^{\beta-1} = 2^{-\alpha j} h^\beta.$$

Noting the commutativity $\partial_1 \Delta_{(h,0)} = \Delta_{(h,0)}\partial_1$, we obtain

$$\|\Delta_{(h,0)}^2 g_j^\eta\|_\infty \lesssim h\|\Delta_{(h,0)}\partial_1 g_j^\eta\|_\infty \lesssim 2^{-j} h^\beta.$$

□

The next lemma gives estimates for $g_j^\eta$ along the edge curve. Here the function $a \in C^\beta(\mathbb{R})$ comes into play, which was defined in (6.18). The following estimates also depend on the properties of $a$, which are summarized in Lemma 6.3.2. Recall that by convention $\Delta_h$ only acts on the variables $t$ or $\tau$.

**Lemma 6.4.4** ([60, Lem. 6.13])**.** *Assume $|\sin\eta| \geq 2\delta_j$. The following estimates hold true for $g_j^\eta$:*

$$\sup_{t\in\mathbb{R}} |\Delta_h g_j^\eta(t,a(t))| \lesssim h|\sin\eta|^{-1}2^{-\alpha j},$$
$$\sup_{t\in\mathbb{R}} |\Delta_h \partial_1 g_j^\eta(t,a(t))|, \sup_{t\in\mathbb{R}} |\Delta_h \partial_2 g_j^\eta(t,a(t))| \lesssim h^{\beta-1}|\sin\eta|^{1-\beta}2^{-j},$$
$$\sup_{t\in\mathbb{R}} |\Delta_h^2 g_j^\eta(t,a(t))| \lesssim h^{\beta}|\sin\eta|^{-1-\beta}2^{-j},$$

*where the implicit constants are independent of $j \in \mathbb{N}_0$ and $h \geq 0$.*

*Proof.* In view of Lemma 6.3.2 it holds

$$\begin{aligned}
\sup_{t\in\mathbb{R}} |\Delta_h g_j^\eta(t,a(t))| &\lesssim h \cdot \sup_{t\in\mathbb{R}} |\frac{d}{dt} g_j^\eta(t,a(t))| \\
&\lesssim h \cdot \Big( \sup_{t\in\mathbb{R}} |\partial_1 g_j^\eta(t,a(t))| + \sup_{t\in\mathbb{R}} |\partial_2 g_j^\eta(t,a(t))a'(t)| \Big) \\
&\lesssim h \cdot |\sin\eta|^{-1}2^{-\alpha j}.
\end{aligned}$$

Considering the transformation behavior of the Hölder constant we obtain with Lemma 6.3.2

$$\begin{aligned}
\sup_{t\in\mathbb{R}} |\Delta_h \partial_1 g_j^\eta(t,a(t))| &\lesssim 2^{-j} \sup_{t\in\mathbb{R}} |(h, a(t+h)-a(t))|_2^{\beta-1} \\
&\lesssim 2^{-j}\big(h^{\beta-1} + \sup_{t\in\mathbb{R}} |a(t+h)-a(t)|^{\beta-1}\big) \\
&\lesssim 2^{-j}h^{\beta-1}|\sin\eta|^{1-\beta}.
\end{aligned}$$

Applying Lemma 6.3.2, the mean value theorem and $\frac{d}{dt}\Delta_h = \Delta_h \frac{d}{dt}$ yields

$$\begin{aligned}
\sup_{t\in\mathbb{R}} |\Delta_h^2 g_j^\eta(t,a(t))| &\lesssim h \cdot \sup_{t\in\mathbb{R}} |\Delta_h \frac{d}{dt} g_j^\eta(t,a(t))| \\
&= h \cdot \sup_{t\in\mathbb{R}} |\Delta_h(\partial_1 g_j^\eta(t,a(t)) + \partial_2 g_j^\eta(t,a(t))a'(t))| \\
&= h \cdot \sup_{t\in\mathbb{R}} |\Delta_h \partial_1 g_j^\eta(t,a(t)) + \Delta_h \partial_2 g_j^\eta(t,a(t))a'(t+h) + \partial_2 g_j^\eta(t,a(t))\Delta_h a'(t)| \\
&\lesssim h^\beta|\sin\eta|^{1-\beta}2^{-j} + h^\beta|\sin\eta|^{-\beta}2^{-j} + \delta_j h^\beta|\sin\eta|^{-1-\beta}2^{-\alpha j}.
\end{aligned}$$

□

### 6.4.2 Estimates for $\omega^\eta$

Similarly, we obtain estimates for the window function $\omega^\eta \in C_0^\infty(\mathbb{R}^2)$, which in contrast to the functions $g_j^\eta$ remains fixed at all scales. This fact and the smoothness of $\omega^\eta$ result in different estimates.

First, we state the trivial estimates $\|\omega^\eta\|_\infty \lesssim 1$, $\|\partial_1\omega^\eta\|_\infty \lesssim 1$, and $\|\partial_2\omega^\eta\|_\infty \lesssim 1$. Next, we apply the forward difference operator $\Delta_{(h,0)}$ to $\omega^\eta$.

**Lemma 6.4.5** ([60, Lem. 6.14])**.** *Let $k \in \mathbb{N}_0$. It holds with implicit constants independent of $h \geq 0$*

$$\|\Delta_{(h,0)}^k \omega^\eta\|_\infty \lesssim h^k \quad \textit{and} \quad \|\Delta_{(h,0)}^k \partial_1\omega^\eta\|_\infty \lesssim h^k.$$

Analogous to Lemma 6.4.4 we establish estimates along the edge curve.

**Lemma 6.4.6** ([60, Lem. 6.15])**.** *Assume* $|\sin\eta| \geq 2\delta_j$. *It holds*

$$\sup_{t\in\mathbb{R}} |\Delta_h \omega^\eta(t, a(t))| \lesssim h|\sin\eta|^{-1},$$
$$\sup_{t\in\mathbb{R}} |\Delta_h \partial_1 \omega^\eta(t, a(t))|, \sup_{t\in\mathbb{R}} |\Delta_h \partial_2 \omega^\eta(t, a(t))| \lesssim h|\sin\eta|^{-1},$$
$$\sup_{t\in\mathbb{R}} |\Delta_h^2 \omega^\eta(t, a(t))| \lesssim h^2|\sin\eta|^{-2} + \delta_j h^\beta |\sin\eta|^{-1-\beta}.$$

*Proof.* This proof is analogous to the proof of Lemma 6.4.4. □

Now we are in the position to give the proof of Lemma 6.3.4.

### 6.4.3 Proof of Lemma 6.3.4

*Proof.* First we differentiate $\mathcal{R}F_j(t,\eta)$ with respect to $t$ and obtain from (6.21)

$$\partial_1(\mathcal{R}F_j)(t,\eta) = a'(t)G_j(t, a(t)) + \int_{-\infty}^{a(t)} \partial_1 G_j(t,u)\, du =: T(t),$$

where on the right-hand side the dependence on $\eta$ is omitted in the notation. In the remainder of the proof, we will also suppress the index $j$ as far as possible. Applying $\Delta_h$ then yields for $t \in \mathbb{R}$

$$\begin{aligned}\Delta_h T(t) &= \Delta_h a'(t)G(t+h, a(t+h)) + a'(t)\Delta_h G(t, a(t)) \\ &\quad + \int_{a(t)}^{a(t+h)} \partial_1 G(t+h, u)\, du + \int_{-\infty}^{a(t)} \Delta_{(h,0)} \partial_1 G(t,u)\, du \\ &=: T_1(t) + T_2(t) + T_3(t) + T_4(t).\end{aligned}$$

Next, we estimate the $L_\infty$-norms of the functions $T_i$ for $i \in \{1,2,3,4\}$. Let us begin with $T_1$. Applying Lemma 6.3.2 we obtain

$$\|T_1\|_\infty \leq \|\Delta_h a'\|_\infty \|G\|_\infty \lesssim \|\Delta_h a'\|_\infty \lesssim \delta_j h^{\beta-1}|\sin\eta|^{-1-\beta} \lesssim h^\beta |\sin\eta|^{-1-\beta}.$$

The estimate of $T_2$ takes some more effort. The product rule yields for $t \in \mathbb{R}$

$$\begin{aligned}T_2(t) = a'(t)\Delta_h G(t, a(t)) &= a'(t)\Delta_h g_j(t, a(t))\omega(t+h, a(t+h)) \\ &\quad + a'(t) g_j(t, a(t))\Delta_h \omega(t, a(t)) =: T_{21}(t) + T_{22}(t).\end{aligned}$$

Using the mean value theorem and Lemmas 6.3.2 and 6.4.4 yields

$$\|T_{21}\|_\infty \leq \|a'\|_\infty \sup_{t\in\mathbb{R}} |\Delta_h g_j(t, a(t))| \|\omega\|_\infty \lesssim h|\sin\eta|^{-2} 2^{-\alpha j} \lesssim h^\beta |\sin\eta|^{-1-\beta}.$$

We take another forward difference of the component $T_{22}$ and obtain

$$\begin{aligned}\Delta_h T_{22}(t) &= \Delta_h a'(t) g_j(t+h, a(t+h))\Delta_h \omega(t+h, a(t+h)) \\ &\quad + a'(t)\Delta_h g_j(t, a(t))\Delta_h \omega(t+h, a(t+h)) + a'(t) g(t, a(t))\Delta_h^2 \omega(t, a(t)) \\ &=: T_{22}^1(t) + T_{22}^2(t) + T_{22}^3(t).\end{aligned}$$

These terms allow the following estimates, where we use Lemmas 6.3.2, 6.4.6 and 6.4.4. Also note $h \lesssim |\sin\eta|$.

$$\begin{aligned}
\|T_{22}^1\|_\infty &\le h^{\beta+1}|\sin\eta|^{-2-\beta} \lesssim h^\beta|\sin\eta|^{-1-\beta},\\
\|T_{22}^2\|_\infty &\le h^2|\sin\eta|^{-3}2^{-\alpha j} \lesssim h^\beta|\sin\eta|^{-1-\beta},\\
\|T_{22}^3\|_\infty &\le h^2|\sin\eta|^{-3} + h^{\beta+1}|\sin\eta|^{-2-\beta} \lesssim h^\beta|\sin\eta|^{-1-\beta}.
\end{aligned}$$

By substitution, the term $T_3$ transforms to

$$\begin{aligned}
T_3(t) &= \int_t^{t+h} \partial_1 G(t+h, a(u))a'(u)\,du\\
&= \int_t^{t+h} \partial_1 g_j(t+h, a(u))\omega(t+h, a(u))a'(u)\,du\\
&\quad + \int_t^{t+h} g_j(t+h, a(u))\partial_1\omega(t+h, a(u))a'(u)\,du\\
&=: T_{31}(t) + T_{32}(t).
\end{aligned}$$

We apply $\Delta_h$ to $T_{31}$. Here $\Delta_h$ acts exclusively on $t$ and $\tau$. We obtain

$$\begin{aligned}
\Delta_h T_{31}(t) &= \int_t^{t+h} \Delta_h\big[\partial_1 g_j(t+h, a(\tau))\omega(t+h, a(\tau))a'(\tau)\big]\,d\tau\\
&= \int_t^{t+h} \Delta_h\partial_1 g_j(t+h, a(\tau))\omega(t+2h, a(\tau+h))a'(\tau+h)\,d\tau\\
&\quad + \int_t^{t+h} \partial_1 g_j(t+h, a(\tau))\Delta_h\omega(t+h, a(\tau))a'(\tau+h)\,d\tau\\
&\quad + \int_t^{t+h} \partial_1 g_j(t+h, a(\tau))\omega(t+h, a(\tau))\Delta_h a'(\tau)\,d\tau\\
&=: T_{31}^1(t) + T_{31}^2(t) + T_{31}^3(t).
\end{aligned}$$

Analogously, we decompose

$$\Delta_h T_{32}(t) = \int_t^{t+h} \Delta_h\big[g_j(t+h, a(\tau))\partial_1\omega(t+h, a(\tau))a'(\tau)\big]\,d\tau =: T_{32}^1(t) + T_{32}^2(t) + T_{32}^3(t).$$

Then we estimate with the results from the appendix

$$\begin{aligned}
\|T_{31}^1\|_\infty &\lesssim h|\sin\eta|^{-1}2^{-j}\big(h^{\beta-1} + h^{\beta-1}|\sin\eta|^{1-\beta}\big) \lesssim h^\beta|\sin\eta|^{-1-\beta},\\
\|T_{31}^2\|_\infty &\lesssim h|\sin\eta|^{-1}2^{-\alpha j}\big(h + h|\sin\eta|^{-1}\big) \lesssim h^\beta|\sin\eta|^{-1-\beta},\\
\|T_{31}^3\|_\infty &\lesssim h2^{-\alpha j}\|\Delta_h a'\|_\infty \lesssim h^\beta|\sin\eta|^{-1-\beta},
\end{aligned}$$

and

$$\begin{aligned}
\|T_{32}^1\|_\infty &\lesssim h|\sin\eta|^{-1}2^{-\alpha j}\big(h + h|\sin\eta|^{-1}\big) \lesssim h^\beta|\sin\eta|^{-1-\beta},\\
\|T_{32}^2\|_\infty &\lesssim h|\sin\eta|^{-1}\big(h + h|\sin\eta|^{-1}\big) \lesssim h^\beta|\sin\eta|^{-1-\beta},\\
\|T_{32}^3\|_\infty &\lesssim h\|\Delta_h a'\|_\infty \lesssim h^\beta|\sin\eta|^{-1-\beta}.
\end{aligned}$$

Finally, we treat the term $T_4$,

$$\begin{aligned}
T_4(t) &= \int_{-\infty}^{a(t)} \Delta_h \partial_1 G(t,u)\, du = \int_{-\infty}^{a(t)} \Delta_h \big(\partial_1 g_j(t,u)\omega(t,u) + g_j(t,u)\partial_1\omega(t,u)\big)\, du \\
&= \int_{-\infty}^{a(t)} \Delta_h \partial_1 g_j(t,u)\omega(t+h,u)\, du + \int_{-\infty}^{a(t)} \big(\partial_1 g_j(t,u)\Delta_h\omega(t,u) \\
&\quad + \Delta_h g_j(t,u)\partial_1\omega(t+h,u)\big)\, du \\
&\quad + \int_{-\infty}^{a(t)} g_j(t,u)\Delta_h\partial_1\omega(t,u)\, du =: T_{41}(t) + T_{42}(t) + T_{43}(t).
\end{aligned}$$

The terms $T_{41}$ and $T_{42}$ can be estimated directly,

$$\begin{aligned}
&\|T_{41}\|_\infty \lesssim h^{\beta-1}\cdot 2^{-j} \le h^\beta, \\
&\|T_{42}\|_\infty \lesssim h \cdot 2^{-\alpha j} \asymp 2^{-j} \le 2^{-j(\beta-1)} \asymp h^\beta.
\end{aligned}$$

The term $T_{43}$ again needs some further preparation,

$$\begin{aligned}
\Delta_h T_{43}(t) = &\int_{a(t)}^{a(t+h)} g_j(t+h,u)\Delta_h\partial_1\omega(t+h,u)\, du \\
&+ \int_{-\infty}^{a(t)} \Delta_h\big(g_j(t,u)\Delta_h\partial_1\omega(t,u)\big)\, du =: T_{43}^1(t) + T_{43}^2(t).
\end{aligned}$$

In the end we arrive at

$$\begin{aligned}
&\|T_{43}^1\|_\infty \lesssim h^2|\sin\eta|^{-1} \lesssim h^\beta|\sin\eta|^{-1}, \\
&\|T_{43}^2\|_\infty \lesssim h^2 \lesssim h^\beta.
\end{aligned}$$

Now we collect the appropriate terms and add them up to obtain $S_1$ and $S_2$. In a last step, we use our $L_\infty$-estimates to obtain the desired $L_2$-estimates. Here we use that $|\operatorname{supp} T_i| \lesssim |I(\eta)| \lesssim |\sin\eta|$ according to Lemma 6.3.3 for $i \in \{1,2,3\}$ and $|\operatorname{supp} T_4| \lesssim 1$. This finishes the proof. $\square$

## 6.5 Appendix B: *Refinement of Theorem 6.3.6*

In this final section ([60, Appendix B]) we prove Theorem 6.3.7, which is a refinement of Theorem 6.3.6. For that we need to analyze the *modified edge fragment* $\widetilde{F}_j$, given for fixed $m \in \mathbb{N}_0$ by

$$\widetilde{F}_j(x) = r(x)^m F_j(x), \quad x \in \mathbb{R}^2, \tag{6.35}$$

where $F_j$ is the function (6.14) and $r : \mathbb{R}^2 \to \mathbb{R}$ shall map a vector $x = (x_1, x_2) \in \mathbb{R}^2$ to its first component $x_1 \in \mathbb{R}$. Alternatively, (6.35) can be written as the product $\widetilde{F}_j(x) = \widetilde{G}_j(x)\mathcal{X}_{\{x_1 \ge E_j(x_2)\}}$ with the function

$$\widetilde{G}_j(x) := r(x)^m G_j(x) = r(x)^m \omega(x) g_j(x), \quad x \in \mathbb{R}^2, \tag{6.36}$$

which is a modified version of $G_j(x) = g_j(x)\omega(x)$ from (6.20).

Rotating by the angle $\eta$ yields $\widetilde{G}^\eta_j(x) = (r^\eta(x))^m G^\eta_j(x) = (r^\eta(x))^m g^\eta_j(x)\omega^\eta(x)$, where $G^\eta_j$ and $r^\eta$ are the functions obtained by rotating $G_j$ and $r$, respectively. The function $r^\eta : \mathbb{R}^2 \to \mathbb{R}$ has the form

$$r^\eta(t,a) := t\cos\eta - a\sin\eta, \quad (t,a) \in \mathbb{R}^2. \tag{6.37}$$

Some important properties of $r^\eta$ and $\widetilde{G}^\eta_j$ are collected below.

### 6.5.1 Estimates for $r^\eta$

First we analyze the function $r^\eta : \mathbb{R}^2 \to \mathbb{R}$ given by (6.37). Clearly $r^\eta \in C^\infty(\mathbb{R}^2)$. Also note that $r^\eta$ is not compactly supported. Since $r^\eta$ only occurs as a factor in products with the window $\omega^\eta$ this does not cause any problems however.

Thanks to the smoothness of $r^\eta$ we have the following result.

**Lemma 6.5.1** ([60, Lem. 6.16])**.** *Let $k,m \in \mathbb{N}_0$ and $K \subset \mathbb{R}^2$ a compact set. Then we have*

$$\|\Delta^k_{(h,0)}(r^\eta)^m\|_{L_\infty(K)} \lesssim h^k.$$

Along the edge curve the following estimates hold. Here $\widetilde{I}(\eta)$ denotes the interval defined in (6.19).

**Lemma 6.5.2** ([60, Lem. 6.17])**.** *Let $|\sin\eta| \geq 2\delta_j$. Then we have $\sup_{t\in\widetilde{I}(\eta)} |r^\eta(t,a(t))| \lesssim \delta_j$. Moreover, for $h \geq 0$ it holds*

$$\sup_{t\in\mathbb{R}} |\Delta_h r^\eta(t,a(t))| \lesssim h \quad \textit{and} \quad \sup_{t\in\mathbb{R}} |\Delta^2_h r^\eta(t,a(t))| \lesssim h^\beta \delta_j |\sin\eta|^{-\beta}.$$

*Proof.* For every $t \in \mathbb{R}$ the point $(t,a(t)) \in \mathbb{R}^2$ in rotated coordinates lies on the (extended) edge curve $\Gamma$. We know that the function $E_j$ deviates little from zero and obeys $\sup_{|x_2|\leq 1} |E_j(x_2)| \leq \delta_j \lesssim 2^{-j(1-\alpha)}$ according to (6.15). Furthermore, the slope of $E_j$ outside of $[-1,1]$ is constant and bounded by $\delta_j$. This yields the estimate $\sup_{t\in\widetilde{I}(\eta)} |r^\eta(t,a(t))| \lesssim \delta_j$.

The other estimates follow from Lemma 6.3.2. In view of this lemma we conclude

$$\sup_{t\in\mathbb{R}} |\Delta_h r^\eta(t,a(t))| \leq h \cdot \sup_{t\in\mathbb{R}} |\cos\eta - a'(t)\sin\eta| \lesssim h|\sin\eta|^{-1}|\sin\eta| = h,$$

and

$$\sup_{t\in\mathbb{R}} |\Delta^2_h r^\eta(t,a(t))| \leq h|\sin\eta|\|\Delta_h a'\|_\infty \lesssim h^\beta \delta_j |\sin\eta|^{-\beta}.$$

□

### 6.5.2 Estimates for $\widetilde{G}^\eta_j$

The function $\widetilde{G}^\eta_j$ is the rotated version of the function $\widetilde{G}_j$ given in (6.36) as the composition of the 'elementary functions' $g_j$, $\omega$, and $r$. Hence we can apply the previous estimates to obtain estimates for $\widetilde{G}^\eta_j$.

**Lemma 6.5.3** ([60, Lem. 6.18]). *Let $|\sin\eta| \geq 2\delta_j$. Let $\widetilde{G}_j^\eta(t,a) = (r^\eta(t,a))^m G_j^\eta(t,a)$ for $(t,a) \in \mathbb{R}^2$, $m \in \mathbb{N}$, $m \neq 0$. Then there are the estimates*

$$\sup_{t\in\mathbb{R}} |\widetilde{G}_j^\eta(t,a(t))| \lesssim \delta_j^m, \qquad \sup_{t\in\mathbb{R}} |\Delta_h \widetilde{G}_j^\eta(t,a(t))| \lesssim \delta_j^{m-1} h,$$
$$\sup_{t\in\mathbb{R}} |\partial_1 \widetilde{G}_j^\eta(t,a(t))| \lesssim \delta_j^{m-1}, \qquad \sup_{t\in\mathbb{R}} |\partial_2 \widetilde{G}_j^\eta(t,a(t))| \lesssim \delta_j^{m-1} |\sin\eta|.$$

*Proof.* We omit the dependence on $j$ and $\eta$ and calculate for $(t,a) \in \mathbb{R}^2$

$$\partial_1 \widetilde{G}(t,a) = \partial_1\big(r(t,a)^m G(t,a)\big) = (\cos\eta) m r(t,a)^{m-1} G(t,a) + r(t,a)^m \partial_1 G(t,a),$$

and $$\partial_2 \widetilde{G}(t,a) = \partial_2\big(r(t,a)^m G(t,a)\big) = -(\sin\eta) m r(t,a)^{m-1} G(t,a) + r(t,a)^m \partial_2 G(t,a).$$

The assertion is then a consequence of the following facts. It holds $\|G\|_\infty \lesssim 1$ and $|r(t,a(t))| \leq \delta_j$ for all $t \in I(\eta)$. Further, for $t \notin I(\eta)$ the expressions $G(t,a(t))$, $\partial_1 G(t,a(t))$, and $\partial_2 G(t,a(t))$ vanish. □

### 6.5.3 Refinement of Lemma 6.3.4

In this subsection we prove the following generalization of Lemma 6.3.4.

**Lemma 6.5.4** ([60, Lem. 6.19]). *For $m \in \mathbb{N}_0$ let $\widetilde{F}_j$ be the modified edge fragment* (6.35). *Further, assume that $|\sin\eta| \geq 2\delta_j$ and $h \asymp 2^{-(1-\alpha)j}$. Then the function $S := \Delta_h \partial_1 \mathcal{R}\widetilde{F}_j(\cdot,\eta)$ admits a decomposition*

*such that*

$$\begin{aligned} S &= S_1^0 + S_2^0, \\ \Delta_h S_2^0 &= S_1^1 + S_2^1, \\ \Delta_h S_2^1 &= S_1^2 + S_2^2, \\ &\;\vdots \\ \Delta_h S_2^{m-1} &= S_1^m + S_2^m, \\ \Delta_h S_2^m &= S_1^{m+1}, \end{aligned}$$

*with the estimates*

$$\|S_1^k\|_2^2 \lesssim 2^{-2jm(1-\alpha)} h^{2\beta} |\sin\eta|^{-1-2\beta} + 2^{-j(1-\alpha)(2\beta+1)}, \qquad k = 0,1,\ldots,m+1.$$

*For convenience we set $S_2^{m+1} = 0$.*

We introduce the following language and say, that the function $S$ *admits a decomposition* $(S_1^k, S_2^k)_k$ *of the form* $(*)$ *of length* $m+1$ *with the estimates*

$$\|S_1^k\|_2^2 \lesssim 2^{-2jm(1-\alpha)} h^{2\beta} |\sin\eta|^{-1-2\beta} + 2^{-j(1-\alpha)(2\beta+1)}, \qquad k = 0,1,\ldots,m+1.$$

Before we come to the proof of Lemma 6.5.4 we need to establish three important technical results.

**Lemma 6.5.5** ([60, Lem. 6.20]). *Let $\widetilde{G}^\eta_j(x) = (r^\eta(x))^m G^\eta_j(x)$ for $x \in \mathbb{R}^2$ and $m \in \mathbb{N}_0$. Further, let $h \asymp 2^{-j(1-\alpha)}$. The function $T : \mathbb{R} \to \mathbb{R}$ defined by $T(t) = a'(t)\Delta_h \widetilde{G}^\eta_j(t, a(t))$ then admits a decomposition $(T_1^k, T_2^k)_k$ of the form $(*)$ of length $(m+1)$ with the estimates*

$$\|T_1^k\|_\infty \lesssim h^m h^\beta |\sin\eta|^{-1-\beta}, \quad k = 0, \dots, m+1,$$
$$\|T_2^k\|_\infty \lesssim h^m |\sin\eta|^{-1}, \quad k = 0, \dots, m,$$

*and subject to the condition* $\operatorname{supp} T_i^k \subset \widetilde{I}(\eta)$, *where $\widetilde{I}(\eta)$ is the interval from* (6.19).

*Proof.* We prove this by induction on $m$. If $m = 0$ we put $T_1^0 = T_{21}$, $T_2^0 = T_{22}$, $T_1^1 = \Delta_h T_{22}$, and $T_2^1 = 0$, with entities $T_{21}$ and $T_{22}$ as defined in the proof of Lemma 6.3.4. The estimates for $T_{21}$ and $\Delta_h T_{22}$ have been carried out there. In view of $h \lesssim \sin\eta$ we can further estimate

$$\|T_2^0\|_\infty = \|T_{22}\|_\infty \lesssim h|\sin\eta|^{-2} \lesssim h^0 |\sin\eta|^{-1}.$$

This proves the case $m = 0$.

We proceed with the induction and assume that the lemma is true for $T$, where $m \in \mathbb{N}_0$ is fixed but arbitrary. The associated decomposition of length $m+1$ shall be denoted by $(T_1^k, T_2^k)_k$. We will show that under this hypothesis also the function $\widetilde{T}(t) := a'(t)\Delta_h \widetilde{G}_+(t, a(t))$, where $\widetilde{G}_+(x) = (r^\eta(x))^{m+1} G^\eta_j(x)$ for $x \in \mathbb{R}^2$, admits a decomposition $(\tilde{T}_1^k, \tilde{T}_2^k)_k$ of the form $(*)$ of length $(m+2)$ with the desired properties.

Subsequently, we simplify the notation by omitting the indices $\eta$ and $j$. First we decompose as follows,

$$\begin{aligned}
\widetilde{T}(t) &= a'(t)\Delta_h\big(r(t,a(t))\widetilde{G}(t,a(t))\big) \\
&= a'(t)\Delta_h r(t,a(t))\widetilde{G}(t+h, a(t+h)) + r(t,a(t))a'(t)\Delta_h \widetilde{G}(t,a(t)) \\
&= \big[r(t,a(t))T_1^0(t)\big] + \big[a'(t)\Delta_h r(t,a(t))\widetilde{G}(t+h,a(t+h)) + r(t,a(t))T_2^0(t)\big] \\
&=: \tilde{T}_1^0(t) + \tilde{T}_2^0(t).
\end{aligned}$$

In view of the properties of $T_1^0$ and Lemma 6.5.2 we see that the function $\tilde{T}_1^0$ satisfies the assertion. The estimate

$$\|\tilde{T}_2^0\|_\infty \lesssim \|a'\|_\infty \sup_{t\in\mathbb{R}} |\Delta_h r(t,a(t))| \sup_{t\in\mathbb{R}} |\widetilde{G}(t,a(t))| \lesssim |\sin\eta|^{-1} \cdot h \cdot \delta_j^m,$$

where Lemmas 6.3.2, 6.5.2 and 6.5.3 were used, shows the claim also for $\tilde{T}_2^0$.

We take another forward difference of the component $\tilde{T}_2^0$ and obtain

$$\begin{aligned}
\Delta_h \tilde{T}_2^0(t) &= \Delta_h a'(t)\Delta_h r(t+h,a(t+h))\widetilde{G}(t+2h,a(t+2h)) \\
&\quad +a'(t)\Delta_h^2 r(t,a(t))\widetilde{G}(t+2h,a(t+2h)) \\
&\quad +\Delta_h r(t,a(t))a'(t)\Delta_h\widetilde{G}(t+h,a(t+h)) + \Delta_h r(t,a(t))T_2^0(t+h) \\
&\quad +r(t,a(t))\Delta_h T_2^0(t) \\
&= \Delta_h a'(t)\Delta_h r(t+h,a(t+h))\widetilde{G}(t+2h,a(t+2h)) \\
&\quad +a'(t)\Delta_h^2 r(t,a(t))\widetilde{G}(t+2h,a(t+2h)) \\
&\quad +\Delta_h r(t,a(t))(T_1^0(t+h)+T_2^0(t+h)) \\
&\quad -\Delta_h r(t,a(t))\Delta_h a'(t)\Delta_h\widetilde{G}(t+h,a(t+h)) \\
&\quad +\Delta_h r(t,a(t))T_2^0(t+h) + r(t,a(t))T_1^1(t) + r(t,a(t))T_2^1(t) \\
&= \big[\Delta_h a'(t)\Delta_h r(t+h,a(t+h))\widetilde{G}(t+2h,a(t+2h)) \\
&\quad +a'(t)\Delta_h^2 r(t,a(t))\widetilde{G}(t+2h,a(t+2h)) \\
&\quad -\Delta_h r(t,a(t))\Delta_h a'(t)\Delta_h\widetilde{G}(t+h,a(t+h)) + r(t,a(t))T_1^1(t) \\
&\quad +\Delta_h r(t,a(t))T_1^0(t+h)\big] \\
&\quad +\big[2\Delta_h r(t,a(t))T_2^0(t+h) + r(t,a(t))T_2^1(t)\big] \\
&=: \tilde{T}_1^1(t) + \tilde{T}_2^1(t).
\end{aligned}$$

For $\tilde{T}_1^1$ we check directly

$$\begin{aligned}
\sup_{t\in\mathbb{R}} |\Delta_h a'(t)\Delta_h r(t,a(t))\widetilde{G}(t+h,a(t+h))| &\lesssim h^{\beta-1}\delta_j|\sin\eta|^{-1-\beta}hh^m \\
&= h^{m+1}h^\beta|\sin\eta|^{-1-\beta}, \\
\sup_{t\in\mathbb{R}} |a'(t)\Delta_h^2 r(t,a(t))\widetilde{G}(t+2h,a(t+2h))| &\lesssim |\sin\eta|^{-1}h^\beta\delta_j|\sin\eta|^{-\beta}h^m \\
&= h^{m+1}h^\beta|\sin\eta|^{-1-\beta}, \\
\sup_{t\in\mathbb{R}} |\Delta_h r(t,a(t))\Delta_h a'(t)\Delta_h\widetilde{G}(t+h,a(t+h))| &\lesssim \delta_j^m h^\beta h|\sin\eta|^{-1-\beta} \\
&\lesssim h^{m+1}h^\beta|\sin\eta|^{-1-\beta}.
\end{aligned}$$

The estimates for the remaining two terms are obvious. Hence $\tilde{T}_1^1$ fulfills the desired properties.

For $\tilde{T}_2^1$ we use the induction hypothesis and Lemma 6.5.2 to obtain

$$\|\tilde{T}_2^1\|_\infty \lesssim \sup_{t\in\mathbb{R}}|r(t,a(t))T_2^1(t)| + \sup_{t\in\mathbb{R}}|\Delta_h r(t,a(t))T_2^0(t+h)| \lesssim h^{m+1}|\sin\eta|^{-1}.$$

Moving forward, this procedure yields terms for $k=1,\dots,m+1$,

$$\begin{aligned}
\tilde{T}_1^{k+1}(t) &= r(t,a(t))T_1^{k+1}(t) + (k+1)\Delta_h r(t+h,a(t+h))T_1^k(t+h) \\
&\quad +(k+1)\Delta_h^2 r(t,a(t))T_2^{k-1}(t+h) + (k+1)\Delta_h^2 r(t,a(t))T_2^k(t+h), \\
\tilde{T}_2^k(t) &= r(t,a(t))T_2^k(t) + (k+1)\Delta_h r(t,a(t))T_2^{k-1}(t+h),
\end{aligned}$$

which satisfy the desired estimates. Here we put $T_1^{m+2} = T_2^{m+2} = 0$ for convenience.

Indeed, using the induction assumptions, we obtain

$$\begin{aligned}
\|\tilde{T}_1^{k+1}\|_\infty &\lesssim \sup_{t\in\mathbb{R}} |r(t,a(t))T_1^{k+1}(t)| + \sup_{t\in\mathbb{R}} |\Delta_h r(t+h,a(t+h))T_1^k(t+h)| \\
&\quad + \sup_{t\in\mathbb{R}} |\Delta_h^2 r(t,a(t))T_2^{k-1}(t+h)| + \sup_{t\in\mathbb{R}} |\Delta_h^2 r(t,a(t))T_2^k(t+h)| \\
&\lesssim h^{m+1}h^\beta |\sin\eta|^{-1-\beta}, \\
\|\tilde{T}_2^k\|_\infty &\lesssim \sup_{t\in\mathbb{R}} |r(t,a(t))T_2^k(t)| + \sup_{t\in\mathbb{R}} |\Delta_h r(t,a(t))T_2^{k-1}(t+h)| \lesssim h^{m+1}|\sin\eta|^{-1}.
\end{aligned}$$

Note, that $T_2^{m+1} = T_1^{m+2} = T_2^{m+2} = 0$. Hence, for $k = m+1$ these expressions read

$$\begin{aligned}
\tilde{T}_1^{m+2}(t) &= (m+2)\Delta_h r(t+h,a(t+h))T_1^{m+1}(t+h) + (m+2)\Delta_h^2 r(t,a(t))T_2^m(t+h), \\
\tilde{T}_2^{m+1}(t) &= (m+2)\Delta_h r(t,a(t))T_2^m(t+h).
\end{aligned}$$

Since $\Delta_h \tilde{T}_2^{m+1} = \tilde{T}_1^{m+2}$ we have $\tilde{T}_2^{m+2} = 0$ and the proof is finished. □

The following Lemma 6.5.6 is in the same spirit as Lemma 6.5.5.

**Lemma 6.5.6** ([60, Lem. 6.21]). *Let $\widetilde{G}_j^\eta(x) = (r^\eta(x))^m G_j^\eta(x)$ for $x \in \mathbb{R}^2$, $m \in \mathbb{N}_0$, and $h \asymp 2^{-j(1-\alpha)}$. Then the function $S : \mathbb{R} \to \mathbb{R}$ defined by $S(t) = a'(t)\Delta_h \partial_1 \widetilde{G}_j^\eta(t,a(t))$ admits a decomposition $(S_1^k, S_2^k)_k$ of the form $(*)$ of length $m+1$ with estimates*

$$\begin{aligned}
&\|S_1^k\|_\infty \lesssim h^{m-1}h^\beta |\sin\eta|^{-1-\beta}, \quad k = 0,\dots,m+1, \\
&\|S_2^k\|_\infty \lesssim h^{m-1}|\sin\eta|^{-1}, \quad k = 0,\dots,m.
\end{aligned}$$

*Moreover, these functions can be chosen such that* $\operatorname{supp} S_i^k \subset \widetilde{I}(\eta)$ *with $\widetilde{I}(\eta)$ from* (6.19).

*Proof.* The proof is by induction on $m$. To enhance readability we again omit the indices $\eta$ and $j$. The assertions are clearly true for $m = 0$.

For the induction we let $m \in \mathbb{N}_0$ be fixed and let $S$ be the function defined in the setting. Further, let us assume that we have a decomposition $(S_1^k, S_2^k)_k$ of length $m+1$ with the desired properties for $S$. We put $S_2^{m+1} = 0$ and for convenience we also define $S_1^{m+2} = S_2^{m+2} = 0$. We will show that under these assumptions the function $\widetilde{S} : \mathbb{R} \to \mathbb{R}$ given by $\widetilde{S}(t) := a'(t)\Delta_h \partial_1 \widetilde{G}_+(t,a(t))$, where $\widetilde{G}_+(x) = r(x)^{m+1}G(x)$ for $x \in \mathbb{R}^2$, admits a decomposition $(\tilde{S}_1^k, \tilde{S}_2^k)_k$ of length $m+2$ of the same form. First we calculate

$$\begin{aligned}
\widetilde{S}(t) &= a'(t)\Delta_h \partial_1 \widetilde{G}_+(t,a(t)) = a'(t)\Delta_h \big(\cos\eta \widetilde{G}(t,a(t)) + r(t,a(t))\partial_1 \widetilde{G}(t,a(t))\big) \\
&= a'(t)\cos\eta \Delta_h \widetilde{G}(t,a(t)) + a'(t)\Delta_h r(t,a(t))\partial_1 \widetilde{G}(t,a(t)) \\
&\quad + r(t+h,a(t+h))a'(t)\Delta_h \partial_1 \widetilde{G}(t,a(t)).
\end{aligned}$$

Using the induction hypothesis we can proceed,

$$\begin{aligned}
\widetilde{S}(t) &= \big[r(t+h,a(t+h))S_1^0(t)\big] \\
&\quad + \big[r(t+h,a(t+h))S_2^0(t) + a'(t)\cos\eta \Delta_h \widetilde{G}(t,a(t)) \\
&\quad + a'(t)\Delta_h r(t,a(t))\partial_1 \widetilde{G}(t,a(t))\big] \\
&=: \tilde{S}_1^0(t) + \tilde{S}_2^0(t).
\end{aligned}$$

The terms $\tilde{S}_1^0$ and $\tilde{S}_2^0$ have the desired properties, which follows from the estimates

$$\begin{aligned}
\sup_{t\in\mathbb{R}} |r(t+h,a(t+h))S_1^0(t)| &\lesssim hh^{m-1}h^{\beta}|\sin\eta|^{-1-\beta},\\
\sup_{t\in\mathbb{R}} |r(t+h,a(t+h))S_2^0(t)| &\lesssim hh^{m-1}|\sin\eta|^{-1},\\
\sup_{t\in\mathbb{R}} |a'(t)\cos\eta\Delta_h\widetilde{G}(t,a(t))| &\lesssim |\sin\eta|^{-1}\delta_j^{m-1}h,\\
\sup_{t\in\mathbb{R}} |a'(t)\Delta_h r(t,a(t))\partial_1\widetilde{G}(t,a(t))| &\lesssim |\sin\eta|^{-1}h\delta_j^{m-1}.
\end{aligned}$$

Taking another forward difference of $\tilde{S}_2^0$ yields

$$\begin{aligned}
\Delta_h\tilde{S}_2^0(t) \quad = \quad & \Delta_h r(t+h,a(t+h))S_2^0(t) + r(t+2h,a(t+2h))\Delta_h S_2^0(t)\\
&+\Delta_h a'(t)\cos\eta\Delta_h\widetilde{G}(t,a(t))\\
&+a'(t+h)\cos\eta\Delta_h^2\widetilde{G}(t,a(t)) + \Delta_h a'(t)\Delta_h r(t,a(t))\partial_1\widetilde{G}(t,a(t))\\
&+a'(t+h)\Delta_h^2 r(t,a(t))\partial_1\widetilde{G}(t,a(t))\\
&+a'(t+h)\Delta_h r(t+h,a(t+h))\Delta_h\partial_1\widetilde{G}(t,a(t)).
\end{aligned}$$

Let $T$ denote the function from Lemma 6.5.5. We observe,

$$\begin{aligned}
a'(t+h)\Delta_h^2\widetilde{G}(t,a(t)) &= a'(t+h)\big(\Delta_h\widetilde{G}(t+h,a(t+h)) - \Delta_h\widetilde{G}(t,a(t))\big)\\
&= a'(t+h)\Delta_h\widetilde{G}(t+h,a(t+h)) - a'(t)\Delta_h\widetilde{G}(t,a(t))\\
&\quad + (a'(t)-a'(t+h))\Delta_h\widetilde{G}(t,a(t))\\
&= a'(t+h)\Delta_h\widetilde{G}(t+h,a(t+h)) - a'(t)\Delta_h\widetilde{G}(t,a(t))\\
&\quad - \Delta_h a'(t)\Delta_h\widetilde{G}(t,a(t))\\
&= T(t+h) - T(t) - \Delta_h a'(t)\Delta_h\widetilde{G}(t,a(t))\\
&= \Delta_h T(t) - \Delta_h a'(t)\Delta_h\widetilde{G}(t,a(t)).
\end{aligned}$$

Now we know by Lemma 6.5.5 that there is a decomposition $(T_1^k, T_2^k)_k$ of $T$ of length $m+1$ with the specific properties given there. This allows to decompose $\Delta_h T = \Delta_h T_1^0 + T_1^1 + T_2^1$ and we obtain

$$a'(t+h)\Delta_h^2\widetilde{G}(t,a(t)) = \Delta_h T_1^0(t) + T_1^1(t) + T_2^1(t) - \Delta_h a'(t)\Delta_h\widetilde{G}(t,a(t)).$$

Using this observation we obtain

$$\begin{aligned}
\Delta_h\tilde{S}_2^0(t) \quad = \quad & \Delta_h r(t+h,a(t+h))S_2^0(t) + r(t+2h,a(t+2h))S_1^1(t)\\
&+r(t+2h,a(t+2h))S_2^1(t)\\
&+\Delta_h a'(t)\cos\eta\Delta_h\widetilde{G}(t,a(t)) + \cos\eta\Delta_h T_1^0(t) + \cos\eta(T_1^1(t)+T_2^1(t))\\
&-\cos\eta\Delta_h a'(t)\Delta_h\widetilde{G}(t,a(t))\\
&+\Delta_h a'(t)\Delta_h r(t,a(t))\partial_1\widetilde{G}(t,a(t)) + a'(t+h)\Delta_h^2 r(t,a(t))\partial_1\widetilde{G}(t,a(t))\\
&+a'(t)\Delta_h r(t+h,a(t+h))\Delta_h\partial_1\widetilde{G}(t,a(t))\\
&+(a'(t+h)-a'(t))\Delta_h r(t+h,a(t+h))\Delta_h\partial_1\widetilde{G}(t,a(t))
\end{aligned}$$

and further

$$\begin{aligned}
\Delta_h \tilde{S}_2^0(t) \quad &= \quad \big[r(t+2h, a(t+2h))S_1^1(t) + \Delta_h a'(t)\cos\eta\Delta_h \widetilde{G}(t,a(t)) \\
&\quad + \cos\eta\Delta_h T_1^0(t) - \cos\eta\Delta_h a'(t)\Delta_h \widetilde{G}(t,a(t)) + \cos\eta T_1^1(t) \\
&\quad + \Delta_h a'(t)\Delta_h r(t,a(t))\partial_1 \widetilde{G}(t,a(t)) + a'(t+h)\Delta_h^2 r(t,a(t))\partial_1 \widetilde{G}(t,a(t)) \\
&\quad + \Delta_h a'(t)\Delta_h r(t+h,a(t+h))\Delta_h\partial_1 \widetilde{G}(t,a(t)) + \Delta_h r(t+h,a(t+h))S_1^0(t)\big] \\
&\quad + \big[r(t+2h,a(t+2h))S_2^1(t) + \cos\eta T_2^1(t) + 2\Delta_h r(t+h,a(t+h))S_2^0(t)\big] \\
&=: \quad \tilde{S}_1^1(t) + \tilde{S}_2^1(t).
\end{aligned}$$

Now we can split $\Delta_h \tilde{S}_2^0 = \tilde{S}_1^1 + \tilde{S}_2^1$ with

$$\begin{aligned}
\tilde{S}_1^1(t) \quad &= \quad r(t+2h,a(t+2h))S_1^1(t) + \cos\eta\Delta_h a'(t)\Delta_h \widetilde{G}(t,a(t)) \\
&\quad + \Delta_h a'(t)\Delta_h r(t,a(t))\partial_1 \widetilde{G}(t,a(t)) \\
&\quad + a'(t+h)\Delta_h^2 r(t,a(t))\partial_1 \widetilde{G}(t,a(t)) + \Delta_h r(t+h,a(t+h))S_1^0(t) + \cos\eta\Delta_h T_1^0(t) \\
&\quad + \Delta_h a'(t)\Delta_h r(t+h,a(t+h))\Delta_h\partial_1 \widetilde{G}(t,a(t)) - \cos\eta\Delta_h a'(t)\Delta_h \widetilde{G}(t,a(t)) \\
&\quad + \cos\eta T_1^1(t), \\
\tilde{S}_2^1(t) \quad &= \quad 2\Delta_h r(t+h,a(t+h))S_2^0(t) + r(t+2h,a(t+2h))S_2^1(t) + \cos\eta T_2^1(t).
\end{aligned}$$

These terms have the desired properties. To see this, we calculate

$$\begin{aligned}
\sup_{t\in\mathbb{R}} |r(t+2h,a(t+2h))S_1^1(t)| &\lesssim h h^{m-1} h^\beta |\sin\eta|^{-1-\beta}, \\
\sup_{t\in\mathbb{R}} |\Delta_h a'(t)\Delta_h \widetilde{G}(t,a(t))| &\lesssim \delta_j h^{\beta-1}|\sin\eta|^{-1-\beta}\cdot \delta_j^{m-1} h, \\
\sup_{t\in\mathbb{R}} |\Delta_h a'(t)\Delta_h r(t,a(t))\partial_1 \widetilde{G}(t,a(t))| &\lesssim \delta_j h^{\beta-1}|\sin\eta|^{-1-\beta}\cdot h\cdot \delta_j^{m-1}, \\
\sup_{t\in\mathbb{R}} |a'(t+h)\Delta_h^2 r(t,a(t))\partial_1 \widetilde{G}(t,a(t))| &\lesssim |\sin\eta|^{-1}\cdot h^\beta \delta_j |\sin\eta|^{-\beta}\cdot \delta_j^{m-1}, \\
\sup_{t\in\mathbb{R}} |\Delta_h r(t+h,a(t+h))S_1^0(t)| &\lesssim h h^{m-1} h^\beta |\sin\eta|^{-1-\beta}, \\
\sup_{t\in\mathbb{R}} |\Delta_h T_1^0(t)| &\lesssim h^m h^\beta |\sin\eta|^{-1-\beta}, \\
\sup_{t\in\mathbb{R}} |\Delta_h a'(t)\Delta_h \widetilde{G}(t,a(t))| &\lesssim \delta_j h^{\beta-1}|\sin\eta|^{-1-\beta}\cdot \delta_j^{m-1} h, \\
\sup_{t\in\mathbb{R}} |\Delta_h a'(t)\Delta_h r(t+h,a(t+h))\Delta_h\partial_1 \widetilde{G}(t,a(t))| &\lesssim \delta_j h^{\beta-1}|\sin\eta|^{-1-\beta}\cdot h\cdot \delta_j^{m-1}, \\
\sup_{t\in\mathbb{R}} |T_1^1(t)| &\lesssim \delta_j^m h^\beta |\sin\eta|^{-1-\beta},
\end{aligned}$$

and

$$\begin{aligned}
\sup_{t\in\mathbb{R}} |2\Delta_h r(t+h,a(t+h))S_2^0(t)| &\lesssim h h^{m-1}|\sin\eta|^{-1}, \\
\sup_{t\in\mathbb{R}} |r(t+2h,a(t+2h))S_2^1(t)| &\lesssim h h^{m-1}|\sin\eta|^{-1}, \\
\sup_{t\in\mathbb{R}} |T_2^1(t)| &\lesssim h h^{m-1}|\sin\eta|^{-1}.
\end{aligned}$$

We proceed with

$$\begin{aligned}\Delta_h \tilde{S}_2^1(t) &= \big[\cos\eta T_1^2(t) + r(t+3h, a(t+3h))S_1^2(t) + 2\Delta_h r(t+2h, a(t+2h))S_1^1(t)\\ &\quad +2\Delta_h^2 r(t+h, a(t+h))S_2^0(t)\big] + \big[\cos\eta T_2^2(t) + r(t+3h, a(t+3h))S_2^2(t)\\ &\quad +3\Delta_h r(t+2h, a(t+2h))S_2^1(t)\big] =: \tilde{S}_1^2(t) + \tilde{S}_2^2(t).\end{aligned}$$

Inductively, we put for $k = 1, \dots, m+1$, where for convenience $T_1^{m+2} = 0$,

$$\begin{aligned}\tilde{S}_1^{k+1}(t) &:= \cos\eta T_1^{k+1}(t) + r(t+(k+2)h, a(t+(k+2)h))S_1^{k+1}(t)\\ &\quad +(k+1)\Delta_h r(t+(k+1)h, a(t+(k+1)h))S_1^k(t)\\ &\quad +(k+1)\Delta_h^2 r(t+kh, a(t+kh))S_2^{k-1}(t),\\ \tilde{S}_2^k(t) &:= \cos\eta T_2^k(t) + r(t+(k+1)h, a(t+(k+1)h))S_2^k(t)\\ &\quad +(k+1)\Delta_h r(t+kh, a(t+kh))S_2^{k-1}(t).\end{aligned}$$

These terms clearly satisfy $\Delta_h \tilde{S}_2^k = \tilde{S}_1^{k+1} + \tilde{S}_2^{k+1}$. They also have the desired properties since

$$\begin{aligned}\sup_{t\in\mathbb{R}} |T_1^{k+1}(t)| &\lesssim h^m h^\beta |\sin\eta|^{-1-\beta},\\ \sup_{t\in\mathbb{R}} |r(t+(k+2)h, a(t+(k+2)h))S_1^{k+1}(t)| &\lesssim h\cdot h^{m-1}h^\beta|\sin\eta|^{-1-\beta},\\ \sup_{t\in\mathbb{R}} |\Delta_h r(t+(k+1)h, a(t+(k+1)h))S_1^k(t)| &\lesssim h\cdot h^{m-1}h^\beta|\sin\eta|^{-1-\beta},\\ \sup_{t\in\mathbb{R}} |\Delta_h^2 r(t+kh, a(t+kh))S_2^{k-1}(t)| &\lesssim h^\beta\delta_j|\sin\eta|^{-\beta}\cdot h^{m-1}|\sin\eta|^{-1},\end{aligned}$$

and

$$\begin{aligned}\sup_{t\in\mathbb{R}} |T_2^k(t)| &\lesssim h^m|\sin\eta|^{-1},\\ \sup_{t\in\mathbb{R}} |r(t+(k+1)h, a(t+(k+1)h))S_2^k(t)| &\lesssim h\cdot h^{m-1}|\sin\eta|^{-1},\\ \sup_{t\in\mathbb{R}} |\Delta_h r(t+kh, a(t+kh))S_2^{k-1}(t)| &\lesssim h\cdot h^{m-1}|\sin\eta|^{-1}.\end{aligned}$$

Since $S_2^{m+1} = S_1^{m+2} = S_2^{m+2} = T_2^{m+1} = T_1^{m+2} = 0$, for $k = m+1$ these expressions read

$$\begin{aligned}\tilde{S}_1^{m+2}(t) &= (m+2)\Delta_h r(t+(m+2)h, a(t+(m+2)h))S_1^{m+1}(t)\\ &\quad +(m+2)\Delta_h^2 r(t+(m+1)h, a(t+(m+1)h))S_2^m(t),\\ \tilde{S}_2^{m+1}(t) &= (m+2)\Delta_h r(t+(m+1)h, a(t+(m+1)h))S_2^m(t).\end{aligned}$$

We see that $\Delta_h \tilde{S}_2^{m+1} = \tilde{S}_1^{m+2}$. Therefore $\tilde{S}_2^{m+2} = 0$ and the proof is finished. □

A slight modification of the previous proof leads to the following lemma.

**Lemma 6.5.7** ([60, Lem. 6.22]). *Let $\widetilde{G}_j^\eta(x) = (r^\eta(x))^m G_j^\eta(x)$ for $x \in \mathbb{R}^2$ and $m \in \mathbb{N}_0$ and $h \asymp 2^{-j(1-\alpha)}$. The function $\widetilde{S} : \mathbb{R}^2 \to \mathbb{R}$ given by $\widetilde{S}(t,\tau) = a'(\tau)\Delta_h\partial_1\widetilde{G}_j^\eta(t, a(\tau))$ for $(t,\tau) \in \mathbb{R}^2$ admits a decomposition $(\tilde{S}_1^k, \tilde{S}_2^k)_k$ of the form $(*)$ of length $m+1$ with estimates*

$$\begin{aligned}\sup_{t\in\mathbb{R}}\sup_{\tau\in[t-h,t+h]} |\tilde{S}_1^k(t,\tau)| &\lesssim h^{m-1}h^\beta|\sin\eta|^{-1-\beta}, \quad k = 0, \dots, m+1,\\ \sup_{t\in\mathbb{R}}\sup_{\tau\in[t-h,t+h]} |\tilde{S}_2^k(t,\tau)| &\lesssim h^{m-1}|\sin\eta|^{-1}, \quad k = 0, \dots, m.\end{aligned}$$

*Proof.* A small adaption of the previous proof is required to account for the little deviation of $\tau$ from $t$. We just make the following remark. For $t, \tau \in \mathbb{R}$ we have $r^\eta(t, a(\tau)) = r^\eta(t, a(t)) + (a(t) - a(\tau)) \sin\eta$. It follows for $h \in \mathbb{R}$

$$\sup_{\tau\in[t-h,t+h]} |r^\eta(t, a(\tau))| \leq |r^\eta(t, a(t))| + |h \sin\eta| \|a'\|_\infty \lesssim |r^\eta(t, a(t))| + |h|.$$

Since $h \asymp 2^{-j(1-\alpha)}$ this additional term poses no problem in the estimations. □

Finally, we have all the tools available to give the proof of Lemma 6.5.4.

*Proof of Lemma 6.5.4.* We have $\widetilde{G}^\eta_j(x) = (r^\eta(x))^m G^\eta_j(x)$ for $x \in \mathbb{R}^2$ and analogous to (6.21)

$$\mathcal{R}\widetilde{F}_j(t, \eta) = \int_{-\infty}^{a(t,\eta)} \widetilde{G}^\eta_j(t, u)\, du.$$

For simplicity we omit the superindex $\eta$ subsequently, and also $j$ wherever possible. Similar to the proof of Lemma 6.3.4 we obtain

$$\begin{aligned} S(t) &= \Delta_h a'(t)\widetilde{G}(t+h, a(t+h)) + a'(t)\Delta_h\widetilde{G}(t, a(t)) + \int_{a(t)}^{a(t+h)} \partial_1\widetilde{G}(t+h, u)\, du \\ &\quad + \int_{-\infty}^{a(t)} \Delta_{(h,0)}\partial_1\widetilde{G}(t, u)\, du \\ &=: \widetilde{T}_1(t) + \widetilde{T}_2(t) + \widetilde{T}_3(t) + \widetilde{T}_4(t). \end{aligned}$$

We will show the assertion for each of these terms separately. Moreover, it suffices to prove $L_\infty$-estimates, which can be transformed to the desired $L_2$-estimates via the corresponding support properties. Note that $|\operatorname{supp}\widetilde{T}_i| \lesssim |\widetilde{I}(\eta)| \lesssim |\sin\eta|$ according to Lemma 6.3.3 for $i \in \{1, 2, 3\}$ and that $|\operatorname{supp}\widetilde{T}_4| \lesssim 1$.

For $\widetilde{T}_1$ the estimate

$$\begin{aligned} \|\widetilde{T}_1\|_\infty &\leq \|\Delta_h a'\|_\infty \sup_{t\in\mathbb{R}} |\widetilde{G}(t, a(t))| \lesssim \|\Delta_h a'\|_\infty \sup_{t\in I(\eta)} |r(t, a(t))|^m \\ &\lesssim \delta_j^{m+1} h^{\beta-1} |\sin\eta|^{-1-\beta} \lesssim h^m h^\beta |\sin\eta|^{-1-\beta} \end{aligned}$$

is sufficient. Next, we show that $\widetilde{T}_2$ and $\widetilde{T}_3$ admit decompositions $(\widetilde{T}_1^k, \widetilde{T}_2^k)_k$ of the form $(*)$ of length $m + 1$ with $\operatorname{supp}\widetilde{T}_i^k \subset \widetilde{I}(\eta)$, $i \in \{1, 2\}$, and the estimates

$$\begin{aligned} &\|\widetilde{T}_1^k\|_\infty \lesssim h^m h^\beta |\sin\eta|^{-1-\beta}, \quad k = 0, \dots, m+1, \\ &\|\widetilde{T}_2^k\|_\infty \lesssim h^m |\sin\eta|^{-1}, \quad k = 0, \dots, m. \end{aligned}$$

The decomposition of the component $\widetilde{T}_2$ is provided by Lemma 6.5.5. Let us turn to $\widetilde{T}_3$. By substitution this term transforms to

$$\widetilde{T}_3(t) = \int_t^{t+h} \partial_1\widetilde{G}(t+h, a(u))a'(u)\, du.$$

We put $\widetilde{T}_1^0 = 0$ and $\widetilde{T}_2^0 = \widetilde{T}_3$. These terms clearly satisfy the assertions. Next we take the forward difference of $\widetilde{T}_2^0$. Here $\Delta_h$ acts on both $t$ and $\tau$. We obtain

$$\begin{aligned}
\Delta_h \widetilde{T}_2^0(t) = \Delta_h \widetilde{T}_3(t) &= \int_t^{t+h} \Delta_h\big(\partial_1 \widetilde{G}(t+h, a(\tau))a'(\tau)\big)\, d\tau \\
&= \int_t^{t+h} \Delta_h \partial_1 \widetilde{G}(t+h, a(\tau))a'(\tau)\, d\tau \\
&\quad + \int_t^{t+h} \partial_1 \widetilde{G}(t+2h, a(\tau+h))\Delta_h a'(\tau)\, d\tau \\
&=: \widetilde{T}_{31}(t) + \widetilde{T}_{32}(t).
\end{aligned}$$

Lemma 6.5.7 then yields a decomposition $(\widetilde{S}_1^k, \widetilde{S}_2^k)_k$, such that we can write

$$\Delta_h \partial_1 \widetilde{G}(t+h, a(\tau))a'(\tau) = \widetilde{S}_1^0(t,\tau) + \widetilde{S}_2^0(t,\tau).$$

This leads to

$$\widetilde{T}_{31}(t) = \int_t^{t+h} \Delta_h \partial_1 \widetilde{G}(t+h, a(\tau))a'(\tau)\, d\tau = \int_t^{t+h} \widetilde{S}_1^0(t,\tau)\, d\tau + \int_t^{t+h} \widetilde{S}_2^0(t,\tau)\, d\tau.$$

We put $\widetilde{T}_1^1(t) := \widetilde{T}_{32}(t) + \int_t^{t+h} \widetilde{S}_1^0(t,\tau)\, d\tau$ and $\widetilde{T}_2^1(t) := \int_t^{t+h} \widetilde{S}_2^0(t,\tau)\, d\tau$. These terms $\widetilde{T}_1^1$ and $\widetilde{T}_2^1$ then satisfy the requirements, i.e.,

$$\begin{aligned}
\|\widetilde{T}_1^1\|_\infty &\lesssim \|\widetilde{T}_{32}\|_\infty + h \cdot \sup_{t\in\mathbb{R}} \sup_{\tau\in[t,t+h]} |\widetilde{S}_1^0(t,\tau)| \lesssim h \cdot \delta_j^{m-1} \cdot h^{\beta-1}\delta_j |\sin\eta|^{-1-\beta}, \\
\|\widetilde{T}_2^1\|_\infty &\lesssim h \sup_{t\in\mathbb{R}} \sup_{\tau\in[t,t+h]} |\widetilde{S}_2^0(t,\tau)| \lesssim h^m |\sin\eta|^{-1}.
\end{aligned}$$

Taking another forward difference of $\widetilde{T}_2^1$ yields

$$\Delta_h \widetilde{T}_2^1(t) = \int_t^{t+h} \Delta_h S_2^0(t,\tau)\, d\tau.$$

Proceeding inductively from here with Lemma 6.5.7 settles the claim for the component $\widetilde{T}_3$.

Finally, we turn to the function $\widetilde{T}_4(t) = \int_{-\infty}^{a(t)} \Delta_{(h,0)} \partial_1 \widetilde{G}(t,u)\, du$. First, we calculate

$$\begin{aligned}
\Delta_h \widetilde{T}_4(t) &= \int_{a(t)}^{a(t+h)} \Delta_{(h,0)} \partial_1 \widetilde{G}(t+h,u)\, du + \int_{-\infty}^{a(t)} \Delta_{(h,0)}^2 \partial_1 \widetilde{G}(t,u)\, du =: \widetilde{T}_{41}(t) + \widetilde{T}_{42}(t), \\
\Delta_h \widetilde{T}_{42}(t) &= \int_{a(t)}^{a(t+h)} \Delta_{(h,0)}^2 \partial_1 \widetilde{G}(t+h,u)\, du + \int_{-\infty}^{a(t)} \Delta_{(h,0)}^3 \partial_1 \widetilde{G}(t,u)\, du =: \widetilde{T}_{43}(t) + \widetilde{T}_{44}(t).
\end{aligned}$$

Next, we show $\|\widetilde{T}_{44}\|_\infty \lesssim h^{\beta+\frac{1}{2}}$ because then, in view of $|\operatorname{supp} \widetilde{T}_{44}| \asymp 1$,

$$\|\widetilde{T}_{44}\|_2^2 \lesssim h^{2\beta+1} \asymp 2^{-j(1-\alpha)(2\beta+1)}.$$

The $L_\infty$-estimate of the term $\widetilde{T}_{44}$ relies on the fact, that for $h \asymp 2^{-j(1-\alpha)}$

$$\|\Delta_{(h,0)}^3 \partial_1 \widetilde{G}\|_\infty \lesssim h^\beta 2^{-\alpha j} \lesssim h^{\beta+1}.$$

This estimate is a consequence of Lemmas 6.4.3, 6.4.5, and 6.5.1 and is analogous to (6.8). Essential is the observation that since $\alpha \geq \frac{1}{2}$ Lemma 6.4.3 yields

$$\|\Delta_{(h,0)}\partial_1 g_j\|_\infty \lesssim 2^{-\alpha j} h^\beta \lesssim h^{\beta+1}.$$

Finally, we take care of the remaining terms $\widetilde{T}_{41}$ and $\widetilde{T}_{43}$. First we note that $|\operatorname{supp} \widetilde{T}_{41}| \lesssim |\widetilde{I}(\eta)| \lesssim |\sin\eta|$ and also $|\operatorname{supp} \widetilde{T}_{43}| \lesssim |\widetilde{I}(\eta)| \lesssim |\sin\eta|$ according to Lemma 6.3.3. Hence, it suffices to prove $\|\widetilde{T}_{41}\|_\infty \lesssim h^m h^\beta |\sin\eta|^{-1-\beta}$ and $\|\widetilde{T}_{41}\|_\infty \lesssim h^m h^\beta |\sin\eta|^{-1-\beta}$. It holds

$$\|\widetilde{T}_{41}\|_\infty \quad \leq \quad \sup_{t\in\mathbb{R}} \Big| \int_{a(t)}^{a(t+h)} \partial_1 \widetilde{G}(t+2h,u)\, du \Big| + \sup_{t\in\mathbb{R}} \Big| \int_{a(t)}^{a(t+h)} \partial_1 \widetilde{G}(t+h,u)\, du \Big|.$$

Analogously, we have

$$\begin{aligned} \|\widetilde{T}_{43}\|_\infty \quad \leq \quad & \sup_{t\in\mathbb{R}} \Big| \int_{a(t)}^{a(t+h)} \partial_1 \widetilde{G}(t+3h,u)\, du \Big| + 2\sup_{t\in\mathbb{R}} \Big| \int_{a(t)}^{a(t+h)} \partial_1 \widetilde{G}(t+2h,u)\, du \Big| \\ & + \sup_{t\in\mathbb{R}} \Big| \int_{a(t)}^{a(t+h)} \partial_1 \widetilde{G}(t+h,u)\, du \Big|. \end{aligned}$$

All these terms on the right-hand side can be estimated in the same way as

$$\widetilde{T}_3(t) = \int_t^{t+h} \partial_1 \widetilde{G}(t+h, a(u)) a'(u)\, du.$$

This finishes the proof. □

### 6.5.4 Proof of Theorem 6.3.7

Lemma 6.5.4 enables us to prove a generalization of Proposition 6.3.5.

**Proposition 6.5.8** ([60, Prop. 6.23]). *We have for $m = (m_1, m_2) \in \mathbb{N}_0^2$ the estimate*

$$\begin{aligned} &\int_{|\lambda|\sim 2^{j(1-\alpha)}} |\partial^m \widehat{F_j}(\lambda\cos\eta, \lambda\sin\eta)|^2\, d\lambda \\ &\qquad \lesssim 2^{-2(1-\alpha)jm_1} 2^{-(1-\alpha)j} \big(1 + 2^{(1-\alpha)j}\sin\eta\big)^{-1-2\beta} + 2^{3\alpha j} 2^{(-1-2\beta)j}. \end{aligned}$$

*Proof.* Observe that $\partial^m \widehat{F_j} = (x^m F_j)^\wedge$. Putting $\overline{F}_j := x_2^{m_2} F_j$, the function $\widetilde{F}_j(x) := x^m F_j(x)$ takes the form of a modified edge fragment as defined in (6.35), i.e., $\widetilde{F}_j = x_1^{m_1} \overline{F}_j$. Analogous to Proposition 6.3.5 we distinguish between the cases $|\sin\eta| < 2\delta_j$ and $|\sin\eta| \geq 2\delta_j$.

In case $|\sin\eta| < 2\delta_j$ we show

$$\int_{|\lambda|\sim 2^{j(1-\alpha)}} |\partial^m \widehat{F_j}(\lambda\cos\eta, \lambda\sin\eta)|^2\, d\lambda \lesssim 2^{-2jm_1(1-\alpha)} 2^{-j(1-\alpha)}.$$

For this let $\overline{F}_j = F_j^0 + F_j^1$ be a decomposition similar to (6.24), where

$$F_j^0(x) := x_2^{m_2} g(2^{-j\alpha} x)\omega(x)\mathcal{X}_{\{x_1 \geq \delta_j\}}.$$

Further, we write $\widetilde{F}_j = \widetilde{F}_j^0 + \widetilde{F}_j^1$ with

$$\widetilde{F}_j^0(x) := x_1^{m_1} F_j^0(x) = x^m g(2^{-j\alpha}x)\omega(x)\mathcal{X}_{\{x_1 \geq \delta_j\}}$$

and $\widetilde{F}_j^1(x) := \widetilde{F}_j(x) - \widetilde{F}_j^0(x)$ the deviation. Note that $\widetilde{F}_j^0$ is a fragment with a straight edge of height about $\delta_j^{m_1}$ and that the function $\widetilde{F}_j^1$ is supported in a vertical strip of width $2\delta_j$.

For $\eta$ satisfying $|\sin\eta| < 2\delta_j$ the Radon transform $\mathcal{R}\widetilde{F}_j^1(\cdot,\eta)$ is $L_\infty$-bounded with $\|\mathcal{R}\widetilde{F}_j^1(\cdot,\eta)\|_\infty \lesssim \delta_j^{m_1}\|\mathcal{R}F_j^1(\cdot,\eta)\|_\infty \lesssim \delta_j^{m_1}$ and it is supported in an interval of length

$$2(\delta_j\cos\eta + \sin\eta) \lesssim \delta_j.$$

It follows $\|\mathcal{R}\widetilde{F}_j^1(\cdot,\eta)\|_2^2 \lesssim \delta_j^{2m_1}\delta_j \lesssim \delta_j^{2m_1}2^{-j(1-\alpha)}$. Therefore

$$\int_{|\lambda|\sim 2^{j(1-\alpha)}} |\mathcal{F}\widetilde{F}_j^1(\lambda\cos\eta,\lambda\sin\eta)|^2\,d\lambda \leq \int_{\mathbb{R}} |\widehat{\mathcal{R}\widetilde{F}_j^1(\cdot,\eta)}(\lambda)|^2\,d\lambda \lesssim \delta_j^{2m_1}2^{-j(1-\alpha)}.$$

It remains to show

$$\int_{|\lambda|\sim 2^{j(1-\alpha)}} |\mathcal{F}\widetilde{F}_j^0(\lambda\cos\eta,\lambda\sin\eta)|^2\,d\lambda \lesssim \delta_j^{2m_1}2^{-j(1-\alpha)}.$$

This follows from the fact, that we have decay $|\mathcal{F}\widetilde{F}_j^0(\lambda,0)| \lesssim \delta_j^{m_1}|\lambda|^{-1/2}$ normal to the straight singularity curve, since the height of the jump is $\delta_j^{m_1}$. Further, the second argument $(\lambda\sin\eta)$ remains bounded due to the condition $|\sin\eta| < 2\delta_j$.

In case $|\sin\eta| \geq 2\delta_j$ we conclude as follows. Let $C_1$, $C_2 > 0$ be the constants specifying the integration domain $[C_1 2^{j(1-\alpha)}, C_2 2^{j(1-\alpha)}]$. We choose $C > 0$ such that $C_2 C < 2\pi$ and fix $h := C2^{-j(1-\alpha)}$. Then there is $c > 0$ such that $|e^{i\lambda h}-1|^{m_1} \geq c$ for $|\lambda| \in [C_1 2^{j(1-\alpha)}, C_2 2^{j(1-\alpha)}]$ at all scales. We obtain

$$\begin{aligned}
&\int_{|\lambda|\sim 2^{j(1-\alpha)}} |\lambda|^2 |\partial^m \widehat{F_j}(\lambda\cos\eta,\lambda\sin\eta)|^2\,d\lambda \\
&\qquad\lesssim \int_{|\lambda|\sim 2^{j(1-\alpha)}} |e^{i\lambda h}-1|^2|\lambda|^2|\widehat{x^m F_j}(\lambda\cos\eta,\lambda\sin\eta)|^2\,d\lambda \\
&\qquad\lesssim \int_{|\lambda|\sim 2^{j(1-\alpha)}} |e^{i\lambda h}-1|^2|\lambda|^2|\left(\mathcal{R}\widetilde{F}_j(\cdot,\eta)\right)^\wedge(\lambda)|^2\,d\lambda \\
&\qquad\lesssim \int_{|\lambda|\sim 2^{j(1-\alpha)}} |[\Delta_h\partial_1\mathcal{R}\widetilde{F}_j(\cdot,\eta)]^\wedge(\lambda)|^2\,d\lambda.
\end{aligned}$$

From Lemma 6.5.4 we know that $S = \Delta_h\partial_1\mathcal{R}\widetilde{F}_j(\cdot,\eta)$ admits a decomposition $(S_1^k, S_2^k)_k$ of length $m_1+1$ with estimates

$$\|S_1^k\|_2^2 \lesssim 2^{-2jm_1(1-\alpha)}h^{2\beta}|\sin\eta|^{-1-2\beta} + 2^{-j(1-\alpha)(2\beta+1)}, \qquad k = 0,1,\dots,m_1+1.$$

Using the same trick as in Proposition 6.3.5, we can then conclude

$$\begin{aligned}
\int_{|\lambda|\sim 2^{j(1-\alpha)}} |\lambda|^2|\partial^m\widehat{F_j}(\lambda\cos\eta,\lambda\sin\eta)|^2\,d\lambda &\lesssim \int_{|\lambda|\sim 2^{j(1-\alpha)}} |[\Delta_h\partial_1\mathcal{R}\widetilde{F}_j(\cdot,\eta)]^\wedge(\lambda)|^2\,d\lambda \\
&\lesssim \sum_{k=0}^{m_1+1}\int_{|\lambda|\sim 2^{j(1-\alpha)}} |\widehat{S_1^k}|^2\,d\lambda \\
&\lesssim \sum_{k=0}^{m_1+1}\|S_1^k\|_2^2 \\
&\lesssim 2^{-2jm_1(1-\alpha)}h^{2\beta}|\sin\eta|^{-1-2\beta} + 2^{-j(1-\alpha)(2\beta+1)}.
\end{aligned}$$

It follows

$$\begin{aligned}\int_{|\lambda|\sim 2^{j(1-\alpha)}} |\partial^m \widehat{F_j}(\lambda\cos\eta, \lambda\sin\eta)|^2 \, d\lambda &\lesssim 2^{-2j(1-\alpha)(m_1+1)} h^{2\beta} |\sin\eta|^{-1-2\beta} + 2^{-j(1-\alpha)(2\beta+3)} \\ &\lesssim 2^{-2(1-\alpha)jm_1} 2^{-(1-\alpha)j} \big(2^{(1-\alpha)j}\sin\eta\big)^{-1-2\beta} + 2^{3\alpha j} 2^{(-1-2\beta)j}.\end{aligned}$$

This finishes the proof. □

By rescaling $F_j$ to the original edge fragment $f_j$ we obtain Theorem 6.3.7, because of the relation $\widehat{f}_j(\xi) = 2^{-2\alpha j}\widehat{F_j}(2^{-\alpha j}\xi)$.

# Chapter 7

# Multivariate $\alpha$-Molecules

The framework of $\alpha$-molecules presented so far is confined to a bivariate setting. Its applicability is thus limited and, since nowadays one often has to deal with higher dimensional data, an extension to higher dimensions is desirable. Such an extension was pursued in [45]. As a main result, a $d$-dimensional version of Theorem 2.2.2 ([45, Thm. 2.5]) could be proved.

Subsequently, we will present the results of [45], whereby we adapt the exposition to the continuous setting. As an exemplary application, we investigate the sparse approximation of video signals, which are instances of 3D data. The multivariate theory allows to derive almost optimal approximation rates for a large class of 3-variate $\frac{1}{2}$-molecule systems.

## 7.1 The Concept of $\alpha$-Molecules in $L_2(\mathbb{R}^d)$

Recalling Definition 2.1.3, a system of bivariate $\alpha$-molecules consists of functions in $L_2(\mathbb{R}^2)$ obtained by applying $\alpha$-scaling, rotations, and translations to a set of generating functions which need to be sufficiently localized in time and frequency. As a consequence, every $\alpha$-molecule is associated with a certain scale, orientation and spatial position, which – in the bivariate case – is conveniently represented by a point in the parameter space $\mathbb{P} = \mathbb{R}^2 \times \mathbb{T} \times \mathbb{R}_+$.

Aiming for a multivariate generalization, we first need to introduce a $d$-dimensional version of $\mathbb{P}$. Let $\mathbb{S}^{d-1}$ denote the unit sphere in $\mathbb{R}^d$. Then we put

$$\mathbb{P}_d := \mathbb{R}^d \times \mathbb{S}^{d-1} \times \mathbb{R}_+.$$

Since $\mathbb{S}^1$ can be identified with $\mathbb{T}$ via (2.2), we have $\mathbb{P}_2 \cong \mathbb{P}$ and $\mathbb{P}_d$ can be regarded as a canonical extension of $\mathbb{P}$ to $d$ dimensions.

As in the bivariate case, each function $m_\lambda \in L_2(\mathbb{R}^d)$ of a system of $d$-variate $\alpha$-molecules $\{m_\lambda\}_{\lambda\in\Lambda}$ shall be associated with a point $\mathbf{x}_\lambda := (x_\lambda, e_\lambda, t_\lambda) \in \mathbb{P}_d$, where the variable $t_\lambda \in \mathbb{R}_+$ represents the scale, the vector $e_\lambda \in \mathbb{S}^{d-1}$ the orientation, and $x_\lambda \in \mathbb{R}^d$ the spatial position of $m_\lambda$. The relation between the index $\lambda$ of a molecule $m_\lambda$ and its *phase space coordinates* $\mathbf{x}_\lambda \in \mathbb{P}_d$ is again described by a so-called *parametrization*.

**Definition 7.1.1** (compare [45, Def. 2.1])**.** A *parametrization* is a pair $(\Lambda, \Phi_\Lambda)$, where $\Lambda$ is an index set and $\Phi_\Lambda$ a mapping

$$\Phi_\Lambda : \Lambda \to \mathbb{P}_d\,, \quad \lambda \mapsto \mathbf{x}_\lambda = (x_\lambda, e_\lambda, t_\lambda),$$

which associates to each $\lambda \in \Lambda$ a *scale* $t_\lambda \in \mathbb{R}_+$, an *orientation* $e_\lambda \in \mathbb{S}^{d-1}$, and a *location* $x_\lambda \in \mathbb{R}^d$.

For practical purposes it is more convenient to represent an orientation $\eta \in \mathbb{S}^{d-1}$ by a set of angles. Therefore we define the rotation matrix $R_\theta$ for $\theta = (\theta_1, \ldots, \theta_{d-2}) \in \mathbb{R}^{d-2}$ by

$$R_\theta := \begin{pmatrix} \cos(\theta_1) & & -\sin(\theta_1) \\ & I_{d-2} & \\ \sin(\theta_1) & & \cos(\theta_1) \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} \cos(\theta_{d-2}) & & -\sin(\theta_{d-2}) & \\ & 1 & & \\ \sin(\theta_{d-2}) & & \cos(\theta_{d-2}) & \\ & & & I_{d-3} \end{pmatrix},$$

where $I_d$ for $d \in \mathbb{N}$ denotes the $d$-dimensional identity matrix. Furthermore, we introduce for $\varphi \in \mathbb{R}$ the matrix

$$R_\varphi := \begin{pmatrix} \cos(\varphi) & \sin(\varphi) & \\ -\sin(\varphi) & \cos(\varphi) & \\ & & I_{d-2} \end{pmatrix}. \tag{7.1}$$

Note that these definitions pose an inconsistency in the notation, since they depend on the particular naming of the index. However, since we always use these particular indices, this will not lead to any problems while improving the readability significantly.

We now observe that each orientation $\eta \in \mathbb{S}^{d-1}$ can be uniquely represented by a set of angles $(\theta_1, \ldots, \theta_{d-2}, \varphi) \in [0, \pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}]^{d-3} \times [0, 2\pi]$ via the relation

$$\eta = R_\varphi^T R_\theta^T e_d, \tag{7.2}$$

where $e_d$ is the $d$th unit vector of $\mathbb{R}^d$. This representation is similar to a representation by Eulerian angles. Explicitly, $\eta$ is given by

$$\eta(\theta, \varphi) = \begin{pmatrix} \eta_1(\theta, \varphi) \\ \vdots \\ \eta_d(\theta, \varphi) \end{pmatrix} = \begin{pmatrix} \cos(\varphi)\cos(\theta_{d-2}) \cdot \cdots \cdot \cos(\theta_2)\sin(\theta_1) \\ \sin(\varphi)\cos(\theta_{d-2}) \cdot \cdots \cdot \cos(\theta_2)\sin(\theta_1) \\ -\sin(\theta_{d-2})\cos(\theta_{d-3}) \cdots \cos(\theta_2)\sin(\theta_1) \\ \vdots \\ -\sin(\theta_3)\cos(\theta_2)\sin(\theta_1) \\ -\sin(\theta_2)\sin(\theta_1) \\ \cos(\theta_1) \end{pmatrix}.$$

We also need to adapt the $\alpha$-scaling matrix (2.4) to the multivariate setting. For $\alpha \in [0, 1]$, we set

$$A_{\alpha,t} := \begin{pmatrix} t^\alpha I_{d-1} & \\ & t \end{pmatrix}, \quad t \in \mathbb{R}_+. \tag{7.3}$$

In case $\alpha = 1$ this matrix scales isotropically, in the range $\alpha \in [0, 1)$ it scales uniformly in all directions except for the $e_d$-direction. Hence, in contrast to the matrix (2.4), where $e_1$ was chosen as the distinguished direction in which the scaling is stronger, here we choose $e_d$ .

After this preparation we are ready to give the definition of a system of $d$-variate $\alpha$-molecules, $d \in \mathbb{N}\backslash\{1\}$, which essentially reduces to Definition 2.1.3 for $d = 2$, except for the interchanged roles of the directions $e_1$ and $e_d$ and an inversion of the utilized rotation matrix since, if $d = 2$, the matrix (7.1) is the inverse of the matrix (2.3).

Recall that we use the notation $[x]_i := \langle x, e_i \rangle$, $i \in \{1, \ldots, d\}$, for the $i$:th component of a vector $x \in \mathbb{R}^d$. Further, we define $|x|_{[d-1]} := |([x]_1, \ldots, [x]_{d-1}, 0)|_2$.

**Definition 7.1.2** (compare [45, Def. 2.2])**.** Let $\alpha \in [0,1]$, $d \in \mathbb{N}\backslash\{1\}$, and $L, M, N_1, N_2 \in \mathbb{N}_0 \cup \{\infty\}$. Further let $(\Lambda, \Phi_\Lambda)$ be a parametrization with $\Phi_\Lambda(\lambda) = (x_\lambda, e_\lambda, t_\lambda) \in \mathbb{P}_d$ for $\lambda \in \Lambda$. The corresponding angles (7.2) for $e_\lambda$ shall be denoted by $(\theta_\lambda, \varphi_\lambda)$. A family of functions $\{m_\lambda\}_{\lambda\in\Lambda} \subseteq L_2(\mathbb{R}^d)$ is called a *system of d-variate $\alpha$-molecules of order* $(L, M, N_1, N_2)$ *with respect to the parametrization* $(\Lambda, \Phi_\Lambda)$, if each $m_\lambda$ is of the form

$$m_\lambda = t_\lambda^{-\frac{1+(d-1)\alpha}{2}} g_\lambda(A_{\alpha,t_\lambda}^{-1} R_{\theta_\lambda} R_{\varphi_\lambda}(\cdot - x_\lambda)) \tag{7.4}$$

with *generators* $g_\lambda \in L_2(\mathbb{R}^d)$ satisfying for every multi-index $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \leq L$ the condition

$$|\partial^\rho \hat{g}_\lambda(\xi)| \lesssim \min\left\{1, t_\lambda + |[\xi]_d| + t_\lambda^{1-\alpha}|\xi|_{[d-1]}\right\}^M \langle|\xi|\rangle^{-N_1}\langle|\xi|_{[d-1]}\rangle^{-N_2}. \tag{7.5}$$

The implicit constant in (7.5) is required to be uniform in $\Lambda$. In case that a control parameter takes the value $\infty$, this shall mean that the condition (7.5) is fulfilled with the respective quantity arbitrarily large.

A system of $d$-variate $\alpha$-molecules $\{m_\lambda\}_{\lambda\in\Lambda}$ is thus obtained in the same way as a system of bivariate $\alpha$-molecules. One applies rotations, translations, and $\alpha$-scaling to a set of generating functions $\{g_\lambda\}_\lambda$ which are required to obey a prescribed time-frequency localization. This localization is specified by (7.5), where the number $L$ describes the spatial localization, $M$ the number of directional almost vanishing moments, and $N_1, N_2$ the smoothness of $g_\lambda$.

Applying $A_{\alpha,t}^{-1}$ with $\alpha < 1$ and $t < 1$ to the unit ball $B := \{x \in \mathbb{R}^d : |x| \leq 1\}$ stretches $B$ in the $e_d$-direction. For small $t \in \mathbb{R}_+$ this results in a plate-like support of the characteristic function $\mathcal{X}_B(A_{\alpha,t}^{-1}\cdot)$. At high scales, $d$-variate $\alpha$-molecules thus resemble plate-like objects in the spatial domain with the 'plate' lying in the plane spanned by the vectors $\{e_1, \dots, e_{d-1}\}$. The approximate frequency support on the other hand is concentrated in a pair of opposite wedges in the direction of the respective orientation.

Also note that the weighting function on the right-hand side of (7.5) is symmetric with respect to rotations around the $e_d$-axis, as well as reflections along this axis.

**Remark 7.1.3.** It may seem more natural to choose a rotation $R_\eta$ from $e_d$ to $\eta \in \mathbb{S}^{d-1}$ in the $(e_d, \eta)$-plane to adjust the orientation in (7.4). Due to the symmetries of the weighting function of the generators in (7.5), this choice is however not necessary. Since it is easier to use fixed rotation planes, we stick to this more pragmatic choice of rotation parameters.

## 7.2 The Index Distance

A central ingredient of the theory of bivariate $\alpha$-molecules is the fact that the parameter space $\mathbb{P}$ can be equipped with a natural phase-space metric $\omega_\alpha$ such that the distance between two points in $\mathbb{P}$ 'anti-correlates' with the size of the cross-correlations of $\alpha$-molecules associated with those points: The greater the distance between two points $\mathbf{x}_\lambda, \mathbf{x}_\mu \in \mathbb{P}$, the smaller the modulus of the scalar product of corresponding $\alpha$-molecules $m_\lambda, \tilde{m}_\mu \in L_2(\mathbb{R}^2)$.

Our next aim is to find a suitable analogon of this phase-space metric for the parameter space $\mathbb{P}_d$. Thereby, for simplicity, we do not generalize $\omega_\alpha$ from Definition 2.2.1 but the simplified version $\omega_\alpha^{\mathrm{sim}}$ from Definition 2.2.3.

For orientations $e_\lambda, e_\mu \in \mathbb{S}^{d-1}$, we define the angle $d_{\mathbb{S}}(e_\lambda, e_\mu) := \arccos(\langle e_\lambda, e_\mu \rangle)$ with $d_{\mathbb{S}}(e_\lambda, e_\mu) \in [0, \pi]$. Again, the angle $d_{\mathbb{S}}(e_\lambda, e_\mu)$ is projected onto the interval $\mathbf{T} := [-\pi/2, \pi/2)$, with $\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}$ being the unique element of the set $\{d_{\mathbb{S}}(e_\lambda, e_\mu) + n\pi \mid n \in \mathbb{Z}\}$ in $\mathbf{T}$. Analogous to the bivariate case, a suitable measure for the distance on $\mathbb{S}^{d-1}$ is then given by the quantity $|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|$. Note that in two dimensions we have $|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}| = |\{\varphi_\lambda - \varphi_\mu\}|$.

We arrive at the following definition which directly generalizes the simplified metric from Definition 2.2.3.

**Definition 7.2.1** ([45, Def. 2.4])**.** Let $\alpha \in [0,1]$, $d \in \mathbb{N}\backslash\{1\}$. The *index distance* $\omega_\alpha : \mathbb{P}_d \times \mathbb{P}_d \to [1, \infty)$ is defined by

$$\omega_\alpha(\mathbf{x}_\lambda, \mathbf{x}_\mu) := \max\left\{\frac{t_\lambda}{t_\mu}, \frac{t_\mu}{t_\lambda}\right\}(1 + d_\alpha(\mathbf{x}_\lambda, \mathbf{x}_\mu)),$$

where $\mathbf{x}_\lambda := (x_\lambda, e_\lambda, t_\lambda)$, $\mathbf{x}_\mu := (x_\mu, e_\mu, t_\mu) \in \mathbb{P}_d$ and with $t_0 := \max\{t_\lambda, t_\mu\}$

$$d_\alpha(\mathbf{x}_\lambda, \mathbf{x}_\mu) := t_0^{-2\alpha}|x_\lambda - x_\mu|^2 + t_0^{-2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2 + t_0^{-1}|\langle e_\lambda, x_\lambda - x_\mu \rangle|.$$

As in the bivariate case, this index distance $\omega_\alpha$ on $\mathbb{P}_d$ is quasi-symmetric and satisfies a quasi-triangle inequality. These properties, shown for the simplified version $\omega_\alpha^{\mathrm{sim}}$ in the 2-dimensional setting, can be found in [65]. Their proofs translate very well to higher dimensions. Other properties analogous to those proved in Section 2.2 can certainly be shown, but this has not been carried out explicitly.

Let us now state the analogon of Theorem 2.2.2 in the multivariate setting. It relates the index distance to the size of the cross-correlations of $\alpha$-molecules.

**Theorem 7.2.2** ([45, Thm. 2.5])**.** *Let $\alpha \in [0,1]$, $d \in \mathbb{N}\backslash\{1\}$, and let $\{m_\lambda\}_{\lambda\in\Lambda}$ and $\{\tilde{m}_\mu\}_{\mu\in\Delta}$ be two systems of $d$-variate $\alpha$-molecules of order $(L, M, N_1, N_2)$ with respect to parametrizations $(\Lambda, \Phi_\Lambda)$ and $(\Delta, \Phi_\Delta)$, respectively. Further, assume that there exists some constant $C > 0$ such that*

$$t_\lambda, t_\mu \leq C \quad \textit{for all } \lambda \in \Lambda, \mu \in \Delta \textit{ with } (x_\lambda, e_\lambda, t_\lambda) := \Phi_\Lambda(\lambda), (x_\mu, e_\mu, t_\mu) := \Phi_\Delta(\mu).$$

*If $N_1 > \frac{d}{2}$ and if there exists some positive integer $N \in \mathbb{N}$ such that*

$$L \geq 2N, \quad M > 3N - d + \frac{1 + (d-1)\alpha}{2}, \quad N_1 \geq N + \frac{1 + (d-1)\alpha}{2}, \quad N_2 \geq 2N + d - 2,$$

*then we have*

$$|\langle m_\lambda, p_\mu \rangle| \lesssim \omega_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu))^{-N} \quad \textit{for all } \lambda \in \Lambda, \mu \in \Delta.$$

*Proof.* The proof is analogous to the proof of Theorem 2.2.2. It is outsourced to Section 7.6. □

Based on Theorem 7.2.2, the same methodology as in the bivariate setting to categorize $\alpha$-molecule frames according to their sparse approximation behavior can be developed. For this we next formulate a $d$-dimensional version of the transfer principle, Theorem 2.3.6.

## 7.3 Transfer Principle and Consistency of Parametrizations

In this section we derive a $d$-dimensional version of Theorem 2.3.6. Beforehand, we need to adapt the notion of $(\alpha, k)$-consistency of parametrizations to $d$ dimensions. The definition is analogue to Definition 2.3.5.

**Definition 7.3.1** ([45, Def. 3.5])**.** Let $\alpha \in [0,1]$, $d \in \mathbb{N}\backslash\{1\}$, and $k > 0$. Two parametrizations $(\Lambda, \Phi_\Lambda)$ and $(\Delta, \Phi_\Delta)$ with $\Phi_\Lambda : \Lambda \to \mathbb{P}_d$, $\Phi_\Delta : \Delta \to \mathbb{P}_d$, are called *$(\alpha, k)$-consistent*, if

$$\sup_{\lambda\in\Lambda} \sum_{\mu\in\Delta} \omega_\alpha\big(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu)\big)^{-k} < \infty \quad \text{and} \quad \sup_{\mu\in\Delta} \sum_{\lambda\in\Lambda} \omega_\alpha\big(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu)\big)^{-k} < \infty.$$

Using Theorem 7.2.2 and Schur's test (Lemma 2.3.4), we obtain the following generalization of Theorem 2.3.6 in $d$ dimensions.

**Theorem 7.3.2** ([45, Thm. 3.7])**.** *Let $\alpha \in [0,1]$, $d \in \mathbb{N}\backslash\{1\}$, $k > 0$, and $0 < p \le 1$. Let $\{m_\lambda\}_{\lambda\in\Lambda}$ and $\{\tilde{m}_\mu\}_{\mu\in\Delta}$ be two frames of $d$-variate $\alpha$-molecules of order $(L, M, N_1, N_2)$ with $(\alpha, k)$-consistent parametrizations $(\Lambda, \Phi_\Lambda)$ and $(\Delta, \Phi_\Delta)$ satisfying*

$$t_\lambda, t_\mu \le C, \ \textit{for all } \lambda \in \Lambda, \mu \in \Delta$$

*and*

$$L \ge 2\frac{k}{p}, \quad M > 3\frac{k}{p} - d + \frac{1 + \alpha(d-1)}{2}, \quad N_1 > \frac{d}{2}, N_1 \ge \frac{k}{p} + \frac{1 + \alpha(d-1)}{2}, \quad N_2 \ge 2\frac{k}{p} + d - 2.$$

*Then $\{m_\lambda\}_{\lambda\in\Lambda}$ and $\{\tilde{m}_\mu\}_{\mu\in\Delta}$ are sparsity equivalent in $\ell^q$ for all $p \le q < 2$.*

*Proof.* By Lemma 2.3.4, it suffices to prove that

$$\max\Big\{ \sup_{\lambda\in\Lambda} \sum_{\mu\in\Delta} |\langle m_\lambda, \tilde{m}_\mu\rangle|^p, \sup_{\mu\in\Delta} \sum_{\lambda\in\Lambda} |\langle m_\lambda, \tilde{m}_\mu\rangle|^p \Big\}^{1/p} < \infty.$$

Since, by Theorem 7.2.2, we have $|\langle m_\lambda, \tilde{m}_\mu\rangle| \lesssim \omega_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu))^{-k/p}$, we can conclude that

$$\begin{aligned} &\max\Big\{ \sup_{\lambda\in\Lambda} \sum_{\mu\in\Delta} |\langle m_\lambda, \tilde{m}_\mu\rangle|^p, \sup_{\mu\in\Delta} \sum_{\lambda\in\Lambda} |\langle m_\lambda, \tilde{m}_\mu\rangle|^p \Big\} \\ &\qquad \lesssim \max\Big\{ \sup_{\lambda\in\Lambda} \sum_{\mu\in\Delta} \omega_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu))^{-k}, \sup_{\mu\in\Delta} \sum_{\lambda\in\Lambda} \omega_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu))^{-k} \Big\}, \end{aligned}$$

with the expression on the right-hand side being finite due to the $(\alpha, k)$-consistency of the parametrizations $(\Lambda, \Phi_\Lambda)$ and $(\Delta, \Phi_\Delta)$. The proof is completed. □

This theorem allows to categorize frames of $d$-variate $\alpha$-molecules according to their sparse approximation behavior, similar to Theorem 2.3.6 in the bivariate case. In the sequel, we will apply this methodology with respect to video data, which is modelled as 3D cartoon functions. Before that, however, we introduce multivariate $\alpha$-shearlet molecules, a large subclass of multivariate $\alpha$-molecule systems.

## 7.4 Multivariate $\alpha$-Shearlet Molecules

In this section, we introduce a very general class of shear-based systems in $d$ dimensions, namely *systems of d-variate $\alpha$-shearlet molecules*. The definition is analogous to the bivariate case [59]. Roughly speaking, they are shear-based systems obtained from variable generators, where similar to $\alpha$-molecules the conditions on the generators have been relaxed to a mere time-frequency localization requirement. The notion of $\alpha$-shearlet molecules comprises many specific shear-based constructions and simplifies the treatment of such systems within the general framework of $\alpha$-molecules.

**Remark 7.4.1.** One might wonder if there also exists a natural generalization of the concept of (discrete) $\alpha$-curvelet molecules to dimensions $d > 2$. Up to now, no such generalization has been put forward. A major difficulty is the question of how to suitably discretize the sphere $\mathbb{S}^{d-1}$ to obtain the discrete rotation parameters. This problem is avoided when using the shearlet approach.

As explained in Section 3.3, shearlet-like constructions are based on anisotropic scaling, shearings, and translations. For the change of scale, we utilize $\alpha$-scaling as defined by (7.3). The change of orientation is provided by shearings, in $d$ dimensions given by the shearing matrices

$$S_h = \begin{pmatrix} I_{d-1} & 0 \\ h^T & 1 \end{pmatrix} \quad \text{and} \quad S_h^T = \begin{pmatrix} I_{d-1} & h \\ 0 & 1 \end{pmatrix}, \quad h \in \mathbb{R}^{d-1},$$

which are the natural generalizations of (1.2). The matrix $S_h^T$ shears parallel to the $(e_1, \dots, e_{d-1})$-plane and the shear vector $h \in \mathbb{R}^{d-1}$ determines the direction of the shearing in this plane. Note that the transformations associated with shearings and $\alpha$-scalings naturally form a group [25].

To avoid directional bias, the frequency domain is divided into cone-like regions along the coordinate axes and a coarse-scale box for the low frequencies. Note that this comes at the cost of the loss of the group properties mentioned above. This division procedure is however crucial for applications, and also, as the subsequent arguments will show, for including $\alpha$-shearlets in the concept of $\alpha$-molecules. The pyramids are defined as

$$\mathcal{P}_\varepsilon = \left\{ (\xi_1, \dots, \xi_d) \in \mathbb{R}^d \mid \forall i = 1, \dots d : |\xi_i| \le |\xi_\epsilon| \right\},$$

where $\varepsilon \in \{1, \dots, d\}$. $\varepsilon = 0$ shall refer to the coarse-scale box $\mathcal{R}$. In the sequel we will always stay in this so-called cone-adapted setting. For an illustration of this specific setting in 3D, we refer to Subsection 7.5.2.

In each cone we require different versions of the scaling and shearing operators. The cyclic permutation matrix

$$Z = \begin{pmatrix} 0 & 1 \\ I_{d-1} & 0 \end{pmatrix} \tag{7.6}$$

allows to elegantly define these operators associated with the respective cones by $Z^\varepsilon S_h Z^{-\varepsilon}$ and $Z^\varepsilon A_{\alpha,t} Z^{-\varepsilon}$.

Before we come to the definition of $\alpha$-shearlet molecules, we need to introduce a set of characteristic parameters, associated with these systems. The resolution of the underlying sampling grid is determined by the parameters $\sigma > 1$, $\tau_1, \ldots, \tau_d > 0$, and a sequence $\Theta = (\eta_j)_{j\in\mathbb{N}_0} \in \mathbb{R}_+^{\mathbb{N}_0}$. The parameter $\sigma$ specifies the fineness of the scale sampling. The parameters $\tau_\varepsilon$, $\varepsilon \in \{1,\ldots,d\}$, determine the spatial resolution in the $e_\varepsilon$-direction. For convenience they are summarized in the diagonal matrix $\mathcal{T} := \operatorname{diag}(\tau_1,\ldots,\tau_d) \in \mathbb{R}^{d\times d}$. The angular resolution at each scale $j \in \mathbb{N}_0$ is given by the value $\eta_j$ of the sequence $\Theta$. Last but not least, in each cone $\varepsilon \in \{1,\ldots,d\}$ and at each scale $j \in \mathbb{N}_0$ the shearing parameter $\ell$ is restricted to a set $\mathscr{L}_{\varepsilon,j}$. These sets are collected in $\mathscr{L} := \{\mathscr{L}_{\varepsilon,j} : \varepsilon \in \{1,\ldots,d\},\ j \in \mathbb{N}_0\}$.

After the introduction of this sampling data $\mathbb{D} := \{\sigma, \Theta, \mathscr{L}, \mathcal{T}\}$ we can now give the definition of a system of $\alpha$-shearlet molecules in $d$ dimensions, depending on $\mathbb{D}$. The scale-dependent step size $\eta_j$ of the directional sampling is assumed to satisfy $\eta_j \asymp \sigma^{-j(1-\alpha)}$ for $j \in \mathbb{N}_0$. Further, we require the upper bounds $\mathbf{L}_j := \max\{|\ell|_\infty : \ell \in \mathscr{L}_{\varepsilon,j}, \varepsilon \in \{1,\ldots,d\}\}$, $j \in \mathbb{N}_0$, to fulfill the complementary condition $\mathbf{L}_j \lesssim \sigma^{j(1-\alpha)}$. We remark, that the translation parameters $\tau_\varepsilon$ may also vary with the indices $(\varepsilon, j, \ell)$, as long as their values are restricted to some fixed interval $[\tau_{min}, \tau_{max}]$ with $0 < \tau_{min} \le \tau_{max} < \infty$. However, this is not indicated in the notation.

**Definition 7.4.2** ([45, Def. 5.1])**.** Let $\alpha \in [0,1]$, $d \in \mathbb{N}\backslash\{1\}$, and $L, M, N_1, N_2 \in \mathbb{N}_0 \cup \{\infty\}$. Further the sampling data $\mathbb{D}$ shall be given as above. For $\varepsilon \in \{1,\ldots,d\}$, a system of functions

$$\Sigma_\varepsilon := \Big\{ m^\varepsilon_{j,\ell,k} \in L_2(\mathbb{R}^d) \ : \ (j,\ell,k) \in \Lambda^s_\varepsilon \Big\},$$

indexed by the set $\Lambda^s_\varepsilon := \{(j,\ell,k) \ : \ j \in \mathbb{N}_0,\ \ell \in \mathscr{L}_{\varepsilon,j} \subseteq \mathbb{Z}^{d-1},\ k \in \mathbb{Z}^d\}$, is called a *system of d-variate $\alpha$-shearlet molecules of order* $(L, M, N_1, N_2)$ *associated with the orientation* $\varepsilon$, if it is of the form

$$m^\varepsilon_{j,\ell,k}(x) = \sigma^{\frac{(1+\alpha(d-1))j}{2}} \gamma^\varepsilon_{j,\ell,k}(Z^\varepsilon A^j_{\alpha,\sigma} S_{\ell\eta_j} Z^{-\varepsilon} x - \mathcal{T}k)$$

with generating functions $\gamma^\varepsilon_{j,\ell,k} \in L_2(\mathbb{R}^d)$ satisfying for every $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \le L$

$$|\partial^\rho \hat{\gamma}^\varepsilon_{j,\ell,k}(\xi)| \lesssim \frac{\min\{1, \sigma^{-j} + \sigma^{-(1-\alpha)j}|Z^{-\varepsilon}\xi|_{[d-1]} + |[Z^{-\varepsilon}\xi]_d|\}^M}{\langle|\xi|\rangle^{N_1} \langle|Z^{-\varepsilon}\xi|_{[d-1]}\rangle^{N_2}}. \tag{7.7}$$

The implicit constant is required to be uniform over $\Lambda^s_\varepsilon$. If one of the parameters $L, M, N_1, N_2$ takes the value $\infty$, this shall mean that condition (7.7) is fulfilled with the respective quantity arbitrarily large.

Combining systems of $\alpha$-shearlet molecules of order $(L, M, N_1, N_2)$ for each orientation $\varepsilon \in \{1,\ldots,d\}$ with a system of coarse-scale elements

$$\Sigma_0 := \Big\{ m^0_{0,\mathbf{0},k} := \gamma^0_{0,\mathbf{0},k}(\cdot - \mathcal{T}k) \ : \ k \in \mathbb{Z}^d \Big\}, \tag{7.8}$$

where the generators $\gamma^0_{0,\mathbf{0},k} \in L_2(\mathbb{R}^d)$ fulfill $|\partial^\rho \hat{\gamma}^0_{0,\mathbf{0},k}(\xi)| \lesssim \langle|\xi|\rangle^{-N_1} \langle|\xi|_{[d-1]}\rangle^{-N_2}$ for every $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \le L$, yields a *system of $\alpha$-shearlet molecules* of order $(L, M, N_1, N_2)$. The associated index set is

$$\Lambda^s_0 := \Big\{ (0, \mathbf{0}, k) : k \in \mathbb{Z}^d \Big\} \subset \mathbb{N}_0 \times \mathbb{Z}^{d-1} \times \mathbb{Z}^d.$$

**Definition 7.4.3** ([45, Def. 5.2])**.** For each $\varepsilon \in \{1, \dots, d\}$, let $\Sigma_\varepsilon$ be a system of $\alpha$-shearlet molecules of order $(L, M, N_1, N_2)$ associated with the respective orientation. Further, let $\Sigma_0$ be a system of coarse-scale elements defined as in (7.8). Then the union

$$\Sigma := \bigcup_{\varepsilon=0}^{d} \Sigma_\varepsilon$$

is called a *system of d-variate $\alpha$-shearlet molecules of order* $(L, M, N_1, N_2)$. The associated *$\alpha$-shearlet index set* is given by

$$\Lambda^s = \{(\varepsilon, j, \ell, k) : \varepsilon \in \{0, \dots, d\},\ (j, \ell, k) \in \Lambda^s_\varepsilon\}.$$

In the next subsection, we will show that the concept of $d$-variate $\alpha$-shearlet molecules fits into the general theory of multivariate $\alpha$-molecules.

### 7.4.1 The $\alpha$-Shearlet Parametrization

As the following theorem shows, $d$-variate $\alpha$-shearlet molecules constitute a subclass of $d$-variate $\alpha$-molecule systems. The respective parametrizations are referred to as *$\alpha$-shearlet parametrizations* and generalize the bivariate notion.

**Theorem 7.4.4** ([45, Thm. 5.3])**.** *Let $\alpha \in [0, 1]$, $d \in \mathbb{N}\backslash\{1\}$, and $\Sigma = \{m_\lambda\}_{\lambda\in\Lambda^s}$ be a system of d-variate $\alpha$-shearlet molecules of order $(L, M, N_1, N_2)$. Then $\Sigma$ constitutes a system of d-variate $\alpha$-molecules of the same order. The associated* $\alpha$-shearlet parametrization $(\Lambda^s, \Phi^s)$ *is given by the map $\Phi^s(\lambda) = (x_\lambda, e_\lambda, t_\lambda) \in \mathbb{P}_d$, where for $\lambda = (\varepsilon, j, \ell, k) \in \Lambda^s$*

$$t_\lambda = \sigma^{-j}, \quad e_\lambda = n_\lambda \cdot Z^\varepsilon \begin{pmatrix} \eta_j \ell \\ 1 \end{pmatrix}, \quad x_\lambda = Z^\varepsilon S^{-1}_{\ell\eta_j} A^{-j}_{\alpha,\sigma} Z^{-\varepsilon} \mathcal{T} k \tag{7.9}$$

*with normalization constant $n_\lambda = (1 + \eta_j^2 |\ell|_2^2)^{-1/2}$.*

In particular, for $\varepsilon = 0$ we have $t_\lambda = 1$, $e_\lambda = e_d$, and $x_\lambda = \mathcal{T}k$ for every $\lambda = (0, 0, \mathbf{0}, k) \in \Lambda^s$.

*Proof.* Since a finite union of systems of $\alpha$-molecules is itself a system of $\alpha$-molecules, we can prove this theorem separately for each system $\Sigma_\varepsilon$, $\varepsilon \in \{0, \dots, d\}$. For $\Sigma_0$ the statement is obvious. For the other systems it suffices to give the proof for $\varepsilon = d$, since they are all related by a mere permutation of indices. We subsequently drop the index $\varepsilon$ to simplify the notation and note $Z^\varepsilon = I$ for $\varepsilon = d$.

For $\lambda = (d, j, \ell, k) \in \Lambda^s_d$ let $m_\lambda$ be an $\alpha$-shearlet molecule with corresponding generating function $\gamma_\lambda$. As usual we denote the angles representing the orientation $e_\lambda$ by $(\theta_\lambda, \varphi_\lambda)$, i.e. $e_\lambda = R^T_{\varphi_\lambda} R^T_{\theta_\lambda} e_d$. The molecule $m_\lambda$ can clearly be written in the form (7.4) with respect to the generator

$$g_\lambda(x) := \gamma^d_{j,\ell,k}(A^j_{\alpha,\sigma} S_{\ell\eta_j} R^T_{\varphi_\lambda} R^T_{\theta_\lambda} A^{-j}_{\alpha,\sigma} x), \quad x \in \mathbb{R}^d.$$

It remains to check condition (7.5) for these functions. On the Fourier side we have

$$\hat{g}_\lambda(\xi) = \hat{\gamma}^d_{j,\ell,k}(A^{-j}_{\alpha,\sigma} S^{-T}_{\ell\eta_j} R^T_{\varphi_\lambda} R^T_{\theta_\lambda} A^j_{\alpha,\sigma} \xi), \quad \xi \in \mathbb{R}^d.$$

For $\lambda = (d, j, \ell, k) \in \Lambda_d^s$ let us first examine the matrix

$$M_\lambda := S_{\ell\eta_j}^{-T} R_{\varphi_\lambda}^T R_{\theta_\lambda}^T. \tag{7.10}$$

A simple calculation shows $M_\lambda e_d = S_{\ell\eta_j}^{-T} R_{\varphi_\lambda}^T R_{\theta_\lambda}^T e_d = S_{\ell\eta_j}^{-T} e_\lambda = S_{\ell\eta_j}^{-T} n_\lambda(\eta_j \ell, 1)^T = n_\lambda e_d$. Hence, the entries of the last column of $M_\lambda$ vanish except for the last one. Next, we prove the uniform boundedness of the set of operators $\{M_\lambda\}_{\lambda \in \Lambda_d^s}$. It holds uniformly for $\lambda \in \Lambda_d^s$

$$\|M_\lambda\|_{2\to2} = \|S_{\ell\eta_j}^{-T}\|_{2\to2} \leq \sqrt{d + \eta_j^2 |\ell|_2^2} \lesssim \sqrt{d + \eta_j^2 \mathbf{L}_j^2} \lesssim 1.$$

Note that this implies that each entry in $M_\lambda$ is bounded in modulus. Since similar considerations hold for the inverse $M_\lambda^{-1} = R_{\theta_\lambda} R_{\varphi_\lambda} S_{\ell\eta_j}^T$, we can conclude that both $\widetilde{M}_\lambda := A_{\alpha,\sigma}^{-j} M_\lambda A_{\alpha,\sigma}^j$ and its inverse $\widetilde{M}_\lambda^{-1}$ have the form

$$\begin{pmatrix} * & \dots & * & 0 \\ * & \ddots & * & \vdots \\ * & \dots & * & 0 \\ \square & \dots & \square & * \end{pmatrix},$$

where the entries $*$ are the same as in $M_\lambda$ (or $M_\lambda^{-1}$) and the entries $\square$ are of the form $\sigma^{-j(1-\alpha)}[M_\lambda e_i]_d$ (or $\sigma^{-j(1-\alpha)}[M_\lambda^{-1} e_i]_d$) for $i \in \{1, \dots, d-1\}$. In particular, the entries of $\widetilde{M}_\lambda$ and $\widetilde{M}_\lambda^{-1}$ are uniformly bounded in modulus. This implies $\|\widetilde{M}_\lambda\|_{2\to2} \lesssim 1$ and $\|\widetilde{M}_\lambda^{-1}\|_{2\to2} \lesssim 1$. Altogether, we obtain

$$|\widetilde{M}_\lambda \xi| \asymp |\xi| \text{ uniformly for } \xi \in \mathbb{R}^d \text{ and } \lambda \in \Lambda_d^s. \tag{7.11}$$

Due to the structure of the last column of $\widetilde{M}_\lambda$ we further have for $\xi = (\xi_1, \dots, \xi_d)^T \in \mathbb{R}^d$

$$|\widetilde{M}_\lambda \xi|_{[d-1]} = |\widetilde{M}_\lambda (\xi_1, \dots, \xi_{d-1}, 0)^T|_{[d-1]} \leq \|\widetilde{M}_\lambda\|_{2\to2} |(\xi_1, \dots, \xi_{d-1}, 0)^T| = \|\widetilde{M}_\lambda\|_{2\to2} |\xi|_{[d-1]}.$$

For the inverse $\widetilde{M}^{-1}$ it holds analogously $|\widetilde{M}_\lambda^{-1} \xi|_{[d-1]} \leq \|\widetilde{M}_\lambda^{-1}\|_{2\to2} |\xi|_{[d-1]}$. We conclude

$$|\widetilde{M}_\lambda \xi|_{[d-1]} \asymp |\xi|_{[d-1]} \text{ uniformly for } \xi \in \mathbb{R}^d \text{ and } \lambda \in \Lambda_d^s. \tag{7.12}$$

Finally, the following estimate holds uniformly for $\xi = (\xi_1, \dots, \xi_d)^T \in \mathbb{R}^d$ and $\lambda \in \Lambda_d^s$,

$$|[\widetilde{M}_\lambda \xi]_d| \leq |[M_\lambda e_d]_d| |\xi_d| + \sum_{i=1}^{d-1} \sigma^{-j(1-\alpha)} |[M_\lambda e_i]_d| |\xi_i| \lesssim \sigma^{-(1-\alpha)j} |\xi|_{[d-1]} + |\xi_d|. \tag{7.13}$$

Finally, we can prove (7.5) for every $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \leq L$,

$$\begin{aligned} |\partial^\rho \hat{g}_\lambda(\xi)| \lesssim \sup_{|\beta|_1 \leq L} |(\partial^\beta \hat{\gamma}_{j,\ell,k}^d)(\widetilde{M}_\lambda \xi)| &\lesssim \frac{\min\{1, \sigma^{-j} + \sigma^{-(1-\alpha)j} |\widetilde{M}_\lambda \xi|_{[d-1]} + |[\widetilde{M}_\lambda \xi]_d|\}^M}{\langle |\widetilde{M}_\lambda \xi| \rangle^{N_1} \langle |\widetilde{M}_\lambda \xi|_{[d-1]} \rangle^{N_2}} \\ &\lesssim \frac{\min\{1, \sigma^{-j} + \sigma^{-(1-\alpha)j} |\xi|_{[d-1]} + |[\xi]_d|\}^M}{\langle |\xi| \rangle^{N_1} \langle |\xi|_{[d-1]} \rangle^{N_2}}. \end{aligned}$$

The first estimate holds true, since $\hat{g}_\lambda(\xi) = \hat{\gamma}_\lambda^d(\widetilde{M}_\lambda \xi)$ and the entries of $\widetilde{M}_\lambda$ are uniformly bounded in $\lambda$. The second estimate is due to (7.7). For the last estimate we used (7.11), (7.12), and (7.13). The observation $t_\lambda = \sigma^{-j}$ finishes the proof. $\square$

### 7.4.2 Consistency of $\alpha$-Shearlet Parametrizations

In view of Theorem 7.3.2 the consistency of parametrizations is of particular interest when comparing the approximation properties of different $\alpha$-molecule systems. In this paragraph we shall prove, in Proposition 7.4.7, that – as in the bivariate setting – the $\alpha$-shearlet parametrizations in $d$ dimensions are all consistent with each other. This allows to establish approximation rates for various shearlet-like constructions simultaneously, as long as they fall under the umbrella of the shearlet-molecule concept.

We start with an auxiliary lemma.

**Lemma 7.4.5** ([45, Lem. 5.4]). *Let $1 \geq c > 0$ be fixed, and consider the* gnomonic projection $\phi : \mathbb{R}^d \setminus \left\{x \in \mathbb{R}^d \mid [x]_d = 0\right\} \to \mathbb{R}^d, x \mapsto \frac{1}{[x]_d} x$. *For $v, w \in \mathbb{S}^{d-1} \cap \left\{x \in \mathbb{R}^d : [x]_d \geq c\right\}$ we then have $|\phi(v) - \phi(w)| \asymp |v - w|$ and $|v - w|_{[d-1]} \asymp |v - w|$.*

*Proof.* First note that $|v - w|_{[d-1]} = |\pi(v) - \pi(w)|$, where $\pi$ is the orthogonal projection of $\mathbb{R}^d$ onto the $(e_1, \dots, e_{d-1})$-plane. On the set $\mathbb{S}^{d-1} \cap \left\{x \in \mathbb{R}^d : [x]_d \geq c\right\}$, the mappings $\phi$ and $\pi$ are diffeomorphisms with bounded derivatives in both directions. This implies the statement. $\square$

We also need the following observation.

**Lemma 7.4.6** ([45, Lem. 6.3]). *Let $c > 0$ be a constant. Then we have for all $v, w \in \mathbb{S}^{d-1}$ with $[v]_d \geq c$ and $[w]_d \geq 0$*

$$|\{d_{\mathbb{S}}(v, w)\}| \asymp |v - w|.$$

*Proof.* Under the assumptions there exists $\varepsilon > 0$ dependent on $c$, such that $0 \leq d_{\mathbb{S}}(v, w) \leq \pi - \varepsilon$. It follows $\frac{\varepsilon}{\pi - \varepsilon} |d_{\mathbb{S}}(v, w)| \leq |\{d_{\mathbb{S}}(v, w)\}| \leq |d_{\mathbb{S}}(v, w)|$. The observation $d_{\mathbb{S}}(v, w) \asymp |v - w|$ finishes the proof. $\square$

After this preparation we are in the position to prove the consistency. Note that the proof is a slightly modified version of the proof given for [45, Prop. 5.6]. In particular, we do not need [45, Lem. 5.5].

**Proposition 7.4.7** ([45, Prop. 5.6]). *Let $\alpha \in [0, 1]$, $d \in \mathbb{N} \setminus \{1\}$, and let $(\Lambda, \Phi_\Lambda)$ and $(\Delta, \Phi_\Delta)$ be two $\alpha$-shearlet parametrizations, possibly with different parameters. Then $(\Lambda, \Phi_\Lambda)$ and $(\Delta, \Phi_\Delta)$ are $(\alpha, k)$-consistent for every $k > d$.*

*Proof.* Due to symmetry, it suffices to prove that for $N > d$ it holds

$$\sup_{\mu \in \Delta} \sum_{\lambda \in \Lambda} \omega_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu))^{-N} < \infty.$$

For this task it is convenient to decompose the shearlet index set $\Lambda = \Lambda_0 \cup \dots \cup \Lambda_d$ into the sets $\Lambda_\varepsilon$ associated with the respective pyramidal regions $\tilde{\mathcal{P}}_\varepsilon$ for $\varepsilon \in \{1, \dots, d\}$ and the low-frequency box $\mathcal{R}$ for $\varepsilon = 0$. The sum then splits accordingly into $d + 1$ parts, which we handle separately below.

$\Lambda_0$**:** Let $\mu \in \Delta$ and $\lambda = (0, 0, \mathbf{0}, k) \in \Lambda_0$ with $k \in \mathbb{Z}^d$. The shearlet parametrization (7.9) yields $t_\lambda = 1$, $e_\lambda = e_d$, and $x_\lambda = \mathcal{T}k$. Furthermore, $t_\mu \leq 1$ for all $\mu \in \Delta$. Hence we have

$$\begin{aligned}\omega_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu)) &= t_\mu^{-1}(1 + |\mathcal{T}k - x_\mu|^2 + |\{d_{\mathbb{S}}(e_d, e_\mu)\}|^2 + |\langle e_d, \mathcal{T}k - x_\mu\rangle|) \\ &\geq t_\mu^{-1}(1 + |\mathcal{T}k - x_\mu|^2).\end{aligned}$$

We conclude

$$\sum_{\lambda\in\Lambda_0} \omega_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu))^{-N} \leq \sum_{k\in\mathbb{Z}^d} t_\mu^N (1 + |\mathcal{T}k - x_\mu|^2)^{-N} \lesssim \sum_{k\in\mathbb{Z}^d} (1 + |k|^2)^{-N},$$

where for $N > d/2$ the sum on the right converges.

$\Lambda_\varepsilon$**,** $\varepsilon \in \{1, \ldots, d\}$**:** We only deal with the special case $\varepsilon = d$, since the other cases can be transformed to this case via rotations. Let $\mu \in \Delta$ and write $t_\mu = \sigma^{-j'}$ with $j' \in \mathbb{R}$. In view of $t_\lambda = \sigma^{-j}$ for $\lambda = (d, j, \ell, k) \in \Lambda_d$ we then have

$$\sum_{\lambda\in\Lambda_d} \omega_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu))^{-N} = \sum_{j\in\mathbb{N}_0} \sigma^{-N|j-j'|} \sum_{\substack{\lambda\in\Lambda_d\\ t_\lambda=\sigma^{-j}}} \left(1 + d_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu))\right)^{-N}.$$

If we can prove that

$$\mathcal{S} := \sum_{\substack{\lambda\in\Lambda_d\\ t_\lambda=\sigma^{-j}}} \left(1 + d_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu))\right)^{-N} \lesssim \sigma^{d|j-j'|}, \tag{7.14}$$

independently of $j \in \mathbb{N}_0$ and $\mu \in \Delta$, we are done, since $\sigma > 1$, $t_\mu = \sigma^{-j'}$, $\max\{t_\lambda/t_\mu, t_\mu/t_\lambda\} = \sigma^{|j-j'|}$ and thus if $N > d$

$$\sum_{\lambda\in\Lambda_d} \omega_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu))^{-N} \lesssim \sum_{j\in\mathbb{N}_0} \sigma^{(d-N)|j-j'|} \leq 2\sum_{j\in\mathbb{N}_0} \sigma^{(d-N)j} = \frac{2}{1 - \sigma^{d-N}} < \infty.$$

Putting in the definition of $d_\alpha$ and abbreviating $j_0 := \min\{j, j'\}$, the sum $\mathcal{S}$ becomes

$$\mathcal{S} = \sum_{\substack{\lambda\in\Lambda_d\\ t_\lambda=\sigma^{-j}}} \Big(1 + \sigma^{2\alpha j_0}|x_\lambda - x_\mu|^2 + \sigma^{2(1-\alpha)j_0}|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2 + \sigma^{j_0}|\langle e_\lambda, x_\lambda - x_\mu\rangle|\Big)^{-N}. \tag{7.15}$$

In order to prove the estimate (7.14) for $\mathcal{S}$, we first study the different terms of the summand independently. Let $\lambda = (d, j, \ell, k) \in \Lambda_d$ and recall the matrix $M_\lambda$ from (7.10). It holds

$$M_\lambda^T = R_{\theta_\lambda} R_{\varphi_\lambda} S_{\ell\eta_j}^{-1},$$

and – according to the discussion of $M_\lambda$ in the proof of Theorem 7.4.4 – its last row is given by $(0, \ldots, 0, n_\lambda)$ with $n_\lambda = (1 + \eta_j^2|\ell|_2^2)^{-\frac{1}{2}}$. Since $\eta_j \asymp \sigma^{-j(1-\alpha)}$ and $|\ell|_2 \lesssim \mathbf{L}_j \lesssim \sigma^{j(1-\alpha)}$ this implies $n_\lambda \asymp 1$ uniformly for all $\lambda \in \Lambda_d$.

As a direct consequence $[M_\lambda^T x]_d = n_\lambda[x]_d \asymp [x]_d$ uniformly for $\lambda \in \Lambda_d$ and $x \in \mathbb{R}^d$. In addition, we have $|M_\lambda^T x| \asymp |x|$ uniformly for $\lambda \in \Lambda_d$ and $x \in \mathbb{R}^d$ since $\|M_\lambda^T\|_{2\to2} = \|M_\lambda\|_{2\to2} \lesssim 1$ and also $\|M_\lambda^{-T}\|_{2\to2} = \|M_\lambda^{-1}\|_{2\to2} \lesssim 1$.

These observations allow the following estimate,

$$\begin{aligned}|x_\lambda - x_\mu| &= |S_{\ell\eta_j}^{-1} A_{\alpha,\sigma}^{-j}\mathcal{T}k - x_\mu| \asymp |A_{\alpha,\sigma}^{-j}\mathcal{T}k - S_{\ell\eta_j}x_\mu| \asymp |A_{\alpha,\sigma}^{-j}k - \mathcal{T}^{-1}S_{\ell\eta_j}x_\mu| \\ &\gtrsim |A_{\alpha,\sigma}^{-j}k - \mathcal{T}^{-1}S_{\ell\eta_j}x_\mu|_{[d-1]} = |\sigma^{-j\alpha}k - \mathcal{T}^{-1}S_{\ell\eta_j}x_\mu|_{[d-1]}. \end{aligned} \tag{7.16}$$

In view of $e_\lambda = R_{\varphi_\lambda}^T R_{\theta_\lambda}^T e_d$ and $S_{\ell\eta_j} = M_\lambda^{-T} R_{\theta_\lambda} R_{\varphi_\lambda}$ we also have the estimate

$$\begin{aligned}|\langle e_\lambda, x_\lambda - x_\mu\rangle| &= |\langle e_\lambda, S_{\ell\eta_j}^{-1} A_{\alpha,\sigma}^{-j}\mathcal{T}k - x_\mu\rangle| = |\langle e_d, R_{\theta_\lambda}R_{\varphi_\lambda}S_{\ell\eta_j}^{-1} A_{\alpha,\sigma}^{-j}\mathcal{T}k - R_{\theta_\lambda}R_{\varphi_\lambda}x_\mu\rangle| \\ &= |\langle e_d, M_\lambda^T(\mathcal{T}A_{\alpha,\sigma}^{-j}k - M_\lambda^{-T}R_{\theta_\lambda}R_{\varphi_\lambda}x_\mu)\rangle| \asymp |\langle e_d, A_{\alpha,\sigma}^{-j}k - \mathcal{T}^{-1}S_{\ell\eta_j}x_\mu\rangle| \\ &= |\langle e_d, \sigma^{-j}k - \mathcal{T}^{-1}S_{\ell\eta_j}x_\mu\rangle|. \end{aligned} \tag{7.17}$$

According to the $\alpha$-shearlet parametrization (7.9) we have $e_\lambda = n_\lambda(\ell\eta_j, 1)^T$, where $n_\lambda \asymp 1$ as shown above. Hence, there is a constant $c > 0$ such that $n_\lambda \geq c$ for all $\lambda \in \Lambda_d$. It follows $[e_\lambda]_d \geq c > 0$ for all $\lambda \in \Delta_d$. Without loss of generality we can further assume that $[e_\mu]_d \geq 0$ since $|\{d_{\mathbb{S}}(e_\lambda, -e_\mu)\}| = |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|$. In this situation Lemma 7.4.6 applies and tells us that $|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}| \asymp |e_\lambda - e_\mu|$.

Moreover, if $|[e_\mu]_d| \geq c/2$ we obtain with Lemma 7.4.5

$$|e_\lambda - e_\mu| \asymp |\phi(e_\lambda) - \phi(e_\mu)| = |(\ell\eta_j, 1)^T - \phi(e_\mu)|,$$

where $\phi$ denotes the gnomonic projection. In this case, we define $\nu_\mu \in \mathbb{R}^{d-1}$ by $\phi(e_\mu) =: (\nu_\mu, 1)^T$. Then

$$|e_\lambda - e_\mu| \asymp |(\ell\eta_j, 1)^T - (\nu_\mu, 1)^T| = |(\ell\eta_j)^T - (\nu_\mu)^T|.$$

If $|[e_\mu]_d| < c/2$ we put $\nu_\mu := 0$. Then, since $[e_\lambda]_d \geq c$,

$$|e_\lambda - e_\mu| \geq |[e_\lambda]_d - [e_\mu]_d| > c/2 \gtrsim |(\ell\eta_j)^T| = |(\ell\eta_j)^T - (\nu_\mu)^T|.$$

Altogether, we arrive at

$$|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}| \gtrsim |\ell\eta_j - \nu_\mu|. \tag{7.18}$$

We now use (7.16), (7.17) and (7.18) to estimate the sum $\mathcal{S}$ in (7.15). Introducing the quantities $q_1(\ell) := \sigma^{j\alpha}\mathcal{T}^{-1}S_{\ell\eta_j}x_\mu$, $q_2(\ell) := \sigma^j\mathcal{T}^{-1}S_{\ell\eta_j}x_\mu$, and $q_3 := \eta_j^{-1}\nu_\mu$, and taking into account $\eta_j \asymp \sigma^{-(1-\alpha)j}$ we obtain with $p := j_0 - j \leq 0$

$$\mathcal{S} \lesssim \sum_{k\in\mathbb{Z}^d}\sum_{\ell\in\mathscr{L}_{d,j}} \left(1 + \sigma^{2\alpha p}|k - q_1(\ell)|_{[d-1]}^2 + \sigma^p|\langle e_d, k - q_2(\ell)\rangle| + \sigma^{2(1-\alpha)p}|\ell - q_3|^2\right)^{-N}.$$

The term $\sigma^{pd}\mathcal{S}$ is thus – up to a multiplicative constant – bounded by

$$\begin{aligned}&\sum_{\ell\in\mathbb{Z}^{d-1}} \sigma^{p(1-\alpha)(d-1)} \sum_{k\in\mathbb{Z}^d} \sigma^{p\alpha(d-1)}\sigma^p \\ &\qquad \cdot \left(1 + \sigma^{2\alpha p}|k - q_1(\ell)|_{[d-1]}^2 + \sigma^p|\langle e_d, k - q_2(\ell)\rangle| + \sigma^{2(1-\alpha)p}|\ell - q_3|^2\right)^{-N}.\end{aligned}$$

The last sum can be interpreted as a Riemann sum, which is bounded – up to a multiplicative constant independent of $p \leq 0$ – by the corresponding integral

$$\int_{y\in\mathbb{R}^{d-1}}\int_{x\in\mathbb{R}^d} \left(1 + |x - \sigma^{\alpha p}q_1(y)|_{[d-1]}^2 + |\langle e_d, x - \sigma^p q_2(y)\rangle| + |y - \sigma^{(1-\alpha)p}q_3|^2\right)^{-N} dx\, dy.$$

All in all we end up with

$$\mathcal{S} \lesssim \sigma^{d(j-j_0)} \int_{y\in\mathbb{R}^{d-1}} \int_{x\in\mathbb{R}^d} \Big(1 + |x|^2_{[d-1]} + |\langle e_d, x\rangle| + |y|^2\Big)^{-N} \, dx\, dy.$$

To see that the integral converges for $N > d$, we carry out the integration over $x_d$, which yields up to a fixed constant

$$\int_{y\in\mathbb{R}^{d-1}} \int_{\tilde{x}\in\mathbb{R}^{d-1}} (1 + |\tilde{x}|^2_{[d-1]} + |y|^2)^{-(N-1)} \, d\tilde{x}\, dy = \int_{z\in\mathbb{R}^{2(d-1)}} (1+|z|^2)^{-(N-1)}.$$

The integral on the right converges precisely for $N > d$. This observation concludes the proof. □

## 7.5 Application: *Sparse Approximation of Video Data*

In this section, we will demonstrate with a specific example how the machinery of $d$-variate $\alpha$-molecules can be applied in practice. In our exemplary application, we are interested in the sparse approximation of video signals modelled by the class of cartoon-like functions $\mathcal{E}^2([0,1]^3,\nu)$ introduced below.

Following the general methodology of the transfer principle, we just need to find a suitable anchor system for which a sparse approximation result with respect to $\mathcal{E}^2([0,1]^3,\nu)$ is known. Utilizing Theorem 7.3.2, the framework can then transfer the approximation rate from this anchor system to other systems. In this way, we will identify a large class of representation systems providing almost optimal sparse approximation for $\mathcal{E}^2([0,1]^3,\nu)$.

### 7.5.1 Cartoon-like Functions in 3D

A suitable model for video data is provided by a 3-dimensional version of the original class of cartoon-like functions introduced by Donoho in [38], namely the following model defined in [83].

**Definition 7.5.1** ([38],[83, Def 2.1])**.** For fixed $\nu > 0$, the *class* $\mathcal{E}^2([0,1]^3,\nu)$ *of cartoon-like functions* consists of functions $f : \mathbb{R}^3 \to \mathbb{C}$ of the form

$$f = f_0 + f_1\mathcal{X}_B,$$

where $B \subset [0,1]^3$ and $f_i \in C^2(\mathbb{R}^3)$ with $\operatorname{supp} f_i \subset [0,1]^3$ and $\|f_i\|_{C^2} \le \nu$ for each $i = 0, 1$. Further, the discontinuity $\partial B$ shall be a closed $C^2$-surface with principal curvatures bounded by $\nu$.

This model is justified by the observation that real-life video data, just like real-life image data, typically consists of smooth regions, separated by piecewise smooth boundaries. Note however that for simplicity we restrict to cartoon-like functions with smooth boundaries in Definition 7.5.1.

### 7.5.2 Pyramid-adapted Shearlet Systems in 3D

In the sequel, we will present some concrete shearlet systems in 3 dimensions which are already on the market. Thereby we restrict our attention to parabolically scaled systems.

In the classic sense [79], a shearlet system in $L_2(\mathbb{R}^3)$ refers to a collection of functions of the form

$$\Big\{\psi_{j,\ell,k} = 2^j\psi(S_\ell A^j_{\frac{1}{2},2}\cdot -k) \ : \ j\in\mathbb{Z}, \ell\in\mathbb{Z}^2, k\in\mathbb{Z}^3\Big\}, \tag{7.19}$$

where $\psi \in L_2(\mathbb{R}^3)$ is some suitable generator. The classic choice for $\psi$ is furthermore a function defined on the frequency domain by

$$\hat{\psi}(\xi) = w(\xi_3)v(\tfrac{\xi_1}{\xi_3})v(\tfrac{\xi_2}{\xi_3}), \quad \xi = (\xi_1,\xi_2,\xi_3)^T \in \mathbb{R}^3,$$

where $v \in C_c^\infty(\mathbb{R})$ is a bump function and $w \in C_c^\infty(\mathbb{R})$ is the Fourier transform of a suitable univariate discrete wavelet. It was shown in [79] that it is possible to choose $v$ and $w$ so that (7.19) becomes a Parseval frame for $L_2(\mathbb{R}^3)$.

Unfortunately, the shearlet system (7.19) is directionally biased due to the fact that for large shearings the frequency support of the shearlets becomes more and more elongated along the $(e_1,e_2)$-plane. This bias has a negative effect on the approximation properties and makes the system (7.19) impractical for most applications.

To avoid this problem, the Fourier domain is partitioned into three pyramidal regions similar to the two cones in 2 dimensions,

$$\begin{aligned}
\mathcal{P}_1 &= \Big\{(\xi_1,\xi_2,\xi_3)\in\mathbb{R}^3 \ : \ |\tfrac{\xi_2}{\xi_1}|\le 1, |\tfrac{\xi_3}{\xi_1}|\le 1\Big\},\\
\mathcal{P}_2 &= \Big\{(\xi_1,\xi_2,\xi_3)\in\mathbb{R}^3 \ : \ |\tfrac{\xi_1}{\xi_2}|\le 1, |\tfrac{\xi_3}{\xi_2}|\le 1\Big\},\\
\mathcal{P}_3 &= \Big\{(\xi_1,\xi_2,\xi_3)\in\mathbb{R}^3 \ : \ |\tfrac{\xi_1}{\xi_3}|\le 1, |\tfrac{\xi_2}{\xi_3}|\le 1\Big\}.
\end{aligned}$$

Then, for each pyramid a separate shearlet system can be used and, since each system now only has to cover one pyramid, large shears are avoided. To take care of low frequencies, as in the 2-dimensional case, it is common to use distinguished coarse-scale elements with frequencies in a centered box around the origin. Subsequently, it will be the cube

$$\mathcal{R} = \Big\{\xi\in\mathbb{R}^3 \ : \ |\xi|_\infty \le \tfrac{1}{8}\Big\}.$$

Note that this cube together with the truncated pyramids $\tilde{\mathcal{P}}_1 := \mathcal{P}_1\backslash\mathcal{R}$, $\tilde{\mathcal{P}}_2 := \mathcal{P}_2\backslash\mathcal{R}$, and $\tilde{\mathcal{P}}_3 := \mathcal{P}_3\backslash\mathcal{R}$ partitions the Fourier domain into 4 distinct regions.

With each of these regions, different operators are associated. The coarse-scale functions are only translated, in the other regions we also scale and shear. The scaling and shearing operators associated with the respective regions $\varepsilon\in\{1,2,3\}$ are given by $A^{(\varepsilon)}_{\frac{1}{2},t} = Z^\varepsilon A_{\frac{1}{2},t}Z^{-\varepsilon}$ and $S^{(\varepsilon)}_h = Z^\varepsilon S_h Z^{-\varepsilon}$, and take the concrete form

$$A^{(1)}_{\frac{1}{2},t} = \begin{pmatrix} t & 0 & 0\\ 0 & t^{\frac{1}{2}} & 0\\ 0 & 0 & t^{\frac{1}{2}}\end{pmatrix}, \quad A^{(2)}_{\frac{1}{2},t} = \begin{pmatrix} t^{\frac{1}{2}} & 0 & 0\\ 0 & t & 0\\ 0 & 0 & t^{\frac{1}{2}}\end{pmatrix}, \quad A^{(3)}_{\frac{1}{2},t} = A_{\alpha,s} = \begin{pmatrix} t^{\frac{1}{2}} & 0 & 0\\ 0 & t^{\frac{1}{2}} & 0\\ 0 & 0 & t\end{pmatrix},$$

for $t > 0$, and for $h \in \mathbb{R}^2$

$$S_h^{(1)} = \begin{pmatrix} 1 & h_1 & h_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad S_h^{(2)} = \begin{pmatrix} 1 & 0 & 0 \\ h_2 & 1 & h_1 \\ 0 & 0 & 1 \end{pmatrix}, \quad S_h^{(3)} = S_h = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ h_1 & h_2 & 1 \end{pmatrix}.$$

Now we are ready to define a modified shearlet system, which is adapted to our partition of the Fourier domain and therefore called *pyramid-adapted*. This system does not exhibit the directional bias as (7.19) and can be considered as a 3D-version of the cone-adapted shearlet system from Definition 3.3.8.

**Definition 7.5.2** ([83, 82])**.** For fixed $\tau_1, \tau_2 > 0$ let $\mathcal{T} = \operatorname{diag}(\tau_1, \tau_2, \tau_2) \in \mathbb{R}^{3\times 3}$. The (affine) *pyramid-adapted 3D shearlet system* generated by the functions $\phi \in L_2(\mathbb{R}^3)$ and $\psi^\varepsilon \in L_2(\mathbb{R}^3)$, $\varepsilon \in \{1, 2, 3\}$, is defined as the union

$$SH(\phi, \psi^1, \psi^2, \psi^3; \tau_1, \tau_2) := \Phi(\phi; \tau_1) \cup \Psi_1(\psi^1; \tau_1, \tau_2) \cup \Psi_2(\psi^2; \tau_1, \tau_2) \cup \Psi_3(\psi^3; \tau_1, \tau_2) \quad (7.20)$$

of the coarse-scale functions $\Phi(\phi; \tau_1) := \{\phi_k = \phi(\cdot - \tau_1 k) : k \in \mathbb{Z}^3\}$ and the functions

$$\Psi_\varepsilon(\psi^\varepsilon; \tau_1, \tau_2) := \Big\{ \psi^\varepsilon_{j,\ell,k} \ : \ j \in \mathbb{N}_0, \ell \in \mathbb{Z}^2, |\ell|_\infty \le \lceil 2^{j/2} \rceil, k \in \mathbb{Z}^3 \Big\},$$

associated with the pyramids $\tilde{\mathcal{P}}_\varepsilon$ for $\varepsilon \in \{1, 2, 3\}$, which are given by

$$\psi^\varepsilon_{j,\ell,k} := 2^j \psi^\varepsilon(Z^\varepsilon S_\ell A^j_{\frac{1}{2},2} Z^{-\varepsilon} \cdot - Z^\varepsilon \mathcal{T} Z^{-\varepsilon} k).$$

These pyramid-adapted affine systems are the prime examples of $\frac{1}{2}$-shearlet-molecules. In practice, one usually wants them to be frames, or even Parseval or tight frames. However, ensuring the frame property of pyramid-adapted shearlets is not trivial.

### Frames of Pyramid-adapted Shearlets

The simplest way to obtain a Parseval frame of pyramid-adapted shearlets builds upon a Parseval shearlet frame of the type (7.19), which is easier to construct. A shearlet system associated with the pyramid $\tilde{\mathcal{P}}_3$ is then obtained by removing all elements, whose frequency support does not intersect $\tilde{\mathcal{P}}_3$. Truncating the remaining functions in the frequency domain outside of $\tilde{\mathcal{P}}_3$, one obtains a Parseval frame for the space

$$L_2(\tilde{\mathcal{P}}_3)^\vee := \{ f \in L_2(\mathbb{R}^3) \ : \ \operatorname{supp} \hat{f} \subset \tilde{\mathcal{P}}_3 \}.$$

A similar procedure yields Parseval frames associated with the the other parts of the Fourier domain, namely for $L_2(\tilde{\mathcal{P}}_\varepsilon)^\vee$, $\varepsilon \in \{1, 2\}$, and $L_2(\mathcal{R})^\vee$. The union of these frames then is a Parseval frame for the whole space $L_2(\mathbb{R}^3)$.

This approach has the drawback that it leads to bad spatial localization of the shearlets due to their lack of smoothness in the frequency domain, which is a consequence of the truncation. A different approach was taken by Candès, Demanet, and Ying in [115]. They gave up on the affine structure of the system and could then find a shearlet-type construction with the Parseval property. Guo and Labate later modified this approach [69, 70] and found another shearlet-type construction, which is even close to affine.

We will subsequently denote this system by $SH$. It is a Parseval frame of band-limited shearlets for $L_2(\mathbb{R}^3)$. Moreover, it is a system of $\frac{1}{2}$-molecules.

**Proposition 7.5.3** ([45, Prop. 5.9]). *Appropriately re-indexed, the smooth Parseval frame of band-limited* $3D$*-shearlets* $SH$ *constructed in [69] constitutes a system of* $3$*-dimensional* $\frac{1}{2}$*-shearlet molecules of order* $(\infty, \infty, \infty, \infty)$.

In particular, $SH$ is a system of 3-dimensional $\frac{1}{2}$-molecules of order $(\infty, \infty, \infty, \infty)$. The associated parametrization $(\Lambda_{SH}, \Phi_{SH})$ is given explicitly in [45, Cor. 5.10]. It is related to a $\frac{1}{2}$-shearlet parametrization by a mere relabelling of the shearlets.

**Remark 7.5.4** ([45, Rem. 5.11]). Although $(\Lambda_{SH}, \Phi_{SH})$ is not a shearlet parametrization, it is $(\frac{1}{2}, k)$-consistent with every $\frac{1}{2}$-shearlet parametrization for $k > 3$. This follows from Proposition 7.4.7 and the observation that relabelling of elements does not make any difference.

There also exist shearlet frames for $L_2(\mathbb{R}^3)$ consisting of compactly supported functions. Such frames have been constructed for example in [83]. As the following proposition shows, they are also instances of $\frac{1}{2}$-shearlet molecules and their order can be controlled by the regularity of the generators.

**Proposition 7.5.5** ([45, Prop. 5.12]). *Let* $\phi, \psi^1, \psi^2, \psi^3 \in L_2(\mathbb{R}^3)$ *be compactly supported and* $L, M, N_1, N_2 \in \mathbb{N}_0 \cup \{\infty\}$. *If* $\phi \in C^{N_1+N_2}(\mathbb{R}^3)$ *and if, for every* $\varepsilon \in \{1, 2, 3\}$,

*(i) the derivatives* $\partial^\gamma \psi^\varepsilon$ *exist and are continuous for every* $\gamma \in \mathbb{N}_0^3$ *with* $[Z^\varepsilon \gamma]_1, [Z^\varepsilon \gamma]_2 \leq N_1 + N_2$ *and* $[Z^\varepsilon \gamma]_3 \leq N_1$, *where* $Z$ *is the cyclic permutation matrix* (7.6),

*(ii) the generator* $\psi^\varepsilon$ *has* $M + L$ *directional vanishing moments in* $e_\varepsilon$*-direction, i.e.*

$$\forall (x_1, x_2) \in \mathbb{R}^2 : \int_{\mathbb{R}} \psi^\varepsilon(Z^\varepsilon x) x_3^N \, dx_3 = 0 \quad \textit{for every } N \in \{0, \ldots, M + L - 1\},$$

*then the system* (7.20) *obtained from these generators is a system of* $\frac{1}{2}$*-shearlet molecules of order* $(L, M, N_1, N_2)$.

*Proof.* Due to the assumptions, the generators are functions in $C_c(\mathbb{R}^3)$ and hence in particular contained in $L_1(\mathbb{R}^3)$. As a consequence, their Fourier transforms $\hat{\phi}, \hat{\psi}^1, \hat{\psi}^2, \hat{\psi}^3$ are bounded. Hence, rightly indexed, the induced system of the form (7.20) constitutes a system of $\frac{1}{2}$-shearlet molecules. It remains to verify the order of the system. For this, little more is needed than utilizing the facts that spatial decay implies smoothness in Fourier domain (and vice versa), and that vanishing moments in spatial domain implies estimates of the form $|\hat{g}(\xi)| \lesssim \min(1, |\xi|)^M$ in Fourier domain. We refer to [62, Prop. 3.11] for details, where a similar two-dimensional version of the theorem is proven. □

### 7.5.3 Quasi-Optimal Approximation with 3D Parabolic Molecules

In [83] the optimal approximation rate under a *polynomial depth search constraint* for $\mathcal{E}^2([0,1]^3, \nu)$ was derived. Recall that a dictionary-based algorithm for sparse approximation is said to satisfy a polynomial depth search constraint if there exists a polynomial $\pi$ such that the algorithm only chooses from the first $\pi(N)$ dictionary elements when forming the $N$:th sparse approximation [38]. If such a constraint is not assumed, one could for instance use the whole of $\mathcal{E}^2([0,1]^3, \nu)$ as a dictionary yielding 1-sparse representations for any element of $\mathcal{E}^2([0,1]^3, \nu)$. But clearly, those have no practical relevance for real-world approximation schemes.

The following benchmark for $\mathcal{E}^2([0,1]^3, \nu)$ was proved in [83] with the same techniques used in Section 5.3.

**Theorem 7.5.6** ([38, Thm 7.2],[83, Thm 3.2]). *The best $N$-term approximation rate for $\mathcal{E}^2([0,1]^3, \nu)$, achieved by an arbitrary dictionary under the restriction of polynomial depth search, cannot exceed*

$$\|f - f_N\|_2^2 \lesssim N^{-1},$$

*where $f_N$ is the best $N$-term approximation of $f \in \mathcal{E}^2([0,1]^3, \nu)$.*

There are several examples of frames which almost provide these optimal rates [69, 83], typically up to log-terms. In particular, it was proven by Guo and Labate [69] that the smooth Parseval frame of 3D-shearlets $SH$ sparsely approximates this class. The obtained approximation result is stated below in (7.21). It is based on the following estimate for the size of the shearlet coefficients.

**Theorem 7.5.7** ([69, Thm 3.1]). *Let $SH = \{\psi_\lambda\}_{\lambda \in \Lambda^s}$ be the smooth Parseval frame of 3D-shearlets defined in [70]. Then the sequence of shearlet coefficients $\theta_\lambda(f) := \langle f, \psi_\lambda \rangle$, $\lambda \in \Lambda^s$, associated with $f \in \mathcal{E}^2([0,1]^3, \nu)$ satisfies*

$$\sup_{f \in \mathcal{E}^2([0,1]^3, \nu)} |\theta_\lambda(f)|_N \lesssim N^{-1} \cdot \log(N),$$

*where $|\theta_\lambda(f)|_N$ denotes the $N$:th largest shearlet coefficient.*

Theorem 7.5.7 shows that the shearlet coefficients belong to $\omega\ell^p(\Lambda^s)$ for every $p > 1$. In view of Lemma 2.3.1, for every $f \in \mathcal{E}^2([0,1]^3, \nu)$, the frame $SH$ therefore provides at least the approximation rate

$$\|f - f_N\|_2^2 \lesssim N^{-1+\varepsilon} \quad , \varepsilon > 0 \text{ arbitrary}, \tag{7.21}$$

where $f_N$ denotes the $N$-term approximation obtained from the $N$ largest coefficients. According to Theorem 7.5.6, this is almost the optimal approximation rate achievable for cartoon-like functions $\mathcal{E}^2([0,1]^3, \nu)$. For small $\varepsilon > 0$, we get arbitrarily close to the optimal rate.

### Transfer of the Approximation Rate

We now come to our final goal of identifying a large class of representation systems which achieve the almost optimal rate (7.21) for the class $\mathcal{E}^2([0,1]^3, \nu)$. For this, we put the machinery of $\alpha$-molecules to work. Concretely, we use Theorem 7.3.2 to transfer the approximation rate (7.21) of the smooth Parseval frame of 3D-shearlets $SH$ to other systems of 3-dimensional $\frac{1}{2}$-molecules. This leads to the following result, whereby $(\Lambda_{SH}, \Phi_{SH})$ shall denote the parametrization of $SH$.

**Theorem 7.5.8** ([45, Thm. 4.4]). *Assume that a frame $\{m_\lambda\}_{\lambda \in \Lambda}$ of 3-dimensional parabolic molecules satisfies, for some $k > 0$, the following two conditions:*

*(i) its parametrization $(\Lambda, \Phi_\Lambda)$ is $(\frac{1}{2}, k)$-consistent with $(\Lambda_{SH}, \Phi_{SH})$,*

*(ii) its order* $(L, M, N_1, N_2)$ *satisfies*

$$L \geq 2k, \quad M \geq 3k-2, \quad N_1 \geq k+1, \quad N_1 > 3/2, \quad N_2 \geq 2k+1.$$

*Then each dual frame* $\{\tilde{m}_\lambda\}_{\lambda\in\Lambda}$ *possesses an almost optimal* $N$*-term approximation rate for the class of cartoon-like functions* $\mathcal{E}^2([0,1]^3,\nu)$*, i.e., for all* $f \in \mathcal{E}^2([0,1]^3,\nu)$

$$\|f - f_N\|_2^2 \lesssim N^{-1+\varepsilon}, \quad \varepsilon > 0 \text{ arbitrary},$$

*where* $f_N$ *denotes the* $N$*-term approximation obtained from the* $N$ *largest frame coefficients.*

*Proof.* The proof is analogous to the proof of Theorem 6.0.2. Due to Proposition 7.5.3, the frame $SH$ is a system of 3-dimensional $\frac{1}{2}$-molecules of order $(\infty,\infty,\infty,\infty)$. It is thus a suitable reference system for the application of the transfer principle, Theorem 7.3.2. The assertion then follows from Theorem 7.5.7. □

Theorem 7.5.8 specifies a large class of multiscale systems with almost optimal approximation performance for video data in the class $\mathcal{E}^2([0,1]^3,\nu)$. According to Remark 7.5.4, condition (i) is in particular fulfilled by every $\frac{1}{2}$-shearlet parametrization (see Section 7.4) for $k > 3$. Hence, due to condition (ii), all systems of 3-dimensional $\frac{1}{2}$-shearlet molecules of order

$$L \geq 7, \quad M \geq 8, \quad N_1 \geq 5, \quad N_2 \geq 8,$$

provide almost optimal approximation for $\mathcal{E}^2([0,1]^3,\nu)$.

Taking into account Proposition 7.5.5, the statement of Theorem 7.5.8 in particular includes the following result for compactly supported shearlet frames.

**Corollary 7.5.9** ([45, Cor. 4.5])**.** *Any dual frame of a shearlet frame of the form* (7.20) *generated by compactly supported functions* $\phi, \psi^1, \psi^2, \psi^3 \in L_2(\mathbb{R}^3)$*, so that* $\phi \in C^{13}(\mathbb{R}^3)$ *and for each permutation* $(i,j,k)$ *of* $(1,2,3)$*, we have*

*(i)* $\partial^\gamma \psi^i$ *exists and is continuous for every* $\gamma \in \mathbb{N}_0^3$ *with* $\gamma_i \leq 5$ *and* $\gamma_j, \gamma_k \leq 13$*,*

*(ii)* $\psi^i$ *has at least 15 vanishing directional moments in direction* $e_i$*,*

*provides the almost optimal approximation rate* (7.21) *for the cartoon class* $\mathcal{E}^2([0,1]^3,\nu)$*.*

A result similar to this corollary was proved in [83]. In comparison, the most intriguing fact about this corollary is the simplicity of its deduction. The framework of $\alpha$-molecules enables a simple transfer of the decay rates.

## 7.6 Appendix: *Proof of Theorem 7.2.2*

This section is devoted to the proof of Theorem 7.2.2. The exposition is essentially the same as in [45, Sec. 6]. It is split into several parts and has the same general structure as the proof of Theorem 2.2.2 in Section 2.5. Let us first collect some simple elementary facts, which turn out to be useful.

### 7.6.1 Auxiliary Lemmas

Subsequently $\mathcal{O}(d, \mathbb{R})$ shall denote the orthogonal group of $\mathbb{R}^d$. Further, recall the 'projection' $\{\theta\}$ of $\theta \in \mathbb{R}$ onto the interval $\mathbf{T} := [-\frac{\pi}{2}, \frac{\pi}{2})$ defined in (2.9). Recall also the notation $d_{\mathbb{S}}(v, w)$ for the angle $\arccos(\langle v, w\rangle) \in [0, \pi]$ between two vectors $v, w \in \mathbb{S}^{d-1}$.

An immediate corollary of Lemma 2.5.1 is the following result.

**Lemma 7.6.1** ([45, Lem. 6.2])**.** *Let $e_d \in \mathbb{R}^d$ be the d:th unit vector. For $\eta \in \mathbb{S}^{d-1}$ we have* $|\,\{d_{\mathbb{S}}(\eta, e_d)\}\,| \asymp |\eta|_{[d-1]}$.

*Proof.* Using a suitable rotation $R \in \mathcal{O}(d, \mathbb{R})$ of the form

$$R = \begin{pmatrix} R_{d-1} & 0 \\ 0 & 1 \end{pmatrix},$$

where $R_{d-1} \in \mathcal{O}(d-1, \mathbb{R})$, we can achieve $R\eta = (\sin(\theta), 0, \ldots, 0, \cos(\theta))^T$ with $\theta = d_{\mathbb{S}}(\eta, e_d)$. Since $|\eta|_{[d-1]} = |\eta - e_d|_{[d-1]} = |R(\eta - e_d)|_{[d-1]} = |R\eta - e_d|_{[d-1]} = |\sin(\theta)|$, it just remains to prove $|\sin(\theta)| \asymp |\,\{\theta\}\,|$, which is true by Lemma 2.5.1. $\square$

We will further need the lemma below.

**Lemma 7.6.2** ([45, Lem. 6.4])**.** *Let $R \in \mathcal{O}(d, \mathbb{R})$ be a rotation and $\theta_0 = d_{\mathbb{S}}(e_d, Re_d) \in [0, \pi]$ the angle between the d:th unit vector $e_d \in \mathbb{R}^d$ and its image $Re_d$ under $R$. Then it holds for all $\eta \in \mathbb{S}^{d-1}$*

$$|R\eta|_{[d-1]} = \sin(d_{\mathbb{S}}(R\eta, e_d)) \geq \min\{|\sin(d_{\mathbb{S}}(\eta, e_d) + \theta_0)|, |\sin(d_{\mathbb{S}}(\eta, e_d) - \theta_0)|\}.$$

Note $d_{\mathbb{S}}(\eta, e_d) = d_{\mathbb{S}}(R\eta, Re_d)$.

*Proof.* Let $\eta = (\eta_1, \ldots, \eta_d)^T \in \mathbb{S}^{d-1}$ and put $\theta_1 := d_{\mathbb{S}}(\eta, e_d) = \arccos(\langle \eta, e_d\rangle) \in [0, \pi]$. The rotation $R \in \mathcal{O}(d, \mathbb{R})$ can be decomposed in the form $R = \tilde{R}R_{\theta_0}$ with $\tilde{R}, R_{\theta_0} \in \mathcal{O}(d, \mathbb{R})$ such that

$$\tilde{R} = \begin{pmatrix} R_{d-1} & \\ & 1 \end{pmatrix} \quad \text{and} \quad R_{\theta_0} = \begin{pmatrix} \cos(\theta_0) & & -\sin(\theta_0) \\ & I_{d-2} & \\ \sin(\theta_0) & & \cos(\theta_0) \end{pmatrix},$$

where $R_{d-1} \in \mathcal{O}(d-1, \mathbb{R})$ is some $(d-1)$-dimensional rotation matrix and $I_{d-2}$ is the $(d-2)$-dimensional identity matrix. The rotation $\tilde{R}$ leaves $|\cdot|_{[d-1]}$ invariant, whence

$$|R\eta|_{[d-1]} = |\tilde{R}R_{\theta_0}\eta|_{[d-1]} = |R_{\theta_0}\eta|_{[d-1]}.$$

Using $\eta_d = \cos(\theta_1)$ and $|\eta|^2_{[d-1]} = \eta_1^2 + \eta_2^2 + \ldots + \eta_{d-1}^2 = 1 - \eta_d^2$, it further follows

$$\begin{aligned} |R_{\theta_0}\eta|^2_{[d-1]} &= (\cos(\theta_0)\eta_1 - \sin(\theta_0)\eta_d)^2 + \eta_2^2 + \cdots + \eta_{d-1}^2 \\ &= \cos^2(\theta_0)\eta_1^2 + \sin^2(\theta_0)\cos^2(\theta_1) - 2\cos(\theta_0)\sin(\theta_0)\eta_1\cos(\theta_1) + (1 - \eta_1^2 - \cos^2(\theta_1)) \\ &= 1 - (\eta_1\sin(\theta_0) + \cos(\theta_1)\cos(\theta_0))^2. \end{aligned}$$

The last expression is a second-degree polynomial in the variable $\eta_1$ with a negative leading coefficient. Since $\eta_1^2 \leq 1 - \eta_d^2 = 1 - \cos^2(\theta_1) = \sin^2(\theta_1)$, the variable $\eta_1$ can take values only in $[-\sin(\theta_1), \sin(\theta_1)]$. The polynomial attains its minimum on this interval at the endpoints. Hence, we can conclude

$$\begin{aligned} |R_{\theta_0}\eta|^2_{[d-1]} &\geq \min_{\epsilon\in\{-1,1\}} \{1 - (\epsilon \sin(\theta_1)\sin(\theta_0) + \cos(\theta_1)\cos(\theta_0))^2\} \\ &= \min_{\epsilon\in\{-1,1\}} \{1 - \cos^2(\theta_1 - \epsilon\theta_0)\} = \min_{\epsilon\in\{-1,1\}} \{\sin^2(\theta_1 - \epsilon\theta_0)\}, \end{aligned}$$

which proves the claim. □

### 7.6.2 Integral Estimates

We start with an estimate which can be used for the generators in (7.4) and allows us to work in polar coordinates.

**Lemma 7.6.3** ([45, Lem. 6.5]). *Let the family of functions $\{g_\lambda\}_{\lambda\in\Lambda}$ satisfy*

$$|\partial^\rho g_\lambda(\xi)| \lesssim \min\left\{1, t_\lambda + |[\xi]_d| + t_\lambda^{1-\alpha}|\xi|_{[d-1]}\right\}^M \langle|\xi|\rangle^{-N_1}\langle|\xi|_{[d-1]}\rangle^{-N_2} \tag{7.22}$$

*uniformly for a multi-index $\rho \in \mathbb{N}_0^d$, and assume that there is a constant $C > 0$ such that $t_\lambda \leq C$ for all $\lambda \in \Lambda$. Then the following estimate holds true uniformly for $\lambda \in \Lambda$ and $\xi \in \mathbb{R}^d$*

$$|(\partial^\rho g_\lambda)(A_{\alpha,t_\lambda}R_{\theta_\lambda}R_{\varphi_\lambda}\xi)| \lesssim \frac{\min\{1, t_\lambda(1+|\xi|)\}^M}{(1+t_\lambda|\xi|)^{N_1}\left(1+t_\lambda^\alpha|R_{\theta_\lambda}R_{\varphi_\lambda}\xi|_{[d-1]}\right)^{N_2}}. \tag{7.23}$$

*Note that* (7.22) *is just the condition* (7.5) *imposed on the Fourier side on the generating set of a system of $\alpha$-molecules.*

*Proof.* We have $|A_{\alpha,t_\lambda}\xi| \geq \min\{t_\lambda, t_\lambda^\alpha\}|\xi| \gtrsim t_\lambda|\xi|$ uniformly for $\xi \in \mathbb{R}^d$ and $\lambda \in \Lambda$, since $t_\lambda^{-1} \geq 1/C > 0$ for every $\lambda \in \Lambda$. It follows $|A_{\alpha,t_\lambda}R_{\theta_\lambda}R_{\varphi_\lambda}\xi| \gtrsim t_\lambda|R_{\theta_\lambda}R_{\varphi_\lambda}\xi| = t_\lambda|\xi|$. Further, we observe $|A_{\alpha,t_\lambda}\xi|_{[d-1]} = t_\lambda^\alpha|\xi|_{[d-1]}$ and $|[A_{\alpha,t_\lambda}\xi]_d| = t_\lambda|[\xi]_d|$. Finally, it holds $\langle|\xi|\rangle \asymp 1+|\xi|$ and $|[\xi]_d| + |\xi|_{[d-1]} \asymp |\xi|$. Collecting all of these estimates, one obtains

$$\begin{aligned} |(\partial^\rho g_\lambda)(A_{\alpha,t_\lambda}R_{\theta_\lambda}R_{\varphi_\lambda}\xi)| &\lesssim \frac{\min\left\{1, t_\lambda + |[A_{\alpha,t_\lambda}R_{\theta_\lambda}R_{\varphi_\lambda}\xi]_d| + t_\lambda^{1-\alpha}|A_{\alpha,t_\lambda}R_{\theta_\lambda}R_{\varphi_\lambda}\xi|_{[d-1]}\right\}^M}{\langle|A_{\alpha,t_\lambda}R_{\theta_\lambda}R_{\varphi_\lambda}\xi|\rangle^{N_1}\langle|A_{\alpha,t_\lambda}R_{\theta_\lambda}R_{\varphi_\lambda}\xi|_{[d-1]}\rangle^{N_2}} \\ &\lesssim \frac{\min\{1, t_\lambda(1+|\xi|)\}^M}{(1+t_\lambda|\xi|)^{N_1}\left(1+t_\lambda^\alpha|R_{\theta_\lambda}R_{\varphi_\lambda}\xi|_{[d-1]}\right)^{N_2}}. \end{aligned}$$

□

The expression on the right-hand side of (7.23) can further be estimated by the function

$$S_{\lambda,M,N_1,N_2}(\xi) := \frac{\min\{1, t_\lambda(1+|\xi|)\}^M}{(1+t_\lambda|\xi|)^{N_1}(1+t_\lambda^{-(1-\alpha)}|R_{\theta_\lambda}R_{\varphi_\lambda}(\xi/|\xi|)|_{[d-1]})^{N_2}}, \quad \xi \in \mathbb{R}^d. \tag{7.24}$$

As already discussed in Lemma 2.5.4 of Subsection 2.5.2, this function can be separated into angular and radial components, which allows to treat these parts independently in the integration later.

**Lemma 7.6.4** ([45, Lem. 6.6]). *Assume that $t_\lambda \le C$ holds for all $\lambda \in \Lambda$. For every $M, N_1, N_2, K \in \mathbb{N}_0$ such that $0 \le K \le N_2$ we have with respect to $\lambda \in \Lambda$ and $\xi \in \mathbb{R}^d$ the uniform estimate*

$$\frac{\min\{1, t_\lambda(1+|\xi|)\}^M}{(1+t_\lambda|\xi|)^{N_1}\left(1+t_\lambda^\alpha|R_{\theta_\lambda}R_{\varphi_\lambda}\xi|_{[d-1]}\right)^{N_2}} \lesssim S_{\lambda,M-K,N_1,K}(\xi).$$

*Proof.* The proof is analogous to the proof of Lemma 2.5.4. □

Next, we want to estimate the scalar product of two functions of the form (7.24). Before the actual result, Lemma 7.6.8, we need some preparation. This is the part of the proof of Theorem 7.2.2 which differs the most from the situation in two dimensions.

As a direct corollary of Lemma 2.5.2, we get the following result.

**Lemma 7.6.5** ([45, Lem. 6.7]). *Let $a \ge a' > 0$, $d \in \mathbb{N}\backslash\{1\}$, and $N > 1$. Then we have uniformly for $y \in \mathbb{R}$*

$$\int_{\mathbb{R}} \frac{|x|^{d-2}dx}{(1+a|x|)^{N+d-2}(1+a'|x-y|)^{N+d-2}} \lesssim a^{-(d-1)}(1+a'|y|)^{-N}.$$

*Proof.* Utilizing Lemma 2.5.2 ,the result from Grafakos [51][Appendix K.1],

$$\int_{\mathbb{R}} \frac{dx}{(1+a|x|)^N(1+a'|x-y|)^N} \lesssim \max\{a,a'\}^{-1}(1+\min\{a,a'\}|y|)^{-N}$$

we can estimate

$$\begin{aligned}
\int_{\mathbb{R}} \frac{|x|^{d-2}dx}{(1+a|x|)^{N+d-2}(1+a'|x-y|)^{N+d-2}} &= a^{-(d-2)}\int_{\mathbb{R}} \frac{|ax|^{d-2}dx}{(1+a|x|)^{N+d-2}(1+a'|x-y|)^{N+d-2}}\\
&\le a^{-(d-2)}\int_{\mathbb{R}} \frac{(1+|ax|)^{d-2}dx}{(1+a|x|)^{N+d-2}(1+a'|x-y|)^{N+d-2}}\\
&\le a^{-(d-2)}\int_{\mathbb{R}} \frac{dx}{(1+a|x|)^{N}(1+a'|x-y|)^{N}}\\
&\lesssim a^{-(d-2)}\max\{a,a'\}^{-1}(1+\min\{a,a'\}|y|)^{-N} = a^{-(d-1)}(1+a'|y|)^{-N}.
\end{aligned}$$

□

We can immediately deduce the following generalization of Lemma 2.5.3.

**Corollary 7.6.6** ([45, Cor. 6.8]). *Let $a \ge a' > 0$, $d \in \mathbb{N}\backslash\{1\}$, and $N > 1$. Then we have uniformly for $\theta_0 \in \mathbb{R}$*

$$\int_0^\pi \frac{|\sin^{d-2}(\theta)|\,d\theta}{(1+a|\sin(\theta)|)^{N+d-2}(1+a'|\sin(\theta-\theta_0)|)^{N+d-2}} \lesssim a^{-(d-1)}\left(1+a'|\{\theta_0\}|\right)^{-N}.$$

*Proof.* Let us call the integral to be estimated $\mathcal{S}$. Since the integrand on the left hand side is $\pi$-periodic, we may change the domain of integration to $[-\pi/2, \pi/2]$. Applying Lemma 2.5.1, we can further conclude

$$\mathcal{S} \asymp \int_{-\pi/2}^{\pi/2} \frac{|\theta|^{d-2}d\theta}{(1+a|\theta|)^{N+d-2}(1+a'|\{\theta-\theta_0\}|)^{N+d-2}},$$

Since $|\{\theta_0\}| \le \frac{\pi}{2}$ we can estimate

$$\mathcal{S} \lesssim \sum_{\vartheta\in\{-\pi,0,\pi\}} \int_{\mathbb{R}} \frac{|\theta|^{d-2} d\theta}{(1+a|\theta|)^{N+d-2}(1+a'|\theta-(\{\theta_0\}+\vartheta)|)^{N+d-2}},$$

We now use Lemma 7.6.5 to estimate this by

$$\mathcal{S} \lesssim \sum_{\vartheta\in\{-\pi,0,\pi\}} a^{-(d-1)}(1+a'|\{\theta_0\}+\vartheta|)^{-N} \lesssim a^{-(d-1)}(1+a'|\,\{\theta_0\}\,|)^{-N}.$$

□

This result is used to estimate the integral of the angular parts of (7.24) over the sphere $\mathbb{S}^{d-1}$.

**Lemma 7.6.7** ([45, Lem. 6.9])**.** *Let $a, a' > 0$, $d \in \mathbb{N}$, $d \ge 2$, $\theta_\lambda, \theta_\mu \in [0,\pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}]^{d-3}$, $\varphi_\lambda, \varphi_\mu \in [0, 2\pi]$ and $N > 1$. Further, let $d\sigma$ denote the standard surface measure on the sphere $\mathbb{S}^{d-1}$. We then have the estimate*

$$\int_{\mathbb{S}^{d-1}} \frac{d\sigma(\eta)}{(1+a|R_{\theta_\mu}R_{\varphi_\mu}\eta|_{[d-1]})^{N+d-2}(1+a'|R_{\theta_\lambda}R_{\varphi_\lambda}\eta|_{[d-1]})^{N+d-2}}$$
$$\lesssim \max\{a,a'\}^{-(d-1)}\big(1+\min\{a,a'\}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|\big)^{-N},$$

*where $e_\lambda = R^T_{\varphi_\lambda}R^T_{\theta_\lambda}e_d$ and $e_\mu = R^T_{\varphi_\mu}R^T_{\theta_\mu}e_d$.*

*Proof.* Note the symmetry of the statement with respect to interchanging the entities $a, a'$ and $\lambda, \mu$. Without loss of generality we can therefore restrict to the case $a \ge a' > 0$.

Since the mapping $R_{\theta_\mu}R_{\varphi_\mu}$ is an isometry, the integral is equal to

$$\mathcal{S} := \int_{\mathbb{S}^{d-1}} \frac{d\sigma(\eta)}{(1+a|\eta|_{[d-1]})^{N+d-2}(1+a'|R_{\theta_\lambda}R_{\varphi_\lambda}R^T_{\varphi_\mu}R^T_{\theta_\mu}\eta|_{[d-1]})^{N+d-2}}.$$

For the integration we parameterize the sphere $\mathbb{S}^{d-1}$ by standard spherical coordinates, i.e. coordinates $(\theta_1,\dots\theta_{d-2},\varphi) \in [0,\pi]^{d-2}\times[0,2\pi)$ such that for $\eta \in \mathbb{S}^{d-1}$

$$\eta(\theta,\varphi) = \begin{pmatrix} \sin(\theta_1)\cdot\dots\cdot\sin(\theta_{d-2})\cos(\varphi) \\ \sin(\theta_1)\cdot\dots\cdot\sin(\theta_{d-2})\sin(\varphi) \\ \sin(\theta_1)\cdots\sin(\theta_{d-3})\cos(\theta_{d-2}) \\ \vdots \\ \cos(\theta_1) \end{pmatrix}.$$

Observe that $\langle \eta, e_d\rangle = \cos(\theta_1)$ and thus $\theta_1 = d_{\mathbb{S}}(\eta, e_d)$. Also note $|\eta|_{[d-1]} = |\sin(\theta_1)|$. Letting $\theta_0 := d_{\mathbb{S}}(e_\lambda, e_\mu) \in [0,\pi]$ denote the angle between $e_\lambda$ and $e_\mu$ we have the equality $\theta_0 = d_{\mathbb{S}}(e_d, R_{\theta_\lambda}R_{\varphi_\lambda}R^T_{\varphi_\mu}R^T_{\theta_\mu}e_d)$. Since $R_{\theta_\lambda}R_{\varphi_\lambda}R^T_{\varphi_\mu}R^T_{\theta_\mu} \in \mathcal{O}(d,\mathbb{R})$ we can apply Lemma 7.6.2 to estimate $|R_{\theta_\lambda}R_{\varphi_\lambda}R^T_{\varphi_\mu}R^T_{\theta_\mu}\eta|_{[d-1]}$. We obtain

$$\mathcal{S} \le \int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi} \frac{\sin^{d-2}(\theta_1)\sin^{d-3}(\theta_2)\dots\sin(\theta_{d-2})d\theta_1 d\theta_2\dots d\theta_{d-2}d\varphi}{(1+a|\sin(\theta_1)|)^{N+d-2}(1+a'\min\{|\sin(\theta_1+\theta_0)|,|\sin(\theta_1-\theta_0)|\})^{N+d-2}}$$
$$\lesssim \int_0^{\pi} \frac{\sin^{d-2}(\theta_1)\,d\theta_1}{(1+a|\sin(\theta_1)|)^{N+d-2}(1+a'\min\{|\sin(\theta_1+\theta_0)|,|\sin(\theta_1-\theta_0)|\})^{N+d-2}}$$
$$\le \sum_{\epsilon\in\{-1,1\}} \int_0^{\pi} \frac{|\sin(\theta_1)|^{d-2}\,d\theta_1}{(1+a|\sin(\theta_1)|)^{N+d-2}(1+a'|\sin(\theta_1-\epsilon\theta_0)|)^{N+d-2}}.$$

Using Corollary 7.6.6 we finally arrive at $\mathcal{S} \lesssim \max\{a,a'\}^{-(d-1)}(1+\min\{a,a'\}|\{\theta_0\}|)^{-N}$. $\square$

With this estimate for the angular components in our toolbox, we proceed to prove the main result concerning the correlation of functions of the form (7.24). It corresponds to Lemma 2.5.5.

**Lemma 7.6.8** ([45, Lem. 6.10]). *Let $\alpha \in [0,1]$, $d \in \mathbb{N}\backslash\{1\}$, and $M, N_1, N_2 \in \mathbb{N}_0$. Further, let $(\Lambda, \Phi_\Lambda)$ and $(\Delta, \Phi_\Delta)$ be parametrizations with $(x_\lambda, e_\lambda, t_\lambda) = \Phi_\Lambda(\lambda)$ and $(x_\mu, e_\mu, t_\mu) = \Phi_\Delta(\mu)$ for $\lambda \in \Lambda$, $\mu \in \Delta$, such that $t_\lambda \le C$ and $t_\mu \le C$ for a fixed constant $C > 0$. Then for $A > 0$ and $B > 1$ satisfying*

$$N_1 > \frac{d}{2}, \quad M + d > N_1 \ge A + \frac{1+(d-1)\alpha}{2}, \quad \textit{and} \quad N_2 \ge B + d - 2$$

*the following estimate holds true with an implicit constant independent of $\lambda \in \Lambda$ and $\mu \in \Delta$,*

$$(t_\lambda t_\mu)^{\frac{1+(d-1)\alpha}{2}} \int_{\mathbb{R}^d} S_{\lambda,M,N_1,N_2}(x) S_{\mu,M,N_1,N_2}(x)\, dx$$
$$\lesssim \max\left\{\frac{t_\lambda}{t_\mu}, \frac{t_\mu}{t_\lambda}\right\}^{-A} \Big(1 + \max\{t_\lambda, t_\mu\}^{-(1-\alpha)} |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|\Big)^{-B}.$$

*Proof.* Without loss of generality we subsequently assume $t_\lambda \ge t_\mu$. The strategy is to separate the integration into an angular and a radial part and estimate these independently. For the estimate of the angular part we can use Lemma 7.6.7, which yields

$$(t_\lambda t_\mu)^{\frac{(1+\alpha(d-1)}{2}} \int_0^\infty \int_{\mathbb{S}^{d-1}} S_{\lambda,M,N_1,N_2}(\eta, r) S_{\mu,M,N_1,N_2}(\eta, r) r^{d-1}\, d\sigma(\eta) dr$$
$$\lesssim (t_\lambda t_\mu)^{\frac{1+\alpha(d-1)}{2}} t_\mu^{(1-\alpha)(d-1)} t_\mu^{-d} (1 + t_\lambda^{-(1-\alpha)} |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|)^{-B} \cdot \mathcal{S}$$

with a remaining radial integral

$$\mathcal{S} := t_\mu^d \int_0^\infty \frac{\min\{1, t_\lambda(1+r)\}^M}{(1+t_\lambda r)^{N_1}} \frac{\min\{1, t_\mu(1+r)\}^M}{(1+t_\mu r)^{N_1}} r^{d-1}\, dr.$$

Note that for the estimate we used the assumptions $t_\lambda \ge t_\mu$, $B > 1$ and $N_2 \ge B + d - 2$. It remains to verify the relation $(t_\mu t_\lambda)^{(1+\alpha(d-1))/2} t_\mu^{(1-\alpha)(d-1)} t_\mu^{-d} \cdot \mathcal{S} \lesssim (t_\mu/t_\lambda)^A$, or equivalently

$$\mathcal{S} \lesssim \left(\frac{t_\mu}{t_\lambda}\right)^{A + \frac{1+\alpha(d-1)}{2}}.$$

To prove this, we split the integration of $\mathcal{S}$ into three parts $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ corresponding to the integration ranges $0 \le r \le 1$, $1 \le r \le s_\mu$, and $s_\mu \le r$ respectively.

$\underline{0 \le r \le 1}$**:** Here we estimate $\min\{1, t_\lambda(1+r)\}^M \le t_\lambda^M (1+r)^M \le 2^M t_\lambda^M$ and $(1+t_\lambda r)^{N_1} \ge 1$, and similarly for the index $\mu$. Hence, the integral over this part can be estimated by

$$\mathcal{S}_1 \lesssim t_\mu^d t_\lambda^M t_\mu^M \int_0^1 r^{d-1}\, dr \asymp t_\mu^{M+d} t_\lambda^M \lesssim \left(\frac{t_\mu}{t_\lambda}\right)^{M+d},$$

where the last inequality holds because of the uniform upper bound $t_\lambda \le C$ for $\lambda \in \Lambda$. Finally observe that the assumed inequalities imply $M + d > A + \frac{1+\alpha(d-1)}{2}$.

$1 \le r \le t_\mu^{-1}$ We estimate the terms involving $\mu$ as follows: $(1+t_\mu r)^{N_1} \ge 1$ and $(r+1) \le 2r$. Hence

$$\min\{1, t_\mu(1+r)\}^M \le t_\mu^M (1+r)^M \le t_\mu^M (r+r)^M \le 2^M t_\mu^M r^M.$$

For the terms with $\lambda$'s, we have $(1+t_\lambda r)^{N_1} \ge t_\lambda^{N_1} r^{N_1}$ and $\min\{1, t_\lambda(1+r)\}^M \le 1$. The integral $\mathcal{S}_2$ hence satisfies

$$\mathcal{S}_2 \lesssim t_\mu^d t_\lambda^{-N_1} t_\mu^M \int_1^{t_\mu^{-1}} r^{M-N_1+d-1}\, dr \lesssim t_\mu^{M+d} t_\lambda^{-N_1} t_\mu^{-M-d+N_1} = \Big(\frac{t_\mu}{t_\lambda}\Big)^{N_1},$$

where it was used that $M+d > N_1$, which implies $M+d-N_1-1 > -1$, for the integration. By assumption $N_1 \ge A + \frac{1+\alpha(d-1)}{2}$, giving the desired result.

$t_\mu^{-1} \le r$ We estimate both terms like the $\lambda$-terms above to obtain

$$\mathcal{S}_3 \lesssim t_\mu^d t_\lambda^{-N_1} t_\mu^{-N_1} \int_{t_\mu^{-1}}^{\infty} r^{d-1-2N_1}\, dr \lesssim t_\mu^{-N_1+d} t_\lambda^{-N_1} t_\mu^{2N_1-d} \lesssim \Big(\frac{t_\mu}{t_\lambda}\Big)^{N_1}.$$

The integral converges since $N_1 > \frac{d}{2}$. Since $N_1 \ge A + \frac{1+\alpha(d-1)}{2}$ the proof is finished. □

### 7.6.3 Cancellation Estimates

Theorem 7.2.2 provides estimates for the scalar products of $\alpha$-molecules. To derive them we evaluate these scalar products on the Fourier side, where we can take advantage of cancellation phenomena. Technically, the method is based on a clever integration by parts involving the following differential operator, depending on $\lambda \in \Lambda$, $\mu \in \Delta$,

$$\mathscr{L}_{\lambda,\mu} := \mathcal{I} - t_0^{-2\alpha}\Delta - \frac{t_0^{-2}}{1 + t_0^{-2(1-\alpha)} |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2} \langle e_\lambda, \nabla\rangle^2, \tag{7.25}$$

where $t_0 = \max\{t_\lambda, t_\mu\}$, $\mathcal{I}$ is the identity operator, $\nabla$ the gradient and $\Delta$ the standard Laplacian.

Lemma 7.6.9 shows how $\mathscr{L}_{\lambda,\mu}$ acts on products of functions $a_\lambda$, $b_\mu$ which satisfy (7.5). It corresponds to Lemma 2.5.6.

**Lemma 7.6.9** ([45, Lem. 6.11])**.** *Let $a_\lambda$ and $b_\mu$ satisfy* (7.22) *for every multi-index $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \le L$ and assume $t_\lambda, t_\mu \le C$. Then we can write the expression*

$$\mathscr{L}_{\lambda,\mu}\Big(a_\lambda(A_{\alpha,t_\lambda} R_{\theta_\lambda} R_{\varphi_\lambda}\xi) b_\mu(A_{\alpha,t_\mu} R_{\theta_\mu} R_{\varphi_\mu}\xi)\Big)$$

*as a finite linear combination of terms of the form*

$$p_\lambda(A_{\alpha,t_\lambda} R_{\theta_\lambda} R_{\varphi_\lambda}\xi) q_\mu(A_{\alpha,t_\mu} R_{\theta_\mu} R_{\varphi_\mu}\xi)$$

*with functions $p_\lambda, q_\mu$, which satisfy* (7.22) *for all multi-indices $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \le L-2$.*

*Proof.* For convenience we introduce the operator $O_\lambda := A_{\alpha,t_\lambda} R_{\theta_\lambda} R_{\varphi_\lambda}$ and the operator $O_\mu := A_{\alpha,t_\mu} R_{\theta_\mu} R_{\varphi_\mu}$. Further, we define the functions $\widetilde{a}_\lambda(\xi) := a_\lambda(O_\lambda \xi)$ and $\widetilde{b}_\mu(\xi) := b_\mu(O_\mu \xi)$. We also abbreviate $\xi_\lambda := O_\lambda \xi$ and $\xi_\mu := O_\mu \xi$. Taking into account $t_\lambda \lesssim 1$, we observe $\|O_\lambda\|_{2\to 2} = \|A_{\alpha,t_\lambda}\|_{2\to 2} = \max\{t_\lambda^\alpha, t_\lambda\} \lesssim t_\lambda^\alpha$. Analogously, it holds $\|O_\mu\|_{2\to 2} = \|A_{\alpha,t_\mu}\|_{2\to 2} \lesssim t_\mu^\alpha$. Finally, we introduce the 'transfer' matrix

$$T_{\lambda,\mu} := R_{\theta_\mu} R_{\varphi_\mu} R_{\varphi_\lambda}^T R_{\theta_\lambda}^T \in \mathcal{O}(d,\mathbb{R}). \tag{7.26}$$

After these remarks we turn to the proof, where we treat the components of $\mathscr{L}_{\lambda,\mu}$ separately.

$\underline{\mathcal{I}}$ This term causes no pain.

$\underline{t_0^{-2\alpha}\Delta}$ By the product rule we have

$$\Delta(\widetilde{a}_\lambda \widetilde{b}_\mu) = \underbrace{2\langle \nabla \widetilde{a}_\lambda, \nabla \widetilde{b}_\mu \rangle}_{\mathbf{A}} + \underbrace{\widetilde{a}_\lambda \Delta \widetilde{b}_\mu + \widetilde{b}_\mu \Delta \widetilde{a}_\lambda}_{\mathbf{B}}.$$

In the following we first treat part **A** and then part **B**.

**A** The chain rule yields $\nabla \widetilde{a}_\lambda(\xi) = O_\lambda^T \nabla a_\lambda(\xi_\lambda)$ for every $\xi \in \mathbb{R}^d$ and an analogous formula for $\widetilde{b}_\mu$. Thus we obtain

$$\big\langle \nabla \widetilde{a}_\lambda(\xi), \nabla \widetilde{b}_\mu(\xi) \big\rangle = \big\langle O_\lambda^T \nabla a_\lambda(\xi_\lambda), O_\mu^T \nabla b_\mu(\xi_\mu) \big\rangle = \big\langle O_\mu O_\lambda^T \nabla a_\lambda(\xi_\lambda), \nabla b_\mu(\xi_\mu) \big\rangle.$$

The expression $\langle O_\mu O_\lambda^T \nabla a_\lambda, \nabla b_\mu \rangle$ is a linear combination of the products $\partial_i a_\lambda \partial_j b_\mu$, where $i,j \in \{1,\dots,d\}$, with the entries of the matrix $O_\mu O_\lambda^T$ as coefficients. The functions $\partial_i a_\lambda$ and $\partial_j b_\mu$ clearly satisfy (7.5) for every $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \le L-1$. Moreover, the entries of the matrix $O_\mu O_\lambda^T$ are bounded in modulus by $\|O_\mu O_\lambda^T\|_{2\to 2}$, which in turn obeys the estimate

$$\|O_\mu O_\lambda^T\|_{2\to 2} = \|A_{\alpha,t_\mu} T_{\lambda,\mu} A_{\alpha,t_\lambda}\|_{2\to 2} \le \|A_{\alpha,t_\mu}\|_{2\to 2} \|A_{\alpha,t_\lambda}\|_{2\to 2} \lesssim (t_\mu t_\lambda)^\alpha \le t_0^{2\alpha},$$

where $t_0 = \max\{t_\lambda, t_\mu\}$. This shows that the function $t_0^{-2\alpha}\mathbf{A}$ can be written as claimed.

**B** Due to symmetry it suffices to treat the term $\widetilde{b}_\mu \Delta \widetilde{a}_\lambda$. Since $\widetilde{b}_\mu(\xi) = b_\mu(\xi_\mu)$ for $\xi \in \mathbb{R}^d$ and since $b_\mu$ fulfills condition (7.5) for every $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \le L$, the function $b_\mu$ is a suitable first factor with the required properties. Let us investigate the second factor $\Delta \widetilde{a}_\lambda$.

The second derivative of $\widetilde{a}_\lambda$ is at each $\xi \in \mathbb{R}^d$ a bilinear mapping $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, which by the chain rule satisfies for $v, w \in \mathbb{R}^d$

$$\widetilde{a}_\lambda''(\xi)[v,w] = a_\lambda''(\xi_\lambda)[O_\lambda v, O_\lambda w].$$

Thus, we have the expansion

$$\Delta \widetilde{a}_\lambda(\xi) = \sum_{i=1}^d \widetilde{a}_\lambda''(\xi)[e_i, e_i] = \sum_{i=1}^d a_\lambda''(\xi_\lambda)[O_\lambda e_i, O_\lambda e_i].$$

Let $\rho \in \mathbb{N}_0^d$ be a multi-index with $|\rho|_1 \leq L-2$. Then the partial derivative with respect to $\rho$ of the function $\xi \mapsto \sum_{i=1}^d t_0^{-2\alpha} a_\lambda''(\xi)[O_\lambda e_i, O_\lambda e_i]$ clearly exists. It remains to prove the frequency localization (7.5).

In view of $\partial^\rho(a_\lambda'') = (\partial^\rho a_\lambda)''$ we can estimate for every $i \in \{1,\dots,d\}$ and every $\xi \in \mathbb{R}^d$

$$t_0^{-2\alpha} |\partial^\rho a_\lambda''(\xi)[O_\lambda e_i, O_\lambda e_i]| \leq t_0^{-2\alpha} |||\partial^\rho a_\lambda''(\xi)||| \|O_\lambda\|_{2\to 2}^2 \lesssim |||\partial^\rho a_\lambda''(\xi)|||.$$

The norm of the bilinear mapping is given by $|||\partial^\rho a_\lambda''(\xi)||| = \sup_{|v|,|w|=1} |\partial^\rho a_\lambda''(\xi)[v,w]|$. This is equal to the spectral norm of the corresponding Hesse matrix. Therefore we can deduce $|||\partial^\rho a_\lambda''(\xi)||| \lesssim \sup_{|\beta|_1=2} |\partial^\beta \partial^\rho a_\lambda(\xi)|$. The functions $\partial^\beta\partial^\rho a_\lambda$ satisfy (7.5) for every $\beta \in \mathbb{N}_0^d$ with $|\beta|_1 = 2$ due to the assumption on $a_\lambda$. The required frequency localization follows.

$\underline{t_0^{-2}(1+t_0^{-2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|^2)^{-1}\langle e_\lambda,\nabla\rangle^2}$ First we define the numbers $w_1 := t_0^{-2}$, $w_2 := t_0^{-2\alpha}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|^{-2}$, and $w_3 := t_0^{-(1+\alpha)}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|^{-1}$ and notice that the pre-factor satisfies

$$t_0^{-2}(1+t_0^{-2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|^2)^{-1} \leq \min\{w_1,w_2,w_3\}. \tag{7.27}$$

The first two estimates are obvious. For the third, recall that $1+t^2 \geq 2t$ for all $t \in \mathbb{R}$. Hence,

$$\begin{aligned} t_0^{-2}(1+t_0^{-2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|^2)^{-1} &\leq \tfrac{1}{2} t_0^{-2}(t_0^{-(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|)^{-1} \\ &\leq t_0^{-(1+\alpha)}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|^{-1}. \end{aligned}$$

We begin with the product rule, which yields

$$\langle e_\lambda,\nabla\rangle^2(\widetilde{a}_\lambda\widetilde{b}_\mu) = \widetilde{b}_\mu\langle e_\lambda,\nabla\rangle^2\widetilde{a}_\lambda + 2(\langle e_\lambda,\nabla\rangle\widetilde{a}_\lambda)(\langle e_\lambda,\nabla\rangle\widetilde{b}_\mu) + \widetilde{a}_\lambda\langle e_\lambda,\nabla\rangle^2\widetilde{b}_\mu. \tag{7.28}$$

Recall that $e_\lambda = R_{\varphi_\lambda}^T R_{\theta_\lambda}^T e_d$. We calculate with the chain rule for $\xi \in \mathbb{R}^d$

$$\langle e_\lambda, \nabla\widetilde{a}_\lambda(\xi)\rangle = \langle O_\lambda e_\lambda, \nabla a_\lambda(\xi_\lambda)\rangle = \langle A_{\alpha,t_\lambda} e_d, \nabla a_\lambda(\xi_\lambda)\rangle = t_\lambda \partial_d a_\lambda(\xi_\lambda),$$

where we used $O_\lambda e_\lambda = A_{\alpha,t_\lambda} e_d$. We similarly obtain, with $T_{\lambda,\mu}$ as in (7.26),

$$\langle e_\lambda, \nabla\widetilde{b}_\mu(\xi)\rangle = \langle O_\mu e_\lambda, \nabla b_\mu(\xi_\mu)\rangle = \langle A_{\alpha,t_\mu} T_{\lambda,\mu} e_d, \nabla b_\mu(\xi_\mu)\rangle.$$

Next, we note that $\langle e_\lambda,\nabla\rangle^2\widetilde{a}_\lambda(\xi) = \widetilde{a}_\lambda''(\xi)[e_\lambda,e_\lambda]$. Together with the chain rule, this implies

$$\langle e_\lambda,\nabla\rangle^2\widetilde{a}_\lambda(\xi) = a_\lambda''(\xi_\lambda)[O_\lambda e_\lambda, O_\lambda e_\lambda] = a_\lambda''(\xi_\lambda)[A_{\alpha,t_\lambda}e_d, A_{\alpha,t_\lambda}e_d] = t_\lambda^2 \partial_d^2 a_\lambda(\xi_\lambda).$$

We also obtain

$$\langle e_\lambda,\nabla\rangle^2\widetilde{b}_\mu(\xi) = b_\mu''(\xi_\mu)[O_\mu e_\lambda, O_\mu e_\lambda] = (\langle A_{\alpha,t_\mu} T_{\lambda,\mu} e_d, \nabla\rangle^2 b_\mu)(\xi_\mu).$$

Let us henceforth use the abbreviation $\eta := T_{\lambda,\mu} e_d \in \mathbb{S}^{d-1}$. Plugging the above calculations into (7.28) leads to the following expression for $\langle e_\lambda,\nabla\rangle^2(\widetilde{a}_\lambda\widetilde{b}_\mu)(\xi)$ at $\xi \in \mathbb{R}^d$

$$t_\lambda^2 b_\mu(\xi_\mu)\cdot\partial_d^2 a_\lambda(\xi_\lambda) + 2t_\lambda\partial_d a_\lambda(\xi_\lambda)\cdot\langle A_{\alpha,t_\mu}\eta, \nabla b_\mu(\xi_\mu)\rangle + a_\lambda(\xi_\lambda)\cdot(\langle A_{\alpha,t_\mu}\eta,\nabla\rangle^2 b_\mu)(\xi_\mu). \tag{7.29}$$

For the first summand of (7.29) we consider the product of the functions $t_\lambda^2 \partial_d^2 a_\lambda$ and $b_\mu$. Since $t_\lambda^2 \le t_0^2$ and in view of (7.27) the pre-factor $w_1$ is compensated. Due to the assumptions on $a_\lambda$ and $b_\mu$ the product is thus of the desired form.

Let us put $\eta_{[d-1]} := (\eta_1, \dots, \eta_{d-1}, 0)^T \in \mathbb{R}^d$ and $\eta_{[d]} := (0, \dots, 0, \eta_d)^T \in \mathbb{R}^d$ and observe that

$$A_{\alpha,t_\mu}\eta = A_{\alpha,t_\mu}(\eta_{[d-1]} + \eta_{[d]}) = t_\mu^\alpha \eta_{[d-1]} + t_\mu \eta_{[d]}.$$

The second summand of (7.29) then becomes – up to the factor 2 –

$$\partial_d a_\lambda(\xi_\lambda) \cdot \big(t_\lambda t_\mu^\alpha \langle \eta_{[d-1]}, \nabla b_\mu(\xi_\mu)\rangle + t_\lambda t_\mu \eta_d \partial_d b_\mu(\xi_\mu)\big).$$

We choose the function $\partial_d a_\lambda$ as the first factor, which clearly has the required properties, and the function

$$\xi \mapsto t_\lambda t_\mu^\alpha \langle \eta_{[d-1]}, \nabla b_\mu(\xi)\rangle + t_\lambda t_\mu \eta_d \partial_d b_\mu(\xi).$$

as the second factor. The second component of this function causes no problems because $|\eta_d| \le 1$ and the pre-factor $w_1$ is compensated due to $t_\lambda t_\mu \le t_0^2$. To deal with the other term, notice that by Lemma 7.6.1 $|\eta_{[d-1]}| = |\eta|_{[d-1]} \asymp |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|$. Thus

$$t_\lambda t_\mu^\alpha |\langle \eta_{[d-1]}, \nabla b_\mu\rangle| \lesssim t_\lambda t_\mu^\alpha |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}||\nabla b_\mu|.$$

The fact that $\partial_i b_\mu$, $i \in \{1, \dots, d\}$, satisfy (7.5) by assumption, and that $t_\lambda t_\mu^\alpha |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|$ compensates $w_3$, implies that also the first component satisfies the required properties.

Let us turn to the last summand of (7.29). The first factor $a_\lambda$ is of the desired form. For the second factor we expand the function $\langle A_{\alpha,t_\mu}\eta, \nabla\rangle^2 b_\mu$ in the form

$$t_\mu^{2\alpha} \langle \eta_{[d-1]}, \nabla\rangle^2 b_\mu + 2 t_\mu^{1+\alpha} \eta_d \langle \eta_{[d-1]}, \nabla\rangle \partial_d b_\mu + t_\mu^2 \eta_d^2 \partial_d^2 b_\mu.$$

Its partial derivatives of order $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \le L-2$ clearly exist, and we get the estimate

$$\begin{aligned} |\langle A_{\alpha,t_\mu}\eta, \nabla\rangle^2 \partial^\rho b_\mu| \lesssim\ & t_0^{2\alpha} |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2 \sum_{i,j=1}^{d-1} |\partial_i \partial_j \partial^\rho b_\mu| \\ & + 2 t_0^{1+\alpha} |\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}||\nabla \partial_d \partial^\rho b_\mu| + t_0^2 |\partial_d^2 \partial^\rho b_\mu|. \end{aligned}$$

We again used Lemma 7.6.1. This estimate completes the proof, taking into account the estimate (7.27) of the pre-factor and the fact that the partial derivatives of $b_\mu$ up to order $L$ satisfy (7.5). □

### 7.6.4 Actual Proof of Theorem 7.2.2

At last we have all the tools available to prove Theorem 7.2.2. Write $\Delta x = x_\lambda - x_\mu$. An application of the Plancherel identity yields

$$\begin{aligned} \langle m_\lambda, p_\mu\rangle &= \langle \hat{m}_\lambda, \hat{p}_\mu\rangle \\ &= (t_\lambda t_\mu)^{\frac{1+(d-1)\alpha}{2}} \int_{\mathbb{R}^d} \hat{a}_\lambda(A_{\alpha,t_\lambda} R_{\theta_\lambda} R_{\varphi_\lambda}\xi) \overline{\hat{b}_\mu(A_{\alpha,t_\mu} R_{\theta_\mu} R_{\varphi_\mu}\xi)} \exp(-2\pi i \langle \xi, \Delta x\rangle)\, d\xi \end{aligned}$$

for two $\alpha$-molecules $m_\lambda$ and $p_\mu$ with respective generators $a_\lambda$ and $b_\mu$. According to Lemma 7.6.3, the functions $\partial^\rho \hat{a}_\lambda$ and $\partial^\rho \hat{b}_\mu$ satisfy (7.23) for every $\rho \in \mathbb{N}_0^d$ with $|\rho|_1 \le L$

Next, we want to exploit cancellation. For this we utilize the differential operator $\mathscr{L}_{\lambda,\mu}$ from (7.25). First, we observe that partial integration yields

$$\Big\langle \mathscr{L}_{\lambda,\mu}^N \exp(-2\pi i\langle\xi,\Delta x\rangle), \hat{a}_\lambda(A_{\alpha,t_\lambda}R_{\theta_\lambda}R_{\varphi_\lambda}\xi)\overline{\hat{b}_\mu(A_{\alpha,t_\mu}R_{\theta_\mu}R_{\varphi_\mu}\xi)}\Big\rangle$$
$$= \Big\langle \exp(-2\pi i\langle\xi,\Delta x\rangle), \mathscr{L}_{\lambda,\mu}^N\big(\hat{a}_\lambda(A_{\alpha,t_\lambda}R_{\theta_\lambda}R_{\varphi_\lambda}\xi)\overline{\hat{b}_\mu(A_{\alpha,t_\mu}R_{\theta_\mu}R_{\varphi_\mu}\xi)}\big)\Big\rangle,$$

since the boundary terms vanish due to the decay properties of the generators and their derivatives. Note that we assume $N_1 > d/2$ and $L \geq 2N$. Second, we calculate for $\xi \in \mathbb{R}^d$

$$\mathscr{L}_{\lambda,\mu}^N\big(\exp(-2\pi i\langle\xi,\Delta x\rangle)\big) = \left(1 + 4\pi^2 t_0^{-2\alpha}|\Delta x|^2 + \frac{4\pi^2 t_0^{-2}\langle e_\lambda,\Delta x\rangle^2}{1 + t_0^{-2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|^2}\right)^N \cdot \exp(-2\pi i\langle\xi,\Delta x\rangle).$$

Consequently, we have

$$\langle m_\lambda, p_\mu\rangle = \left(1 + 4\pi^2 t_0^{-2\alpha}|\Delta x|^2 + \frac{4\pi^2 t_0^{-2}\langle e_\lambda,\Delta x\rangle^2}{1 + t_0^{-2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|^2}\right)^{-N} \cdot \mathcal{S}_{\lambda,\mu},$$

with

$$\mathcal{S}_{\lambda,\mu} := (t_\lambda t_\mu)^{\frac{1+(d-1)\alpha}{2}} \int_{\mathbb{R}^d} \mathscr{L}_{\lambda,\mu}^N\big(\hat{a}_\lambda(A_{\alpha,t_\lambda}R_{\theta_\lambda}R_{\varphi_\lambda}\xi)\overline{\hat{b}_\mu(A_{\alpha,t_\mu}R_{\theta_\mu}R_{\varphi_\mu}\xi)}\big)\exp(-2\pi i\langle\xi,\Delta x\rangle)\,d\xi.$$

Since $L \geq 2N$ by assumption, Lemma 7.6.9 can iteratively be applied $N$ times, and we conclude that

$$\mathscr{L}_{\lambda,\mu}^N\big(\hat{a}_\lambda(A_{\alpha,t_\lambda}R_{\theta_\lambda}R_{\varphi_\lambda}\xi)\overline{\hat{b}_\mu(A_{\alpha,t_\mu}R_{\theta_\mu}R_{\varphi_\mu}\xi)}\big)$$

can be written as a finite linear combination of terms of the form

$$p_\lambda(A_{\alpha,t_\lambda}R_{\theta_\lambda}R_{\varphi_\lambda}\xi)q_\mu(A_{\alpha,t_\mu}R_{\theta_\mu}R_{\varphi_\mu}\xi),$$

where $p_\lambda$ and $q_\mu$ satisfy (7.22) (for the multi-index $\rho \in \mathbb{N}_0^d$ just containing zeros).

Using Lemma 7.6.3 and putting $K = 2N + d - 2 \leq N_2$ in Lemma 7.6.4 then yields

$$|\mathscr{L}_{\lambda,\mu}^N\big(\hat{a}_\lambda(A_{\alpha,t_\lambda}R_{\theta_\lambda}R_{\varphi_\lambda}\xi)\overline{\hat{b}_\mu(A_{\alpha,t_\mu}R_{\theta_\mu}R_{\varphi_\mu}\xi)}\big)|$$
$$\lesssim S_{\lambda,M-(2N+d-2),N_1,2N+d-2}(\xi)S_{\mu,M-(2N+d-2),N_1,2N+d-2}(\xi).$$

Due to the assumptions, we can further choose a number $\widetilde{N} \leq N_1$ which satisfies

$$(M - (2N + d - 2)) + d > \widetilde{N} \geq N + \frac{1 + (d-1)\alpha}{2}. \tag{7.30}$$

Since $\widetilde{N} \leq N_1$ we have the estimate $S_{\eta,M-(2N+d-2),N_1,2N+d-2} \leq S_{\eta,M-(2N+d-2),\widetilde{N},2N+d-2}$ for $\eta = \lambda,\ \mu$. Hence, we obtain

$$\begin{aligned}|\mathcal{S}_{\lambda,\mu}| &\lesssim (t_\lambda t_\mu)^{\frac{1+(d-1)\alpha}{2}} \int_{\mathbb{R}^d} S_{\lambda,M-(2N+d-2),N_1,2N+d-2)}(\xi)S_{\mu,M-(2N+d-2),N_1,2N+d-2}(\xi)\,d\xi \\ &\lesssim (t_\lambda t_\mu)^{\frac{1+(d-1)\alpha}{2}} \int_{\mathbb{R}^d} S_{\lambda,M-(2N+d-2),\widetilde{N},2N+d-2}(\xi)S_{\mu,M-(2N+d-2),\widetilde{N},2N+d-2}(\xi)\,d\xi \\ &\lesssim \max\left\{\frac{t_\lambda}{t_\mu},\frac{t_\mu}{t_\lambda}\right\}^{-N}(1 + t_0^{-(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda,e_\mu)\}|)^{-2N}.\end{aligned}$$

Here we used (7.30) and Lemma 7.6.8 in the last line (using this $S$ and setting $\tilde{M} = M - (2N + d - 2)$, $A = N$ and $B = 2N$ ($B > 1$, $A > 0$ since $N > 1$)).

Altogether, we arrive at the desired estimate

$$
\begin{aligned}
|\langle m_\lambda, p_\mu \rangle| &\lesssim \max\left\{\frac{t_\lambda}{t_\mu}, \frac{t_\mu}{t_\lambda}\right\}^{-N} \left(1 + t_0^{-2\alpha}|\Delta x|^2 + \frac{t_0^{-2}\langle e_\lambda, \Delta x\rangle^2}{1 + t_0^{-2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2}\right)^{-N} \\
&\quad \cdot \left(1 + t_0^{-(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|\right)^{-2N} \\
&\lesssim \max\left\{\frac{t_\lambda}{t_\mu}, \frac{t_\mu}{t_\lambda}\right\}^{-N} \\
&\quad \cdot \left(1 + t_0^{-2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2 + t_0^{-2\alpha}|\Delta x|^2 + \frac{t_0^{-2}\langle e_\lambda, \Delta x\rangle^2}{1 + t_0^{-2(1-\alpha}|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2}\right)^{-N} \\
&\lesssim \omega_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu))^{-N}.
\end{aligned}
$$

For the last estimate observe that the inequality between the arithmetic and the geometric mean

$$
(1 + t_0^{-2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2) + \frac{t_0^{-2}\langle e_\lambda, \Delta x\rangle^2}{1 + t_0^{-2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2} \geq 2t_0^{-1}|\langle e_\lambda, \Delta x\rangle|
$$

implies

$$
\begin{aligned}
&1 + t_0^{-2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2 + t_0^{-2\alpha}|\Delta x|^2 + \frac{t_0^{-2}\langle e_\lambda, \Delta x\rangle^2}{1 + t_0^{-2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2} \\
&\geq \frac{1}{2}\left(1 + t_0^{-2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2 + t_0^{-2\alpha}|\Delta x|^2\right) \\
&\quad + \frac{1}{2}\left(1 + t_0^{-2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2 + \frac{t_0^{-2}\langle e_\lambda, \Delta x\rangle^2}{1 + t_0^{-2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2}\right) \\
&\gtrsim 1 + t_0^{-2(1-\alpha)}|\{d_{\mathbb{S}}(e_\lambda, e_\mu)\}|^2 + t_0^{-2\alpha}|\Delta x|^2 + t_0^{-1}|\langle e_\lambda, \Delta x\rangle| = 1 + d_\alpha(\Phi_\Lambda(\lambda), \Phi_\Delta(\mu)).
\end{aligned}
$$

This concludes the proof. $\square$

# Bibliography


[1] S. T. Ali, J.-P. Antoine, and J.-P. Gazeau. Continuous frames in Hilbert space. *Ann. Physics*, 222(1):1–37, 1993.

[2] H. W. Alt. *Lineare Funktionalanalysis.* Springer-Verlag, Berlin-Heidelberg, 5. edition, 2006.

[3] T. Aoki. Locally bounded linear topological spaces. *Proc. Imp. Acad. Tokyo*, 18:588–594, 1942.

[4] P. Balazs and N. Holighaus. Discretization in generalized coorbit spaces: extensions, annotations and errata for " Continuous frames, function spaces, and the discretization problem" by M. Fornasier and H. Rauhut. Online available: `https://www.univie.ac.at/nonstatgab/warping/baho15.pdf`.

[5] T. Berger. *Rate Distortion Theory.* Wiley Online Library, 1971.

[6] L. Borup and M. Nielsen. Frame decompositions of decomposition spaces. *J. Fourier Anal. Appl.*, 13:39–70, 2007.

[7] J. Cai, B. Dong, S. Osher, and Z. Shen. Image restoration: total variation, wavelet frames, and beyond. *J. Amer. Math. Soc.*, 25:1033–1089, 2012.

[8] E. Candès. Ridgelets: theory and applications. PhD thesis, Stanford University, 1998.

[9] E. Candès and D. L. Donoho. Ridgelets: a key to higher-dimensional intermittency? *Phil. Trans. R. Soc. Lond. A.*, 357:2495–2509, 1999.

[10] E. Candès and D. L. Donoho. Continuous curvelet transform: I. Resolution of the wavefront set. *Appl. Comput. Harmon. Anal.*, 19(2):162–197, 2005.

[11] E. Candès and D. L. Donoho. Continuous curvelet transform: II. Discretization and frames. *Appl. Comput. Harmon. Anal.*, 19(2):198–222, 2005.

[12] E. J. Candès. Ridgelets and the representation of mutilated Sobolev functions. *SIAM J. Math. Anal.*, 33(2):347–368, 2001.

[13] E. J. Candès and L. Demanet. The curvelet representation of wave propagators is optimally sparse. *Comm. Pure Appl. Math.*, 58:1472–1528, 2005.

[14] E. J. Candès and D. L. Donoho. Curvelets - a surprisingly effective nonadaptive representation for objects with edges. In C. Rabut, A. Cohen, and L. Schumaker, editors, *Curves and Surfaces*, pages 105–120. Vanderbilt University Press, 2000.

[15] E. J. Candès and D. L. Donoho. New tight frames of curvelets and optimal representations of objects with $C^2$ singularities. *Comm. Pure Appl. Math.*, 56:219–266, 2004.

[16] V. Chandrasekaran, M. B. Wakin, D. Baron, and R. G. Baraniuk. Compressing piecewise smooth multidimensional functions using surflets: rate-distortion analysis. Technical report, Department of Electrical and Computer Engineering, Rice University, Mar. 2004.

[17] V. Chandrasekaran, M. B. Wakin, D. Baron, and R. G. Baraniuk. Compression of higher dimensional functions containing smooth discontinuities. In *Conference on Information Sciences and Systems*, Princeton, Mar. 2004.

[18] V. Chandrasekaran, M. B. Wakin, D. Baron, and R. G. Baraniuk. Surflets: a sparse representation for multidimensional functions containing smooth discontinuities. In *IEEE Symposium on Information Theory*, Chicago, Jul. 2004.

[19] V. Chandrasekaran, M. B. Wakin, D. Baron, and R. G. Baraniuk. Representation and compression of multidimensional piecewise functions using surflets. *IEEE Trans. Inform. Theory*, 55(1):374–400, 2009.

[20] O. Christensen. *An Introduction to Frames and Riesz Bases.* Birkhäuser, 2003.

[21] C. Christopoulos, A. Skodras, and T. Ebrahimi. The JPEG2000 still image coding system: an overview. *IEEE Trans. Consum. Electron.*, 46(4):1103–1127, 2000.

[22] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods for elliptic operator equations: convergence rates. *Math. Comp.*, 70:27–75, 2001.

[23] S. Dahlke, F. De Mari, E. De Vito, D. Labate, G. Steidl, G. Teschke, and S. Vigogna. Coorbit spaces with voice in a Fréchet space. *J. Fourier Anal. Appl.*, 23(1):141–206, 2017.

[24] S. Dahlke, M. Fornasier, H. Rauhut, G. Steidl, and G. Teschke. Generalized coorbit theory, Banach frames, and the relation to alpha-modulation spaces. *Proc. London Math. Soc. (3)*, 96:464–506, 2008.

[25] S. Dahlke, G. Kutyniok, P. Maass, C. Sagiv, and H.-G. Stark. The uncertainty principle associated with the continuous shearlet transform. *Int. J. Wavelets Multiresolut. Inf. Process.*, 6:157–181, 2008.

[26] S. Dahlke, G. Kutyniok, G. Steidl, and G. Teschke. Shearlet coorbit spaces and associated Banach frames. *Appl. Comput. Harmon. Anal.*, 27(2):195–214, 2009.

[27] S. Dahlke, G. Steidl, and G. Teschke. Coorbit spaces and Banach frames on homogeneous spaces with applications to the sphere. *Adv. Comput. Math.*, 21(1-2):147–180, 2004.

[28] S. Dahlke, G. Steidl, and G. Teschke. Weighted coorbit spaces and Banach frames on homogeneous spaces. *J. Fourier Anal. Appl.*, 10(5):507–539, 2004.

[29] W. Dahmen, C. Huang, G. Kutyniok, C. Schwab, and G. Welper. Efficient resolution of anisotropic structures. In *Extraction of Quantifiable Information from Complex Systems*, volume 102 of *Lecture Notes in Computational Science and Engineering*, pages 25–51. Springer, 2014.

[30] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Inf. Theor.*, 36(5):961–1005, 1990.

[31] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.

[32] M. V. de Hoop, K. Gröchenig, and J. L. Romero. Exact and approximate expansions with pure Gaussian wavepackets. *SIAM J. Math. Anal.*, 46(3):2229–2253, 2014.

[33] M. N. Do and M. Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Trans. Image Proc.*, 14:2091–2106, 2005.

[34] D. Donoho and G. Kutyniok. Microlocal analysis of the geometric separation problem. *Comm. Pure Appl. Math.*, 66:1–47, 2013.

[35] D. L. Donoho. Wedgelets: nearly-minimax estimation of edges. *Ann. Statist.*, 27:859–897, 1999.

[36] D. L. Donoho. Orthonormal ridgelets and linear singularities. *SIAM J. Math. Anal.*, 31(5):1062–1099, 2000.

[37] D. L. Donoho. Ridge functions and orthonormal ridgelets. *J. Approx. Theory*, 111(2):143–179, 2001.

[38] D. L. Donoho. Sparse components of images and optimal atomic decomposition. *Constr. Approx.*, 17:353–382, 2001.

[39] D. L. Donoho and X. Huo. Beamlet pyramids: a new form of multiresolution analysis suited for extracting lines, curves, and objects from very noisy image data. In *Wavelet Applications in Signal and Image Processing VIII (San Diego, CA, 2000), Proc. SPIE*, volume 4119, pages 434–444. SPIE, 2000.

[40] G. Easley and D. Labate. *Shearlets: Multiscale Analysis for Multivariate Data*, chapter Image Processing using Shearlets, pages 283–320. Birkhäuser Boston, 2012.

[41] H. G. Feichtinger. Modulation spaces on locally compact Abelian groups. Technical report, University Vienna, January 1983.

[42] H. G. Feichtinger and K. Gröchenig. A unified approach to atomic decompositions via integrable group representations. In *Function spaces and applications (Lund, 1986)*, volume 1302 of *Lecture Notes in Math.*, pages 52–73. Springer, Berlin, 1988.

[43] H. G. Feichtinger and K. Gröchenig. Banach spaces related to integrable group representations and their atomic decompositions, I. *Journ. Funct. Anal.*, 86(2):307–340, 1989.

[44] A. Flinth. 3D parabolic molecules. Bachelor's thesis, TU Berlin, 2013.

[45] A. Flinth and M. Schäfer. Multivariate $\alpha$-molecules. *J. Approx. Theory*, 202:64–108, 2016.

[46] M. Fornasier and H. Rauhut. Continuous frames, function spaces, and the discretization problem. *J. Fourier Anal. Appl.*, 11(3):245–287, 2005.

[47] M. Frazier, B. Jawerth, and G. Weiss. Littlewood-paley theory and the study of function spaces. In *Conference Board of the Mathematical Sciences, Regional Conference Series in Mathematics*, volume 79. American Mathematical Society, Providence, R.I, 1991.

[48] H. Führ. Coorbit spaces and wavelet coefficient decay over general dilation groups. *Trans. Amer. Math. Soc.*, 367:7373–7401, 2015.

[49] H. Führ. Vanishing moment conditions for wavelet atoms in higher dimensions. *Adv. Comput. Math.*, 42(1):127–153, 2015.

[50] H. Führ and F. Voigtlaender. Wavelet coorbit spaces viewed as decomposition spaces. *J. Funct. Anal.*, 269:80–154, 2015.

[51] L. Grafakos. *Classical Fourier Analysis.* Springer, 2 edition, 2008.

[52] R. Gribonval and M. Nielsen. Non-linear approximation with dictionaries. I. Direct estimates. *J. Fourier Anal. Appl.*, 10:51–71, 2004.

[53] A. Grigoreva. Eine hinreichende Bedingung für die Frame-Eigenschaft von $\alpha$-Molekülsystemen. Bachelor's thesis, TU Berlin, 2017.

[54] K. Gröchenig. Unconditional bases in translation and dilation invariant function spaces on $\mathbb{R}^n$. In *Constructive theory of functions (Varna, 1987)*, pages 174–183. Publ. House Bulgar. Acad. Sci., Sofia, 1988.

[55] K. Gröchenig. Describing functions: atomic decompositions versus frames. *Monatsh. Mathem.*, 112:1–41, 1991.

[56] K. Gröchenig. *Foundations of Time-Frequency Analysis.* Birkhäuser, Boston, 2001.

[57] P. Grohs. Ridgelet-type frame decompositions for Sobolev spaces related to linear transport. *J. Fourier Anal. Appl.*, 18(2):309–325, 2012.

[58] P. Grohs. Intrinsic localization of anisotropic frames. *Appl. Comput. Harmon. Anal.*, 2013. to appear.

[59] P. Grohs, S. Keiper, G. Kutyniok, and M. Schäfer. $\alpha$-Molecules. *Appl. Comput. Harmon. Anal.*, 41(1):297–336, 2016.

[60] P. Grohs, S. Keiper, G. Kutyniok, and M. Schäfer. Cartoon approximation with $\alpha$-curvelets. *J. Fourier Anal. Appl.*, 22(6):1235–1293, 2016.

[61] P. Grohs and Z. Kereta. Continuous parabolic molecules. Technical Report 2015-17, Seminar for Applied Mathematics, ETH Zürich, Switzerland, June 2015.

[62] P. Grohs and G. Kutyniok. Parabolic molecules. *Found. Comput. Math.*, 14(2):299–337, 2014.

[63] P. Grohs and A. Obermeier. On the approximation of functions with line singularities by ridgelets. Technical Report 2016-4, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2016.

[64] P. Grohs and A. Obermeier. Optimal adaptive ridgelet schemes for linear advection equations. *Appl. Comput. Harmon. Anal.*, 41(3):768–814, 2016.

[65] P. Grohs and S. Vigogna. Intrinsic localization of anisotropic frames II: $\alpha$-molecules. *J. Fourier Anal. Appl.*, 21(1):182–205, 2015.

[66] K. Guo, G. Kutyniok, and D. Labate. Sparse multidimensional representations using anisotropic dilation and shear operators. In *Wavelets and Splines (Athens, GA, 2005)*, volume 14, pages 189–201. Nashboro Press, Nashville, TN, 2006.

[67] K. Guo and D. Labate. Optimally sparse multidimensional representation using shearlets. *SIAM J. Math. Anal.*, 39:298–318, 2007.

[68] K. Guo and D. Labate. Representation of Fourier integral operators using shearlets. *J. Fourier Anal. Appl.*, 14:327–371, 2008.

[69] K. Guo and D. Labate. Optimally sparse representations of $3D$ data with $C^2$ surface singularities using Parseval frames of shearlets. *SIAM J. Math. Anal.*, 44:851–886, 2012.

[70] K. Guo and D. Labate. The construction of smooth Parseval frames of shearlets. *Math. Model. Nat. Phenom.*, 8(1):82–105, 2013.

[71] W. Hackbusch, H. R. Schwarz, and E. Zeidler. *Teubner-Taschenbuch der Mathematik.* B. G. Teubner Stuttgart, Leipzig, 1996.

[72] L. Jantscher. *Distributionen.* Walter de Gruyter, Berlin-New York, 1971.

[73] S. Keiper. A flexible shearlet transform - sparse approximation and dictionary learning. Bachelor's thesis, TU Berlin, 2013.

[74] H. Kempka, M. Schäfer, and T. Ullrich. General coorbit space theory for quasi-Banach spaces and inhomogeneous function spaces with variable smoothness and integrability. *J. Fourier Anal. Appl.*, 23(6):1348–1407, 2017.

[75] Z. Kereta. *Continuous Parabolic Molecules.* Doctoral thesis, ETH Zürich, Germany, 2016.

[76] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Construction of compactly supported shearlet frames. *Constr. Approx.*, 35(1):21–72, 2012.

[77] J. Krommweh. Image approximation by adaptive tetrolet transform. In *International conference on sampling theory and applications*, Marseille, France, May 2009.

[78] G. Kutyniok and W.-Q Lim. Compactly supported shearlets are optimally sparse. *J. Approx. Theory*, 163(11):1564–1589, 2011.

[79] G. Kutyniok and D. Labate, editors. *Shearlets: Multiscale Analysis for Multivariate Data.* Birkhäuser, Boston, 2012.

[80] G. Kutyniok and D. Labate. *Shearlets: Multiscale Analysis for Multivariate Data*, chapter Introduction to Shearlets, pages 1–38. Birkhäuser Boston, 2012.

[81] G. Kutyniok, D. Labate, W.-Q Lim, and G. Weiss. Sparse multidimensional representation using shearlets. In *Wavelets XI(San Diego, CA), SPIE Proc.*, volume 5914, pages 254–262. SPIE, Bellingham, WA, 2005.

[82] G. Kutyniok, J. Lemvig, and W.-Q Lim. Compactly supported shearlets. In M. Neamtu and L. Schumaker, editors, *Approximation Theory XIII: San Antonio 2010*, volume 13 of *Springer Proceedings in Mathematics*, pages 163–186. Springer New York, 2012.

[83] G. Kutyniok, J. Lemvig, and W.-Q Lim. Optimally sparse approximations of 3D functions by compactly supported shearlet frames. *SIAM J. Math. Anal.*, 44(4):2962–3017, 2012.

[84] G. Kutyniok, V. Mehrmann, and P. Petersen. Regularization and numerical solution of the inverse scattering problem using shearlet frames. *J. Inverse Ill-Posed Probl.*, 25(3):287–309, 2017.

[85] D. Labate, L. Mantovani, and P. Negi. Shearlet smoothness spaces. *J. Fourier Anal. Appl.*, 19(3):577–611, 2013.

[86] E. Le Pennec and S. Mallat. Bandelet image approximation and compression. *Multiscale Model. Simul.*, 4(3):992–1039, 2005.

[87] E. Le Pennec and S. Mallat. Sparse geometric image representations with bandelets. *IEEE Trans. Image Process.*, 14(4):423–438, 2005.

[88] Y. Liang, Y. Sawano, T. Ullrich, D. Yang, and W. Yuan. New characterizations of Besov-Triebel-Lizorkin-Hausdorff spaces including coorbits and wavelets. *J. Fourier Anal. Appl.*, 18(5):1067–1111, 2012.

[89] Y. Liang, Y. Sawano, T. Ullrich, D. Yang, and W. Yuan. A new framework for generalized Besov-type and Triebel-Lizorkin-type spaces. *Diss. Math.*, 489, 2013.

[90] A. Lisowska. Smoothlets – multiscale functions for adaptive representation of images. *IEEE Trans. Image Process.*, 20(7):1777–1787, 2011.

[91] A. Lisowska. Multiwedgelets in image denoising. In J. Park, J. Ng, H.-Y. Jeong, and B. Waluyo, editors, *Multimedia and Ubiquitous Engineering: MUE 2013*, pages 3–11. Springer Netherlands, Dordrecht, 2013.

[92] J. Ma. Seismic denoising with nonuniformly sampled curvelets. *Comput. Sci. Eng.*, 8(3):16–25, 2006.

[93] J. Ma. Curvelets for surface characterization. *Appl. Phys. Lett.*, 90:054109, 2007.

[94] J. Ma, M. Hussaini, O. Vasilyev, and F.-X. Le Dimet. Multiscale geometric analysis of turbulence by curvelets. *Physics of Fluids*, 21:075104, 2009.

[95] J. Ma and G. Plonka. Combined curvelet shrinkage and nonlinear anisotropic diffusion. *IEEE Trans. Image Process.*, 16(9):2198–2206, 2007.

[96] J. Ma and G. Plonka. A review of curvelets and recent applications. In *IEEE Signal Processing Magazine*, 2009.

[97] S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way.* Academic Press, 2nd edition, 2008.

[98] S. Mallat. Geometrical grouplets. *Appl. Comput. Harmon. Anal.*, 26(2):161–180, 2009.

[99] K. N. Rasmussen and M. Nielsen. Compactly supported curvelet-type systems. *J. Funct. Spaces Appl.*, pages Art. ID 876315, pp. 18, 2012.

[100] H. Rauhut. Coorbit space theory for quasi-Banach spaces. *Studia Math.*, 180(3):237–253, 2007.

[101] S. Rolewicz. On a certain class of linear metric spaces. *Bull. Acad. Polon. Sci. Cl. III.*, 5:471–473, 1957.

[102] M. Schäfer. The role of $\alpha$-scaling for cartoon approximation. 2016. submitted.

[103] B. Scharf. Atomare Charakterisierungen vektorwertiger Funktionenräume. Diplomarbeit, Friedrich-Schiller-Universität Jena, Germany, 2009.

[104] H. Smith. A Hardy space for Fourier integral operators. *J. Geom. Anal.*, 8(4):629–653, 1998.

[105] R. S. Strichartz. *A Guide to Distribution Theory and Fourier Transforms.* World Scientific Publishing Co. Pte. Ltd., 1994.

[106] B. Sun, J. Ma, H. Chauris, and H. Yang. Solving the Wave Equation Using Curvelets. In *72nd EAGE Conference and Technical Exhibition, Eur. Ass. of Geoscientists and Engineers*, page C044, Barcelone, Spain, June 2010.

[107] H. Triebel. *Theory of Function Spaces.* Birkhäuser, Basel, 1983.

[108] H. Triebel. Characterizations of Besov-Hardy-Sobolev spaces: a unified approach. *Journ. of Approx. Theory*, 52:162–203, 1988.

[109] H. Triebel. *Theory of Function Spaces II.* Birkhäuser, Basel, 1992.

[110] T. Ullrich. Continuous characterizations of Besov-Lizorkin-Triebel spaces and new interpretations as coorbits. *Journ. Funct. Spaces Appl.*, page Art. ID 163213, 2012.

[111] T. Ullrich and H. Rauhut. Generalized coorbit space theory and inhomogeneous function spaces of Besov-Lizorkin-Triebel type. *J. Funct. Anal.*, 260(11):3299–3362, 2011.

[112] D. Werner. *Funktionalanalysis.* Springer, Berlin-Heidelberg-New York, 5. edition, 2004. erweiterte Auflage.

[113] R. M. Willet and R. D. Nowak. Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging. *IEEE Trans. Med. Imag.*, 22(3):332–350, 2003.

[114] P. Wojtaszczyk. *A Mathematical Introduction to Wavelets.* Cambridge University Press, Cambridge, 1997.

[115] L. Ying, L. Demanet, and E. Candès. 3D discrete curvelet transform. In *Wavelets XI(San Diego, CA), SPIE Proc.*, volume 5914. SPIE, 2005.