Pave the way with circles: Efficient algorithms for the sampled short-time Fourier transform on nonseparable lattices

N. Holighaus
joint work with C. Wiesmeyr and P.L. Søndergaard

Acoustics Research Institute (ARI), Numerical Harmonic Analysis Group (NuHAG)

November 1, 2013

Overview

1. Short-time Fourier transforms and frames
2. Nonseparable sampling sets
3. Computational aspects

Short-time Fourier transforms

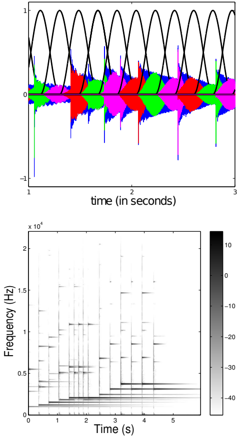


Figure: Signal and STFT

Short-time Fourier transform

The *short-time Fourier transform* of $f \in \mathbb{C}^L$, with respect to the *window* $g \in \mathbb{C}^L$ is defined as

$$\begin{aligned} \mathcal{V}_g f(m,n) &= \langle f, \mathbf{M}_m \mathbf{T}_n g \rangle \\ &= \mathcal{F}(f \mathbf{T}_n \overline{g})(m), \ (n,m)^T \in \mathbb{Z}_L^2. \end{aligned}$$

Here, $\mathbf{T}_n$ and $\mathbf{M}_n$ denote circular translation $\mathbf{T}_n g(l) = g\,(\mathrm{mod}(l-n, L))$ and modulation $\mathbf{M}_m g(l) = g(l) e^{2\pi i m l / L}$, respectively.

Sampling the STFT

The STFT is highly redundant and usually only a subset of the coefficients is computed.

- Standard sampling considers a time step a and frequency step b, leading to sampling sets of the form $a\mathbb{Z}_L \times b\mathbb{Z}_L$.
- Efficient algorithms, based on FFT or matrix factorization, exist for this case.

Goal of the presentation: Generalized sampling on subgroups $\Lambda \leq \mathbb{Z}_L \times \mathbb{Z}_L$ has the potential to improve representation quality and standard algorithms can be used at little additional computational cost.

Lattices

The sampling sets can be visualized in the (2D) time-frequency plane, forming a *lattice*. We call a lattice *separable*, if it is of the form $a\mathbb{Z}_L \times b\mathbb{Z}_L$ and *nonseparable* otherwise.

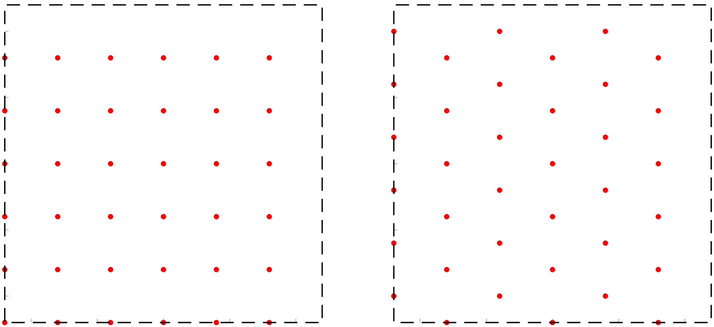
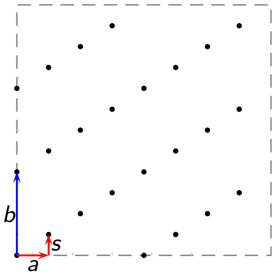


Figure: A separable lattice (left) and an example of a nonseparable lattice with the same time and frequency steps.

Describing nonseparable lattices



Any lattice Λ in $\mathbb{Z}_L \times \mathbb{Z}_L$ can be described by a time step a, frequency step b and frequency offset s.

Alternatively, we can categorize Λ by its lattice type (λ_1, λ_2) where $\lambda_1/\lambda_2 = s/b$ and λ_1, λ_2 are coprime.

In any case, we denote the number of frequency and time steps by

$$M = L/b \quad \text{and} \quad N = L/a. \tag{1}$$

An alternative point of view

A Gabor system on Λ is the collection of functions

$$\mathcal{G}(g,\Lambda) = (g_\lambda)_{\lambda\in\Lambda}, \text{ where } g_\lambda = \mathbf{M}_m\mathbf{T}_n g, \lambda = (n,m)^T. \tag{2}$$

Stable invertibility of the sampled STFT on Λ is equivalent to $\mathcal{G}(g,\Lambda)$ being a frame, i.e. a spanning set such that $0 < A \leq B < \infty$ exist with

$$A\|f\|^2 \leq \sum_{\lambda\in\Lambda} |\langle f, g_\lambda\rangle|^2 \leq B\|f\|^2. \tag{3}$$

The frame bound ratio B/A is a measure of the systems quality and how well signals are represented by it. A frame is called tight, if $A = B$.

More on frames

A Gabor frame is associated with the (invertible) *frame operator*

$$\mathbf{S}_{\mathcal{G}(g,\Lambda)} f = \sum_{\lambda \in \Lambda} \langle f, g_\lambda \rangle g_\lambda, \text{ for all } f \in \mathbb{C}^L. \tag{4}$$

There exist *dual windows* $h \in \mathbb{C}^L$ such that

$$\begin{aligned} f(l) &= \sum_{\lambda \in \Lambda} \langle f, g_\lambda \rangle h_\lambda(l) \\ &= \sum_{n=0}^{N-1} \mathbf{T}_n h(l) \sum_{m=0}^{M-1} \mathcal{F}(f \mathbf{T}_n \overline{g})(mb + ns) e^{2\pi i (mb+ns)l/L}, \text{ for all } f \in \mathbb{C}^L. \end{aligned} \tag{5}$$

A particular dual window is so-called *canonical dual* $\tilde{g} = \mathbf{S}_{g,\Lambda}^{-1} g$. If $\mathcal{G}(g, \Lambda)$ is *tight*, then $\tilde{g} = A^{-1} g$.

Lattices and time frequency concentration

STFT windows are usually designed without considering the lattice, resulting in suboptimal frame bounds and dual frames. Matching the lattice to the window can improve this.

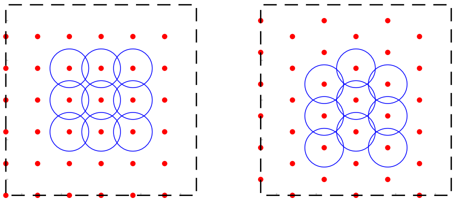


Figure: Interplay of TF concentration and the lattice: (left) separable, (right) (1,2)-type or quincunx lattice

Simultaneous time-frequency concentration can be quantified by the squared magnitude of the (discrete) *ambiguity function*

$$\mathbf{A}_g(m,n) = V_g g(m,n). \tag{6}$$

Frame quality as a covering problem

Proposition

Let $\Lambda = A\mathbb{Z}_L^2$ be a lattice. If $\mathcal{G}(g,\Lambda)$ is a frame with frame bounds $0 < A \leq B < \infty$, then

$$B/A \geq \frac{\max\limits_{m,n\in\{0,\dots,L-1\}} \Pi(n,m)}{\min\limits_{m,n\in\{0,\dots,L-1\}} \Pi(n,m)}, \tag{7}$$

where

$$\Pi(n,m) = \sum_{k=0}^{N-1}\sum_{l=0}^{M-1} |\mathbf{A}_g(n-ka, m-lb-ks)|^2. \tag{8}$$

Moreover, $\Pi(n,m) = A > 0$ for all n, m is equivalent to $\mathcal{G}(g,\Lambda)$ being a tight frame

From window to covering problem

A a consequence of the previous proposition, a sampling of the STFT only has a chance of forming a *good* frame, if the *periodized ambiguity function* is close to constant.

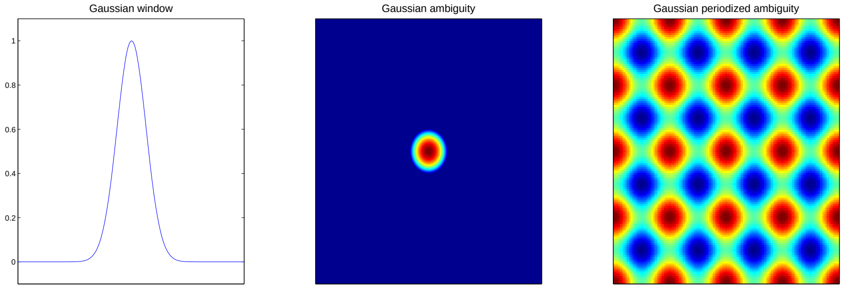
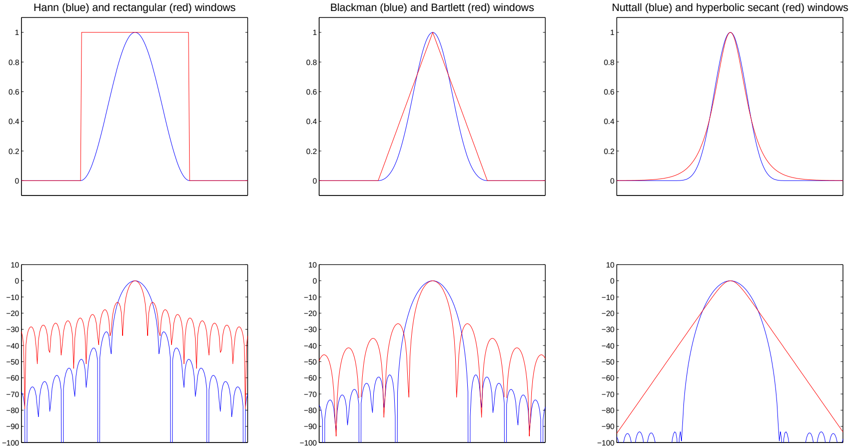


Figure: A Gaussian window, its ambiguity function and periodized ambiguity function of a separable lattice.

Some well-known window prototypes



Their time-frequency concentration

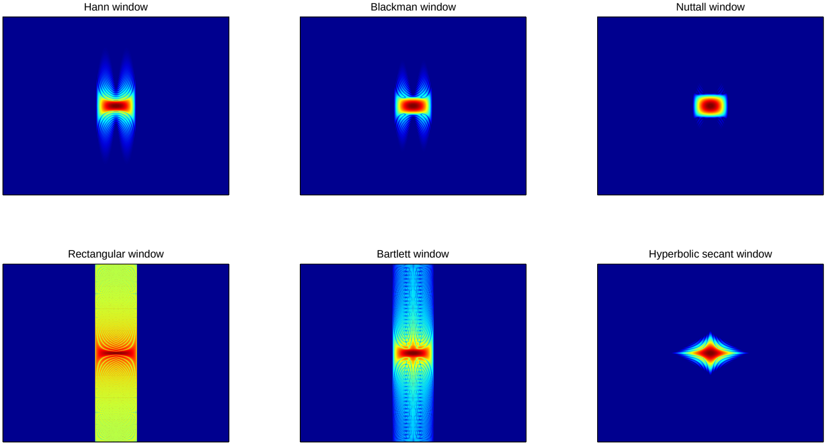


Figure: Ambiguity functions of various windows

Periodized ambiguity functions - I

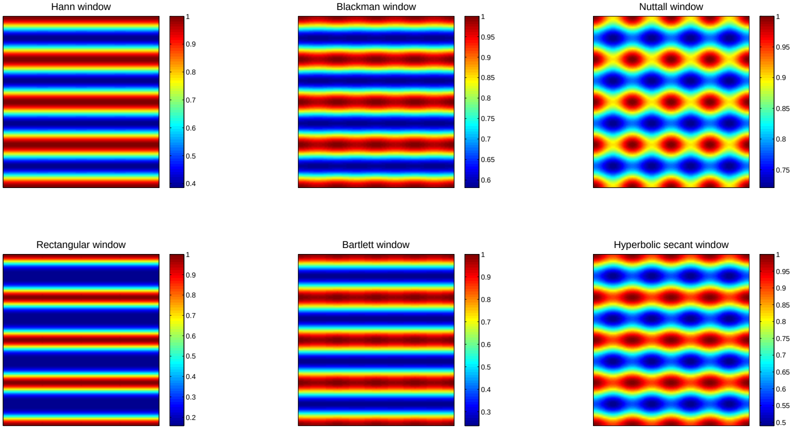


Figure: Periodized ambiguity functions on the rectangular (0/1) lattice

Periodized ambiguity functions - II

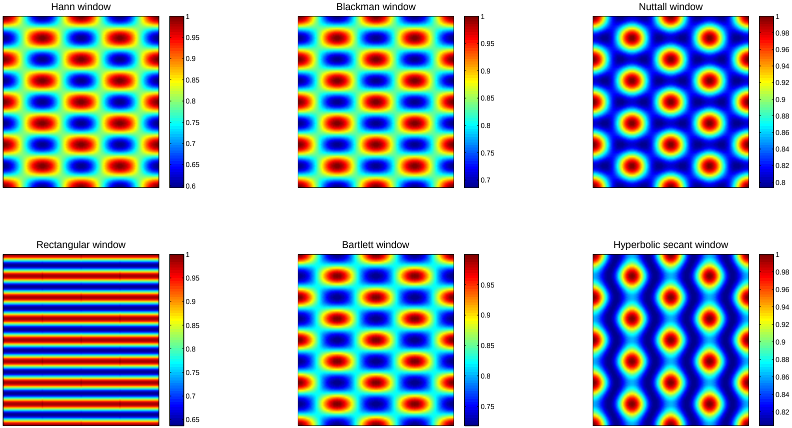


Figure: Periodized ambiguity functions on the quincunx (1/2) lattice

Periodized ambiguity functions - III

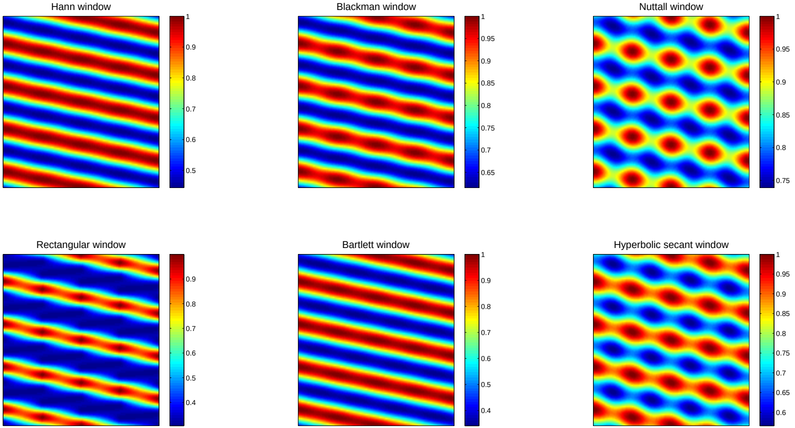
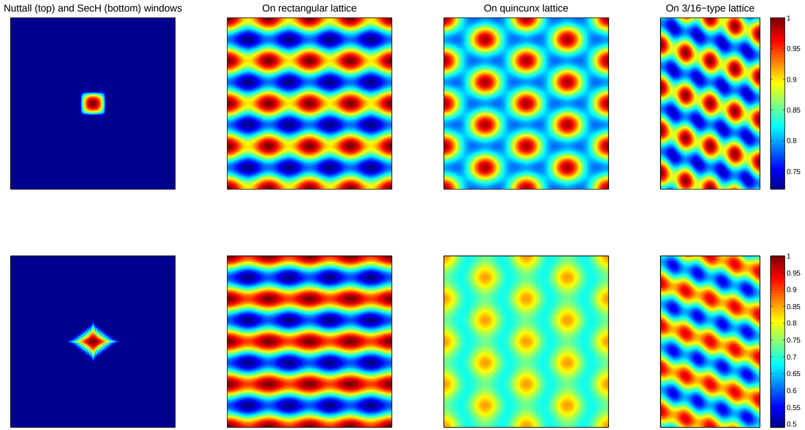
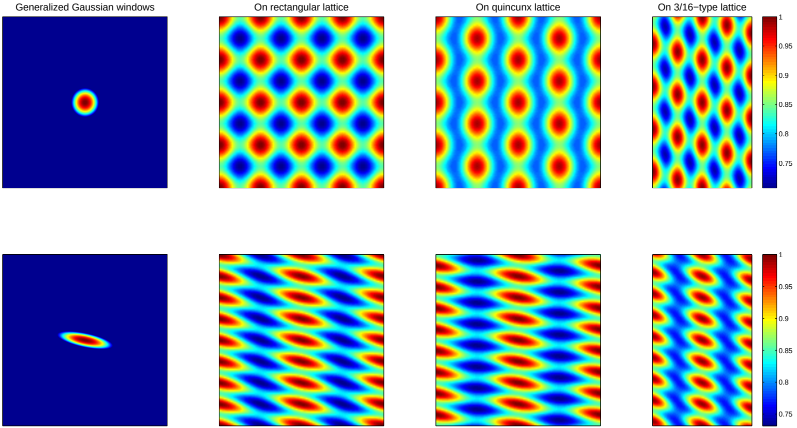


Figure: Periodized ambiguity functions on the '3/16' lattice

Different configuration, same window - I



Different configuration, same window - II



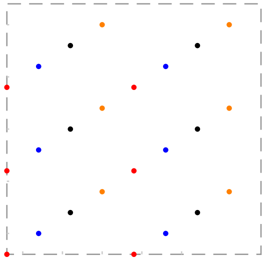
Computation of STFT on lattices

Both STFT analysis and inversion on general lattices can be computed using classical, efficient algorithms for separable lattices plus some inexpensive pre- and post-processing. Two main method types exist to achieve this:

(i) Multiwindow methods

(ii) Deformation methods

Multiwindow methods



Multiwindow techniques decompose a lattice into cosets of the sparser, separable lattice $\tilde{\Lambda} = a\lambda_2\mathbb{Z}_L \times b\mathbb{Z}_L$ and use the following equality, which holds up to a phase factor.

$$\mathcal{G}(g,\Lambda) = \{\mathbf{M}_{mb+ns}\mathbf{T}_{na}g\}_{m,n}$$

$$= \bigcup_{k=0}^{\lambda_2-1} \{\mathbf{M}_{mb}\mathbf{T}_{n\lambda_2 a}\mathbf{M}_{\mathrm{mod}(ks,b)}\mathbf{T}_{ka}g\}_{m,n}$$

$$= \bigcup_{k=0}^{\lambda_2-1} \mathcal{G}(\mathbf{M}_{\mathrm{mod}(ks,b)}\mathbf{T}_{ka}g,\tilde{\Lambda})$$

Deformation methods - I

Deformation methods use a correspondence between certain unitary operators on the time-frequency plane and the signal space, known under the name *metaplectic representation*.
This allows us to find an equivalence between a given system on Λ and another system on a separable lattice $\bar{\Lambda}$.

Deformation methods - II

More explicitly, there exist a pair of unitary operators $\mathbf{D}$ on $\mathbb{Z}_L^2$ and $\mathbf{U}_D$ on $\mathbb{C}^L$, such that

$$\mathbf{D}\bar{\Lambda} = \{D(n,m)^T \ : \ (n,m)^T \in \bar{\Lambda}\} = \Lambda \tag{9}$$

and moreover

$$\tilde{g} = \mathbf{S}_{g,\Lambda}^{-1} g = \mathbf{U}_D^{-1} \mathbf{S}_{\mathbf{U}_D^{-1} g,\bar{\Lambda}}^{-1} \mathbf{U}_D g \text{ and} \tag{10}$$

$$V_g f(m,n) = \phi(m,n) V_{\mathbf{U}_D^{-1} g} \mathbf{U}_D^{-1} f(\bar{m},\bar{n}), \tag{11}$$

with $(\bar{n},\bar{m})^T = \mathbf{D}^{-1}(n,m)^T$. Here, ϕ is just a phase factor and $\mathbf{U}_D$ is some composition of (inverse) Fourier transforms and pointwise scalar multiplication.

Efficiency of the proposed algorithms

We compare the multiwindow method with two distinct deformation methods, denoted as *Smith* and *Shear*.

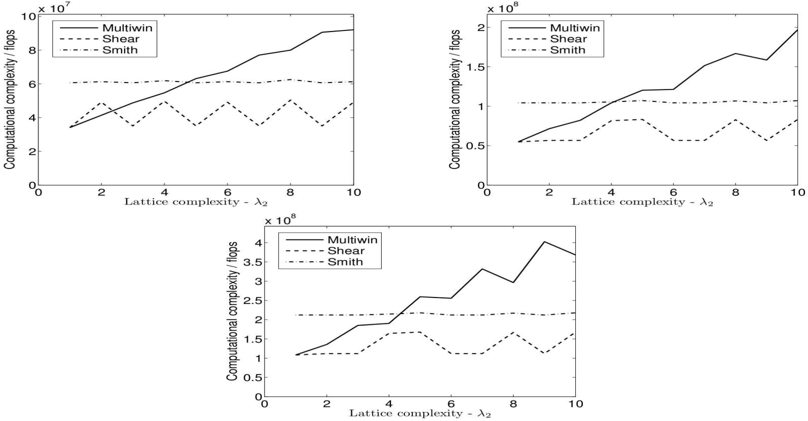


Figure: Performance of the algorithms for 3 different choices of a, b and L. The value at $\lambda_2 = 1$ corresponds to the separable case.

Conclusion

- Good pairings of window and lattice have the potential to improve frame quality and therefore processing results.
- For standard windows, the (1,2)-type (quincunx) lattice seems most promising.
- The additional cost of computing the operations associated with sampled STFTs on nonseparable lattices amounts to some inexpensive pre- and post-processing.
- Efficient implementations are freely available in the LTFAT toolbox.
- From a computational or frame theoretic point of view, there are *no good reasons not to* work on general lattices.

Thank you for your attention

LTFAT URL: http://ltfat.sourceforge.net/

Selected references:

[1] O. Christensen, "An Introduction to Frames and Riesz Bases," ser. Appl. and Numer. Harmon. Anal. Boston: Birkhuser, 2003.

[2] H. G. Feichtinger, M. Hazewinkel, N. Kaiblinger, E. Matusiak, and M. Neuhauser, "Metaplectic operators on $\mathbb{C}^n$," Quart. J. Math. Oxford Ser., vol. 59, no. 1, pp. 15-28, 2008.

[3] K. Gröchenig, "Foundations of Time-Frequency Analysis," ser. Appl. Numer. Harmon. Anal. Boston, MA: Birkhäuser Boston, 2001.

[4] G. Kutyniok and T. Strohmer, "Wilson bases for general time-frequency lattices," SIAM J. Math. Anal., vol. 37, no. 3, pp. 685-711, 2005.

[5] P. L. Søndergaard, "Efficient Algorithms for the Discrete Gabor Transform with a long FIR window," J. Fourier Anal. Appl., vol. 18, no. 3, pp. 456-470, 2012.

[6] C. Wiesmeyr, N. Holighaus, and P. Søndergaard, "Efficient algorithms for the discrete Gabor transform on a nonseparable lattice," IEEE Trans. Sig. Proc., vol. 61, no. 20, pp. 5131-5142, 2013.