B-Splines und Frames

Wolfgang Kreuzer[1,2], Johannes Brand[1]
[1] *Österreichische Akademie der Wissenschaften, Institut für Schallforschung, 1040 Wien, Österreich*
[2] *Email: wolfgang.kreuzer@oeaw.ac.at*

Einleitung

Eine der größten Herausforderungen in der Numerik ist die effiziente und genaue Darstellung von Funktionen $f(x) = \sum_{i=1}^{\infty} c_i g_i(x)$ mittels einfacher Bausteine $g_i(x)$ (z.B. mittels Polynomen, komplexen Exponentialfunktionen, usw.). Die Anwendungsgebiete reichen hierbei von Polynominterpolation, über geeignete Ansatzfunktionen für FEM/BEM bis zur Signalverarbeitung und Computergraphik. Im Allgemeinen wird versucht, $g_i(x)$ aus einer Basis, d.h. eines linear unabhängigen Erzeugendensystems, zu wählen. Im Speziellen werden gerne orthonormale Basisfunktionen verwendet, weil die unbekannten (eindeutigen) Koeffizienten c_i mittels einfacher Multiplikation im Vektorraum bestimme werden können: $c_i = \langle f, g_i \rangle$. Die lineare Unabhängigkeit hat aber auf der anderen Seite den Nachteil, dass Basen mit speziellen Eigenschaften oft (wenn überhaupt) nur schwer konstruiert werden können (z.B. spezielle Waveletbasen auf dem Intervall, Ansatzfunktionen für hochfrequente Helmholtz-BEM, etc.). Es stellt sich also die Frage, ob es nicht Sinn machen würde, z.B. auf die lineare Unabhängigkeit zu verzichten, um im Gegenzug dafür leichter Erzeugendensysteme mit anderen speziellen Eigenschaften konstruieren zu können? Aus dieser Frage heraus wurde das Konzept der s.g. Frames entwickelt.

In der Signalverarbeitung spielen Frames bereits eine sehr wichtige Rolle, z.B. bei der Zeit-Frequenzdarstellung (Spektogramm) eines Signals. Im Allgemeinen sind Frames linear abhängige Erzeugendensysteme mit speziellen Eigenschaften. Frei interpretiert sind Frames das "nächstbeste" System zu einer orthonormalen Basis.

Gerade im Bereich der Helmholtz BEM könnten Framefunktionen, die durch Kombination von herkömmlichen Funktionen wie B-Splines mit komplex-wertigen Exponentialfunktion $e^{\mathrm{i}kx}$ gebildet werden, effiziente Ansatzfunktionen sein, da es mit ihnen auch möglich ist, oszillierende Komponenten der Lösung darzustellen.

B-Splines

Auf Grund ihrer einfachen Konstruktion

$$N_1(x) = \begin{cases} 1 & x \in [0,1], \\ 0 & \text{sonst} \end{cases} \tag{1}$$

$$N_{n+1}(x) = N_n(x) * N_1(x) = \int_0^1 N_n(x-t)dt \tag{2}$$

und ihres beschränkten Trägers $\mathrm{supp}(N_n) = [0, n]$ spielen B-Splines sowohl in der Computergraphik als auch in der numerischen Mathematik eine wichtige Rolle. Darüber

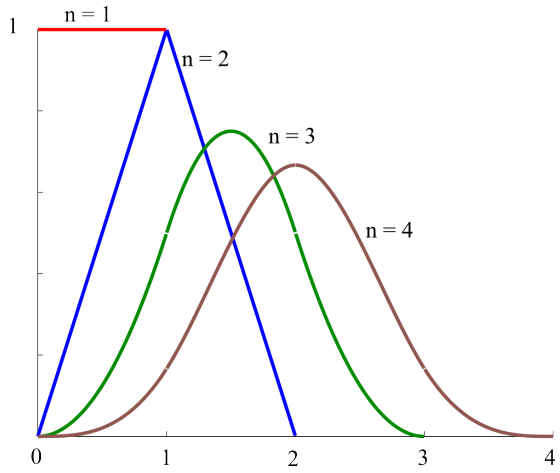

Abbildung 1: B-Spline Funktionen N_n der Ordnungen $n = 1$ bis $n = 4$.

hinaus besitzen B-Splines noch die angenehme Eigenschaften, dass

$$\int_{\infty}^{\infty} N_n(x) f(x) dx = \int_{[0,1]^n} f(x_1 + \cdots + x_n) dx_1 \cdots dx_n \tag{3}$$

und dass

$$\sum_{k \in \mathbb{Z}} N_n(x-k) = 1 \quad \forall x \in \mathbb{R}. \tag{4}$$

Frames

Per Definition handelt es sich bei Frames um (meist abzählbare) Familien von Funktionen $\{g_i\}_{i \in \mathbb{N}}$ eines Hilbertraums $\mathcal{H}$, für die es zwei Konstanten $A, B > 0$ gibt, sodass

$$A \, ||f||^2 \leq \sum_{i=1}^{\infty} |\langle f, g_i \rangle|^2 \leq B \, ||f||^2 \quad \forall f \in \mathcal{H}. \tag{5}$$

Weiters existiert für jeden Frame $\mathcal{G} := \{g_i\}_{i \in \mathbb{N}}$ mindestens ein (nicht notwendigerweise eindeutiger) dualer Frame $\tilde{\mathcal{G}} := \{\tilde{g}_i\}_{i \in \mathbb{N}}$, sodass jede Funktion des Raums durch dieses Framepaar dargestellt werden kann:

$$f = \sum_{i=1}^{\infty} \langle f, \tilde{g}_i \rangle g_i = \sum_{i=1}^{\infty} \langle f, g_i \rangle \tilde{g}_i, \tag{6}$$

wobei bei den Summen eine unbedingte Konvergenz gegeben ist.

Frames (und ihre dualen Frames) bilden ein Erzeugendensystem des Hilbertraums $\mathcal{H}$ und verallgemeinern somit den Begriff einer Basis. Unter den möglichen dualen

Frames spielt der kanonische duale Frame, der durch die Invertierung des Frameoperators $S : f \to \sum_{i\in\mathbb{N}}\langle f, g_i\rangle g_i$ erzeugt wird, eine besondere Rolle, weil er in vielen Fällen mit relativ einfachen Mitteln numerisch approximiert werden kann. In der Praxis lässt sich dieser Frame durch die Pseudoinverse der Matrix A, deren Einträge $A_{ij} := g_j(x_i)$ durch die auf einem Gitter abgetastete Werte der Frameelemente gegeben sind, numerisch approximieren.

Gegenüber einer Basis liegt der Vorteil eines Frames in der Redundanz der Darstellung, die eine leichtere und flexiblere Konstruktion für Erzeugendensysteme mit speziellen Eigenschaften bietet. So ermöglichen Frames beispielsweise eine Darstellung eines Signals, die in Zeit *und* Frequenz konzentriert werden kann, was mit einer Basis nicht möglich ist. Natürlich bringt diese Redundanz auch Nachteile mit sich: Durch die größere Anzahl von Framefunktionen werden System in der Regel um einiges größer, und es empfehlen sich spezielle Verfahren, die zur Best-N-Term Approximation verwendet werden können.

Gabor Frames

Gabor Frames sind Frames, die mittels Modulationen E_{mb} und Translationen T_{na} einer festen Fensterfunktion $g(x)$ konstruiert werden, wobei

$$T_a g := g(x-a), \quad E_b g := g(x)e^{2\pi i b x}. \tag{7}$$

Das System

$$\mathcal{G}(g,a,b) = \{E_{mb}T_{na}g(x)\}_{m,n\in\mathbb{Z}} \tag{8}$$

bildet unter bestimmten Voraussetzungen an das Fenster und die Parameter a (die Schrittweite im Ort oder in der Zeit) und b (die Schrittweite in der Frequenz) einen s.g. Gabor Frame.

B-Spline Frames

Auf Grund ihrer Eigenschaften sind B-Splines hervorragende Fensterfunktionen $g(x)$, mit denen Gabor Frames gebildet werden können. Es gibt einfache Regeln zur Bestimmung der Frameparameter a und b und zur einfachen Konstruktion von dualen Frames [2]:

Theorem 1 *Seien $N \in \mathbb{N}$, $g(x)$ eine reelwertige Funktion mit beschränktem Träger $supp(g) \subseteq [0,N]$ und mit $\sum_{k\in\mathbb{Z}} g(x-k) = 1$. Seien $a = 1, b \in (0, \frac{1}{2N-1}]$, und*

$$\tilde{g} := bg(x) + 2b\sum_{n=1}^{N-1} g(x+n). \tag{9}$$

Dann bilden die Systeme $\{E_{mb}T_{na}g\}_{m,n\in\mathbb{Z}}$ und $\{E_{mb}T_{na}\tilde{g}\}_{m,n\in\mathbb{Z}}$ ein Framepaar für $L^2(\mathbb{R})$.

Theorem 2 *Geben $N \in \mathbb{N}$. Sei $g(x) \in L^2(\mathbb{R})$ eine reelwertige, beschränkte Funktion mit $supp(g) \subseteq [0,N]$ und $\sum_{k\in\mathbb{Z}} g(x-k) = 1$. Seien $a = 1$ und $b \in (0, \frac{1}{2N-1}]$. Dann erzeugt die Funktion*

$$h(x) = \sum_{k=-N+1}^{N-1} g(x+k) \tag{10}$$

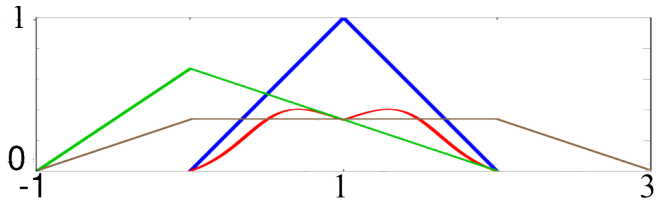

Abbildung 2: Blau: B-Spline der Ordnung 2, Rot: Kanonisches Duale Fenster, Grün: Duales Fenster nach Theorem 1, Braun: Duales Fenster nach Theorem 2. Für die dualen Fenster wurden die Frameparameter $a = 1$ und $b = 1/3$ gewählt.

einen dualen Gabor Frame zum Frame $\{E_{mb}T_{na}g\}$. Ist g symmetrisch, so ist auch h symmetrisch.

B-Splines erfüllen die Voraussetzungen beider Theoreme, und es ist somit möglich, auf einfache Weise basierend auf B-Splines Gabor Framepaare zu generieren (siehe Abb. 2 für ein Beispiel für drei mögliche duale Fenster zum B-Spline N_2).

Implementierung und Numerische Beispiele

Die oben erwähnte Konstruktion erlaubt es (B-Spline) Frames im Vektorraum $L^2(\mathbb{R})$ der auf ganz $\mathbb{R}$ quadratisch integrierbaren Funktionen zu konstruieren. In praktischen Anwendungen werden aber in der Regel nur beschränkte Intervalle[1] betrachten, in vielen Fällen sind Funktionen in diesen Intervallen nur auf einem Abtastgitter gegeben. In [3] werden einige Methoden vorgestellt, wie durch Periodisierung und Abtastung basierend auf Frames in $L^2(\mathbb{R})$ Frames auf dem endlich dimensionalen Vektorraum $\mathbb{C}^N$ der komplexwertigen Vektoren der Länge N konstruieren werden können. Diese Umformulierung hat unter anderem den Vorteil, dass die Koeffizienten im abgetasteten Raum $\mathbb{C}^N$ durch ein einfaches skalares Vektorprodukt berechnet werden können

$$\langle \mathbf{f}, \mathbf{g}\rangle = \mathbf{g}^H\mathbf{f} = \sum_{j=1}^{N} \bar{g}_j f_j, \tag{11}$$

während im originalen Vektorraum das Produkt durch ein Integral gegeben ist:

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)\bar{g}(x)dx. \tag{12}$$

$\bar{g}(x)$ bezeichnet hierbei den konjugiert komplexen Wert zu $g(x)$, $f_j = f(x_j)$.

Im originalen Vektorraum wäre es notwendig eine passende Quadraturformel zu verwenden. Bei der Quadratur sollte besonders auf die Exponentialfunktionen, die durch die Modulation des Fensters entstehen, Rücksicht genommen werden. Es empfehlen sich daher spezielle Quadratur-Verfahren für hoch oszillierende Funktionen wie z.B. Filon Quadratur. Im Prinzip wäre es möglich diesen Ansatz zu verwenden, und die Framefunktionen

[1] Als Beispiel wählen wir das Intervall $[0, L]$.

einfach am Rand des zu betrachteten Intervalls abzuschneiden, bzw. die Funktion und die Frameelemente außerhalb dieses Intervalls zu ignorieren. Die Formulierung in $\mathbb{C}^N$ hat den Vorteil, dass keine Quadratur verwendet werden muss, nimmt im Gegenzug dafür aber an, dass die Fenster, mit denen Frame und duale Frames konstruiert werden, periodisiert sind (siehe z.B. [3]). Darüber hinaus ist es auch notwendig, Frameparameter und Abtastgitter aufeinander abzustimmen.

Alle Framefunktionen für die in Folge verwendeten Beispiele sind für $\mathbb{C}^N$ konstruiert, als Fensterfunktion wird das abgetastete periodisierte B-Spline N_2 (siehe Abb. 2) verwendet. Alle Funktionen werden im Intervall $[0, 3)$, das mit 300 Punkten abgetastet wird, betrachtet. Für die Konstruktion der dualen Fenster werden die Frameparameter $a_0 = 1$ und $b_0 = \frac{1}{3}$ angenommen. Durch die Transformation von $L^2(\mathbb{R}) \to \mathbb{C}^N$ ergeben sich damit die Parameter $a = 100$ ($\Delta x = 1$ entspricht 100 Samples) und $b = 1$ (siehe auch [3]). Insgesamt besteht der Frame dadurch aus 900 verschiedenen Framefunktionen und die untere und obere Frameschranken (siehe Eq. (5)) ergeben $A = 1.5$ und $B = 3.0$. Als Fensterfunktionen für die dualen Frames wurden das unsymmetrische Fenster aus Theorem 1 als auch das symmetrische Fenster aus Theorem 2 sowie das kanonische duale Fenster betrachtet. Zusammengefasst bestehen die numerischen Experiment aus den folgenden Punkten, `g` ist das B-Spline N_2, `gd` bezeichnet das jeweilige duale Fenster:

- Gitter: `x = linspace(0,3,301); x(end) = []`
- Frame: `F = createframe(g, 100, 1, x)`, matrix `nsamples` $\times$ `nframeelemente`
- Dual: `Fd1 = createframe(gd, 100, 1, x)`
- Kan. Dual Frame: `Fdc = inv(F*F')*F`
- Approximierende Funktion: `f = H0`
- Berechung der Frame Koeffizienten: `c1 = Fd1' * f`
- Fehler: `abs(f - F*c1)` $\approx$ `0`

Die Eigenschaften der Frames wurde an zwei Funktionen getestet. Als erste Testfunktion wurde die Hankelfunktion

$$f(x) = H_0(10|e^{2\pi i x/3} - 1.5|) \tag{13}$$

betrachtet. Diese Funktion kann als der akustische Schalldruck auf einem Kreis mit Radius 1, der durch eine punktförmige Schallquelle im Punkt $\mathbf{x}_0 = [1.5, 0]$ bei einer Wellenzahl von $k = 10$ entsteht, interpretiert werden.

Der Fehler zwischen Zielfunktion und Darstellung mittels der Framefunktionen ist bei Verwendung aller 900 Framefunktionen im Rundungsfehlerbereich, wobei es keinen Unterschied macht, mit welchem dualen Frame die Framekoeffizienten berechnet wurden. Es stellt sich jedoch die Frage, ob zur effizienten Darstellung alle 900 Funktionen wirklich notwendig ist. Betrachten wir nur die Frameelemente mit den 100, bzw. 200 im Absolutwert größten Koeffizienten, sehen wir, dass bereits mit dieser Anzahl eine akzeptierbare Genauigkeit erreicht werden

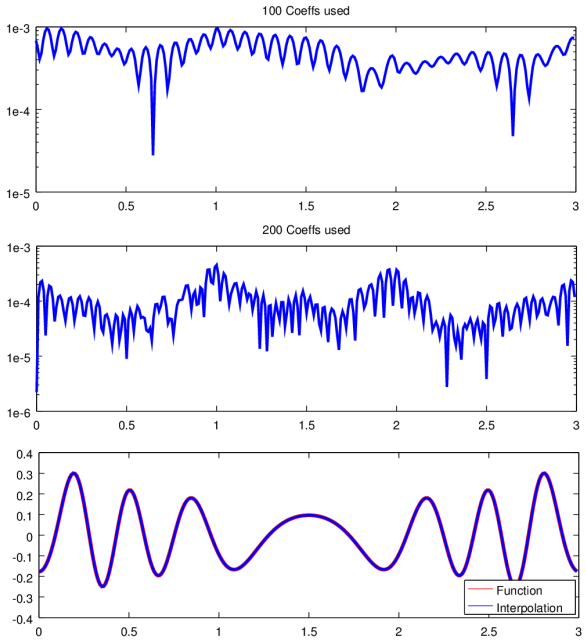

Abbildung 3: Fehler der Darstellung der Hankelfunktion mit den 100 größten Framekoeffizienten (oben), und den 200 höchsten Koeffizienten (mitte), und die Zielfunktion (unten). Als duales Fenster wurde das unsymmetrische Fenster nach Theorem 1 verwendet ($a = 1, b = \frac{1}{3}$).

kann. Darüber hinaus ist es auch ersichtlich, dass es einen nicht unwesentlichen Unterschied macht, ob die Framekoeffizienten mit dem "symmetrischen" dualen Frame (Theorem 2) oder mit dem "unsymmetrischen" dualen Frame (Theorem 1) berechnet wurden. Beim Vergleich von Abb. 3 und Abb. 4 fällt auf, dass der Approximationsfehler im Fall, dass die Koeffizienten mit dem "symmetrischen" dualen Frame berechnet werden, "schneller" kleiner werden, als mit dem "nichtsymmetrischen" dualen Frame. In beiden Abbildungen wird der Approximationsfehler bei Verwendung von nur der 100, bzw. 200 im Absolutwert größten Koeffizienten und die Zielfunktion dargestellt. Dieser Unterschied wird klar, wenn man einen Blick auf die verschiedenen Framekoeffizienten wirft (siehe Abb. 5). Während die Koeffizienten, die mit dem "symmetrischen" dualen Frame berechnet werden (rote 'o'), sehr schnell abfallen, sind die Framekoeffizienten, die mit dem "unsymmetrischen" dualen Frame (blau 'x') und dem kanonischen dual Frame (magenta '*') berechnet werden, nur sehr schwach abfallend.

Ein anderes Bild ergibt sich, wenn wir Funktionen approximieren wollen, die Unstetigkeiten aufweisen. Wird die Funktion

$$f(x) = \frac{x+1}{\lfloor x \rfloor + 1} \sin(2\pi \frac{3}{4} x), \tag{14}$$

die an den Stellen $x = 1$ und $x = 2$ und an den Intervallenden Unstetigkeiten aufweist, angenähert, ergibt sich bei Verwendung der höchsten 200 Koeffizienten (berechnet mit dem "symmetrischen" dual Frame) einen um einiges höheren Fehler (siehe Abb. 6) besonders rund um

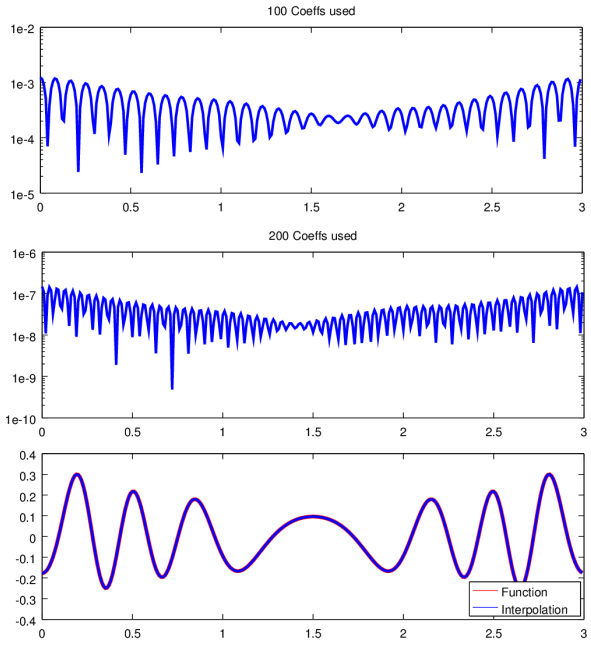

Abbildung 4: Fehler der Darstellung der Hankelfunktion mit den 100 größten Framekoeffizienten (oben), und den 200 höchsten Koeffizienten (mitte), und die Zielfunktion (unten). Als duales Fenster wurde das symmetrische Fenster nach Theorem 2 verwendet ($a = 1, b = \frac{1}{3}$).

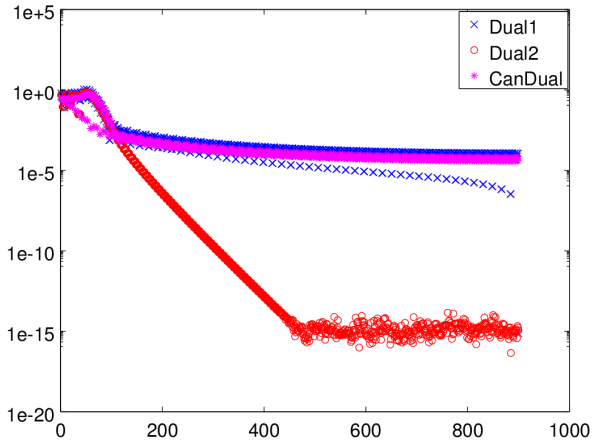

Abbildung 5: Absolutwert der verschiedenen Framekoeffizienten für die Darstellung der Hankelfunktion. Die Koeffizienten wurden mit dem "nichtsymmetrischen" dualen Frame (blau 'x'), mit dem "symmetrischen" dualen Frame (rot 'o'), und mit dem kanonischen dual Frame (magenta '*') berechnet

die Unstetigkeiten. Um einen Fehler im Rundungsfehlerbereich zu erreichen, müssen alle 900 Frameelemente benutzt werden.

Zusammenfassung

Gabor Frames, die auf B-Splines basieren, bieten eine interessante Alternative zur Darstellung und Interpolation von Funktionen mit schwingenden Komponenten, wie sie zum Beispiel in Verbindung mit Lösungen der Helmholtz-

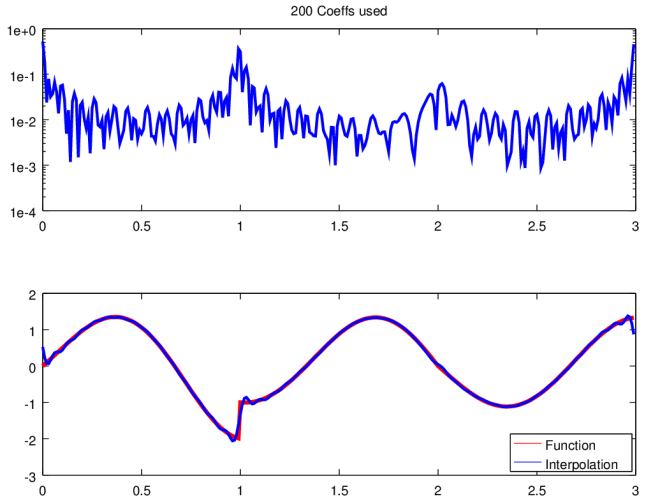

Abbildung 6: Fehler zwischen der unstetigen Zielfunktion und der Framedarstellung, bei der nur die im Absolutwert größten 200 Framekoeffizienten benutzt wurden. Die Koeffizienten wurden mit Hilfe des "symmetrischen" dualen Frames berechnet.

Gleichung auftreten. B-Splines als Fensterfunktionen haben den Vorteil, dass Frameparameter einfach bestimmt werden können, und dass es einfache Konstruktionen für duale Fenster gibt, mit deren Hilfe die Koeffizienten der Darstellung ermittelt werden können. Unstetige Funktionen könnenn zwar immer noch mit diesen Gabor Frames dargestellt werden, aber die Darstellung ist relativ ineffizient, weil alle Frameelemente dazu benutzt werden müssen, um die Unstetigkeiten aufzulösen. In diesem Fall würde es sich empfehlen, einen Frame zu benutzen, der zusätzlich auch noch Wavelet-Eigenschaften aufweist (z.B. α-modulations Frames [4]).

Danksagung

Dieses Arbeit wurde im Zuge des FWF-Projekts: BIOTOP, Adaptive Wavelet and Frame techniques for acoustic BEM (I-1018-N25) durchgeführt

Literatur

[1] Christensen, O.: Frames and Basis. An Introductory Course, Birkenhäuser, 2008

[2] Christensen, O.: B-Spline Generated Frames, in: Forster, B., Massopust P. (Edt), Four Short Courses on Harmonic Analysis, Applied an Numerical Hamonic Analsysis, Birkenhäuser, Boston

[3] Sondergaard, P.L.: Gabor Frames by Sampling and Periodization, Advances in Computational Mathematics, 27(4) (2007), 355–373

[4] Speckbacher, M., Bayer, D., Dahlke, S., Balazs, P.: The α-Modulation Transform: Admissibility, Coorbit Theory and Frames of Compactly Supported Functions, arXiv:1603.00324