A survey of uncertainty principles and some signal processing applications

Benjamin Ricaud · Bruno Torrésani

Abstract The goal of this paper is to review the main trends in the domain of uncertainty principles and localization, highlight their mutual connections and investigate practical consequences. The discussion is strongly oriented towards, and motivated by signal processing problems, from which significant advances have been made recently. Relations with sparse approximation and coding problems are emphasized.

1 Introduction

Uncertainty inequalities generally express the impossibility for a function (or a vector in the discrete case) to be simultaneously *sharply concentrated* in two different

B. Ricaud (✉)
Signal Processing Laboratory 2, Ecole Polytechnique Fédérale de Lausanne (EPFL), Station 11, 1015 Lausanne, Switzerland
e-mail: benjamin.ricaud@epfl.ch

B. Torrésani
Aix-Marseille Université, CNRS, Centrale Marseille, LATP, UMR7353, 13453 Marseille, France
e-mail: bruno.torresani@univ-amu.fr

representations, provided the latter are *incoherent enough*. Such a loose definition can be made concrete by further specifying the following main ingredients:

– **A global setting**, generally a couple of Hilbert spaces (of functions or vectors) providing two representations for the objects of interest (e.g. time and frequency, or more general phase space variables).
– **An invertible linear transform** (operator, matrix) mapping the initial representation to the other one, without information loss.
– **A concentration measure** for the elements of the two representation spaces: variance, entropy, L^p norms,...

Many such settings have been proposed in the literature during the last century, for various purposes. The first formulation was proposed in quantum mechanics where the uncertainty principle is still a major concern. However it is not restricted to this field and appears whenever one has to represent functions and vectors in different manners, to extract some specific information. This is basically what is done in signal processing where the uncertainty principle is of growing interest.

The basic (quantum mechanical) prototype is provided by the so-called Robertson-Schrödinger inequality, which establishes a lower bound for the product of variances of any two self-adjoint operators on a generic Hilbert space. The most common version of the principle is as follows:

Theorem 1 *Let $f \in \mathcal{H}$ (Hilbert space), with $\|f\| = 1$. Let A and B be (possibly unbounded) self-adjoint operators on $\mathcal{H}$ with respective domains $D(A)$ and $D(B)$. Define the mean and variance of A in state $f \in D(A)$ by*

$$e_f(A) = \langle Af, f \rangle , \qquad v_f(A) = e_f\left(A^2\right) - e_f(A)^2 .$$

Setting $[A, B] = AB - BA$ and $\{A, B\} = AB + BA$, we have $\forall f \in D(AB) \cap D(BA)$,

$$v_f(A)v_f(B) \geq \frac{1}{4}\left[\left|e_f([A, B])\right|^2 + \left|e_f\left(\{A - e_f(A), B - e_f(B)\}\right)\right|^2\right] .$$

The quantities $v_f(A)$ and $v_f(B)$ can also be interpreted as the variances of two representations of f given by its projection onto the respectives bases of (possibly generalized) eigenvectors of A and B. From the self-adjointness of A and B, there exists a unitary operator mapping one representation to the other.

The proof of this result is quite generic and carries over many situations. However, the choice of the variance to measure concentration properties may be quite questionable in a number of practical situations, and several alternatives have been proposed and studied in the literature.

The goal of this paper is to summarize a part of the literature on this topic, discuss a few recent results and focus on specific signal processing applications. We shall first describe the continuous setting, before moving to discrete formulations and emphasizing the main differences. Given the space limitations, the current paper cannot be exhaustive. We have selected a few examples which highlight the structure and some important aspects of the uncertainty principle. We refer for example to [15] for

a very good and complete account of classical uncertainty relations, focused on time-frequency uncertainty. An information theory point of view of the uncertainty principle may be found in [5] and a review of entropic uncertainty principles has been given in [30]. More recent contributions, mainly in the sparse approximation literature, introducing new localization measures will be mentioned in the core of the paper.

2 Some fundamental aspects of the uncertainty principle

2.1 Signal representations

The uncertainty principle is usually understood as a relation between the simultaneous spreadings of a function and its Fourier transform. More generally, as expressed in Theorem 1 an uncertainty principle also provides a relation between any two representations of a function, here the ones given by its projections onto the (possibly generalized) eigenbases of A and B. Can a representation be something else than the projection onto a (generalized) eigenbasis? The answer is yes: representations can be made by introducing frames. A set of vectors $\mathcal{U} = \{u_k\}_k$ in a Hilbert space $\mathcal{H}$ is a frame of $\mathcal{H}$ if for all $f \in \mathcal{H}$:

$$A\|f\|^2 \le \sum_k |\langle f, u_k \rangle|^2 \le B\|f\|^2, \tag{1}$$

where A, B are two constants such that $0 < A \le B < \infty$. Since $A > 0$, any $f \in \mathcal{H}$ can be recovered from its frame coefficients $\{\langle f, u_k \rangle\}_k$. This is a key point: in order to compare two representations, information must not be lost in the process. Orthonormal bases are particular cases of frames for which $A = B = 1$ and the frame vectors are orthogonal.

Denote by $U : f \in \mathcal{H} \to \{\langle f, u_k \rangle\}_k$ the so-called analysis operator. U is left invertible, which yields inversion formulas of the form

$$f = \sum_k \langle f, u_k \rangle \tilde{u}_k \,,$$

where $\widetilde{\mathcal{U}} = \{\tilde{u}_k\}_k$ is an other family of vectors in $\mathcal{H}$, which can also be shown to be a frame, termed dual frame of $\mathcal{U}$. Choosing as left inverse the Moore-Penrose pseudo-inverse $U^{-1} = U^\dagger$ yields the so called canonical dual frame $\widetilde{\mathcal{U}}^\circ = \{\tilde{u}_k^\circ = (UU^*)^{-1}u_k\}_k$, but other choices are possible.

The uncertainty principle can be naturally extended to frame representations, i.e. representations of vectors $f \in \mathcal{H}$ by their frame coefficients. As before, uncertainty inequalities limit the extend to which a vector can have two arbitrarily concentrated frame representations. Since variances are not necessarily well defined in such a case, other concentrations measures such as entropies have to be used. For example, bounds for the entropic uncertainty principle are derived in [28].

2.2 The mutual coherence: how different are two representations?

A second main aspect of uncertainty inequalities is the heuristic remark that *the more different the representations, the more constraining the bounds*. However, one needs

to be able to measure how different two representations are. This is where the notion of coherence enters.

Let us first stick to frames in the discrete setting. Let $\mathcal{U} = \{u_k\}_k$ and $\mathcal{V} = \{v_k\}_k$ be two frames of $\mathcal{H}$. Let us define the operator $T = VU^{-1}$ which allows one to pass from the representation of f in $\mathcal{U}$ to the one in $\mathcal{V}$. It is given by:

$$TUf(j) = \left\langle \sum_k (Uf)(k)\, \tilde{u}_k, v_j \right\rangle = \sum_k (Uf)(k)\, \langle \tilde{u}_k, v_j \rangle.$$

This relation shows that in finite dimension T is represented by a matrix $G = G(\widetilde{\mathcal{U}}, \mathcal{V})$ (the cross Gram matrix of $\widetilde{\mathcal{U}}$ and $\mathcal{V}$) defined by $G_{j,k} = \langle \tilde{u}_k, v_j \rangle$. The matrix G encodes the differences of the two frames. The latter can be measured by various norms of T, among which the so-called *mutual coherence*:

$$\mu = \mu(\widetilde{\mathcal{U}}, \mathcal{V}) = \max_{j,k} \left| \langle \tilde{u}_k, v_j \rangle \right| = \max_{j,k} \left| G_{j,k} \right| = \|T\|_{\ell^1 \to \ell^\infty} . \qquad (2)$$

This quantity encodes (to some extend) the algebraic properties of T.

Remark 1 This particular quantity (norm) may be generalized to other kinds of norms which would be, depending on the setting, more appropriate for the estimation of the correlation between the two representations. Indeed, it is the characterization of the matrix which quantify how close two representations actually are.

Remark 2 In the standard case (N-dimensional) where the uncertainty is stated between the Kronecker and Fourier bases, $|T_{j,k}| = 1/\sqrt{N}$ for all j, k. These bases are said to be mutually unbiased and $\mu = 1/\sqrt{N}$ is the smallest possible value of μ.

In the case of the entropic uncertainty principle, the demonstration of the inequality is based on the Riesz interpolation theorem and it rely on bounds of T as an operator from $\ell^1 \to \ell^\infty$ and from $\ell^2 \to \ell^2$ (see Section 4.3). As we shall see, this notion of mutual coherence appears in most of the uncertainty relations. A noticeable exception is the variance-based uncertainty principle. In this case it is replaced by the commutation relation between the two self-adjoint operators and the connection with the coherence is not straightforward.

2.3 The notion of phase space

Standard uncertainty principles are associated with pairs of representations: time localization vs frequency localization, time localization vs scale localization,... However, in some situations, it is possible to introduce directly a *phase* space, which involves jointly the two representation domains, in which (non-separable) uncertainty principles can be directly formulated: joint time-frequency space, joint time-scale space,...

Uncertainty principles associated with pairs of representations often have counterparts defined directly in the joint space. We shall see a few examples in the course of the current paper. In such situations, the mutual coherence is replaced with a notion of phase space coherence.

3 Uncertainty inequalities in continuous settings: a few remarkable examples

To get better insights on the uncertainty principle we state here a few remarkable results which illustrate the effect of changing (even slightly) the main ingredients. This helps understanding the choices made below in discrete settings.

The most popular and widespread form of the uncertainty principle uses the variance as spreading measure of a function and its Fourier transform. This leads to the inequality stated in Theorem 1, where $A = X$ is the multiplication operator $Xf(t) = tf(t)$ and $B = P = -i\partial_t/2\pi$ is the derivative operator. This first instance of uncertainty inequalities is associated to the so-called canonical phase space, i.e. the time-frequency, or position-momentum space. Let us first introduce some notations. Given $f \in L^2(\mathbb{R})$, denote by $\hat{f}$ its Fourier transform, defined by

$$\hat{f}(v) = \int_{-\infty}^{\infty} f(t)e^{-2i\pi vt}\, dt\, .$$

With this definition, the Fourier transformation is an unitary operator $L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$. The classical uncertainty inequalities state that for any $f \in L^2(\mathbb{R}), f$ and $\hat{f}$ cannot be simultaneously sharply localized.

Heisenberg's inequality Let $\mathcal{H} = L^2(\mathbb{R})$ and consider the self-adjoint operators X and P, defined by $Xf(t) = tf(t)$ and $Pf(t) = -if'(t)/2\pi$. X and P satisfy the commutation relations $[X, P] = i\,\mathbf{1}$, where $\mathbf{1}$ is the identity operator. For $f \in L^2(\mathbb{R})$, denote by e_f and v_f its expectation and variance (see Theorem 1):

$$e_f = e_f(X) = \frac{1}{\|f\|^2}\int_{-\infty}^{\infty} t|f(t)|^2\, dt\, , \quad v_f = v_f(X) = \frac{1}{\|f\|^2}\int_{-\infty}^{\infty} (t - e_f)^2|f(t)|^2\, dt. \tag{3}$$

Then the Robertson-Schrödinger inequality takes the form

Corollary 1 *For all $f \in L^2(\mathbb{R})$,*

$$v_f \cdot v_{\hat{f}} \geq \frac{1}{16\pi^2}\, , \tag{4}$$

with equality if and only if $f : t \rightarrow f(t) = ae^{-b(t-\mu)^2/2}$ is a Gaussian function, up to time shifts, modulations, rescalings and chirping ($a, b, \mu \in \mathbb{C}$, with $\Re(b) > 0$).

3.1 Variance time-frequency uncertainty principles on different spaces

Usual variance inequalities are defined for functions on the real line, or on Euclidian spaces. It is important to stress that these inequalities do not generalize easily to other settings, such as periodic functions, or more general functions on bounded domains.

First, the definition of mean and variance themselves can be difficult issues.[1] For example, the definition of the mean of a function on the circle S^1 is problematic. Sticking to the above notations, the operator X is not well defined on $L^2(S^1)$ because of the periodicity, the meaning of $e_f(X)$ is not clear, and does definitely not represent the mean value of f. Adapted definitions of mean and variance are required. For example, the case $\mathcal{H} = L^2(M)$, where M is a Riemannian manifold, has been studied by various authors (see [11] and references therein).

For example, in the case of the circle one definition of the mean value is given by $e_f = \arg\langle f, Ef \rangle$ where $E\psi(t) = \exp(i2\pi t)\psi(t)$ (the so-called von Mises's mean, see [4]). From this, an *angle-momentum* uncertainty inequality has been obtained in [4, 21]. Yet, additional difficulties appear: first the bound of the uncertainty principle is modified (compared to the $L^2(\mathbb{R})$ case) and depends non trivially on the function f involved. This implies that functions whose uncertainty product attains the bounds are not necessarily minimizers and the strict positivity of the lower bound may not be guaranteed. The answer of the authors is to suggest to modify the definition of the variance (in addition to the modification of the mean).

Similar problems are encountered in various different situations, such as the affine uncertainty which we account for below. All this shows that alternatives to variance-based spreading measures are necessary. We will address these in Section 3.3 below.

3.2 Different representations

In the Robertson-Schrödinger formulation, the two representation spaces under consideration (which form the phase space) are L^2 spaces of the spectrum of two self-adjoint operators A and B. The spectral theorem establishes the existence of two unitary maps U_A and U_B mapping $\mathcal{H}$ to the two L^2 spaces; the images of elements of $\mathcal{H}$ by these operators yield the two representations, for which uncertainty inequalities can be proven. It is worth noticing that these representations can be (possibly formally) interpreted as scalar products of elements of $\mathcal{H}$ with (possibly generalized) eigenbases of A and B.

This allows one to go beyond the time-frequency representation and introduce generalized phase spaces. We shall assume that the generalized phase space is associated with self-adjoint operators $A_1, \ldots A_k$, which are infinitesimal generators of generalized translations, acting on some signal (Hilbert) space $\mathcal{H}$. Whenever two operators A_j, A_l are such that there exists a unitary transform U which turn these two operators into the standard case (operator X, P defined above), one can obtain time-frequency type uncertainty inequalities. In such cases, the lower bound is attained for specific choices of f, which are the images of gaussian functions by the unitary transformation U. We will refer to this construction as a canonization process. An example where canonization is possible can be found in Remark 3 below.

When this is not the case, the commutator $[A_j, A_l]$ is not a multiple of the identity and the lower bound generally depends on f. This implies phenomena described in

---

[1]The definition and properties of the variance (and other moments) on compact manifolds is by itself a well defined field of research named *directional statistics*.

Section 3.1. Even worse, if the spectrum of the operator $i[A_j, A_l]$ include zero then the lower bound is zero, revealing that the variance may not be a spreading measure in this case.

3.2.1 Time-scale variance inequality

The classical affine variance inequality is another particular instance of the Robertson-Schrödinger inequality: let $A = X$ and $B = D = (XP + PX)/2$ denote the infinitesimal generators of translations and dilations, acting on the Hardy space $H^2(\mathbb{R}) = \{f \in L^2(\mathbb{R}), \hat{f}(\nu) = 0 \; \forall \nu \le 0\}$, which is the natural setting here.

Explicit calculation shows that $[X, D] = iX$, and it is worth introducing the **scale transform** $f \in H^2(\mathbb{R}) \to \tilde{f}$, which is a unitary mapping $H^2(\mathbb{R}) \leftrightarrow L^2(\mathbb{R})$ defined by

$$\tilde{f}(s) = \int_0^\infty \hat{f}(\nu)e^{2i\pi s} \frac{d\nu}{\sqrt{\nu}} = \int_{-\infty}^\infty \hat{f}(e^u)e^{u/2}e^{2i\pi us} \, du \; . \tag{5}$$

The corresponding Robertson-Schrödinger inequality states

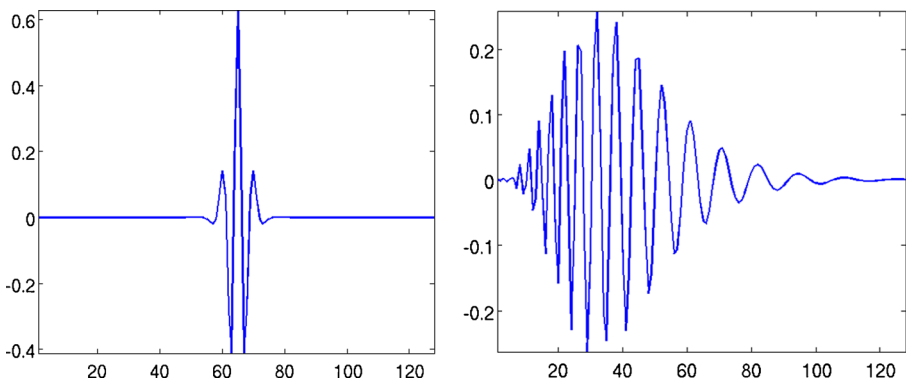**Corollary 2** *For all* $f \in H^2(\mathbb{R})$,

$$v_{\hat{f}}.v_{\tilde{f}} \ge \frac{1}{16\pi^2}e_{\hat{f}}^2 \; , \tag{6}$$

with equality if and only if $\hat{f}$ is a Klauder waveform (Fig. 1), *defined by*

$$\hat{f}(\nu) = K \exp\{a \ln(\nu) - b\nu + i(c \ln(\nu) + d)\} \; , \qquad \nu \in \mathbb{R}^+ \tag{7}$$

for some constants $K \in \mathbb{C}$, $a > -1/2$, $b \in \mathbb{R}^+$ and $c, d \in \mathbb{R}$.

It is worth noticing that the right hand side explicitly depends on f, so that the Klauder waveform, which saturates this inequality, is not necessarily a minimizer of the product of variances, as analyzed in [24].

Fig. 1 Examples of Klauder waveform (*left*) and Altes waveform (*right*)

3.2.2 Modified time-scale inequality

The above remark prompted several authors (see [14] for a review) to seek different forms of averaging, adapted to the affine geometry. This led to the introduction of adapted means and variances: for $f \in H^2(\mathbb{R})$, set

$$\tilde{e}_{\hat{f}} = \exp\left\{\frac{1}{\|\hat{f}\|^2}\int_0^\infty |\hat{f}(\nu)|^2 \ln(\nu)\,d\nu\right\} \quad \tilde{v}_{\hat{f}} = \frac{1}{\|\hat{f}\|^2}\int_0^\infty \left[\ln(\nu/\tilde{e}_{\hat{f}})\right]^2 |\hat{f}(\nu)|^2\,d\nu. \tag{8}$$

In this new setting, one obtains a more familiar inequality

Proposition 1 *For all $f \in H^2(\mathbb{R})$,*

$$\tilde{v}_{\hat{f}}.v_{\tilde{f}} \geq \frac{1}{16\pi^2}, \tag{9}$$

with equality if and only if f takes the form of an Altes waveform (Fig. 1)*, defined by*

$$\hat{f}(\nu) = K\exp\left\{-\frac{1}{2}\ln(\nu) - a\ln^2(\nu/b) + i(c\ln(\nu) + d)\right\}, \qquad \nu \in \mathbb{R}^+, \tag{10}$$

which is now a variance minimizer.

Remark 3 (*Canonization*) The connection between Klauder's construction and Altes' can also be interpreted in terms of canonization. Let $U : H^2(\mathbb{R}) \to L^2(\mathbb{R})$ denote the unitary linear operator defined by $\widehat{Uf}(\nu) = e^{\nu/2}\hat{f}(e^\nu)$, for $\nu \in \mathbb{R}_+$. The adjoint operator reads $U^*f(s) = f(\ln(s))/\sqrt{s}$ (for $s \in \mathbb{R}_+$), and it is readily verified that U is unitary. Consider now the linear operators $\tilde{X}$ and $\tilde{P}$ on $H^2(\mathbb{R})$ defined by $\tilde{X} = U^*XU$ and $\tilde{P} = U^*PU$. Simple calculations show that $\tilde{X} = D/2\pi$ and $\tilde{P} = 2\pi\ln(P/2\pi)$, these two operators being well defined on $H^2(\mathbb{R})$. Hence $\tilde{X}$ and $\tilde{P}$ satisfy the canonical commutation relations on $H^2(\mathbb{R})$:

$$[D, \ln(P)] = [D, \ln(P/2\pi)] = [\tilde{X}, \tilde{P}] = U^*[X, P]U = i\,1_H^2(\mathbb{R}).$$

Now, given any self adjoint operator A on $H^2(\mathbb{R})$, and for any $f \in H^2(\mathbb{R})$, set $g = Uf$, and one has $e_f(A) = e_g(UAU^*)$ and $v_f(A) = v_g(UAU^*)$. Therefore,

$$v_f(D).v_f(\ln(P)) = v_g(X).v_g(P) \geq \frac{1}{16\pi^2},$$

with equality if and only if g is a Gaussian function, i.e. f is an Altes wavelet.

3.3 Different dispersion measures

As stressed above, variance is not always well defined, and even when it is so, variance inequalities may not yield meaningful informations. Alternatives have been proposed in the literature, and we review some of them here. Some of then show better stability to generalizations, and will be more easily transposed to the discrete case.

3.3.1 Hirschman-Beckner entropic inequality

Following a conjecture by Everett [12] and Hirschman [18], Beckner [2] proved an inequality involving entropies. Assume $\|f\| = 1$, and define Shannon's differential entropy by

$$H(f) = \int |f(t)|^2 \ln\left(|f(t)|^2\right) dt . \tag{11}$$

Then the Hirschman-Beckner uncertainty principle states

Theorem 2 *For all $f \in L^2(\mathbb{R})$,*

$$H(f) + H(\hat{f}) \geq 1 - \ln(2) , \tag{12}$$

with equality if and only if f is a Gaussian function (up to the usual modifications).

The proof originates from the *Babenko-Beckner inequality* (also called sharp Hausdorff-Young inequality) [2]: for $f \in L^p(\mathbb{R})$, let $1/p + 1/p' = 1$; then $\|\hat{f}\|_{p'} \leq A_p \|f\|_p$, where $A_p = \sqrt{p^{1/p}/p'^{1/p'}}$. Taking logarithms after suitable normalization yields an inequality involving Rényi entropies (see below for a definition), that reduces to the Hirschman-Beckner inequality for $p = p' = 2$. As remarked in [14], for the time-scale uncertainty, the canonization trick applies in this case as well, and yields a corresponding entropic uncertainty inequality for time and scale variables.

3.3.2 Concentration on subsets, the Donoho-Stark inequalities

In [9], Donoho and Stark prove a series of uncertainty inequalities, in both continuous and discrete settings, using different concentration measures. One of these is the following: for $f \in L^2(\mathbb{R})$, and $\epsilon > 0$, f is said to be ϵ-concentrated in the measurable set U if there exists g supported in U such that $\|f - g\| \leq \epsilon$. Donoho and Stark prove

Theorem 3 *Assume that f is ϵ_T concentrated in T and $\hat{f}$ is ϵ_F concentrated in F; then*

$$|T| \cdot |F| \geq (1 - (\epsilon_T + \epsilon_F))^2 . \tag{13}$$

Remark 4 (*Gerchberg-Papoulis algorithm*) This uncertainty inequality is used to prove the convergence of the Gerchberg-Papoulis algorithm for missing samples restoration for band-limited signals, as follows. Let F, T be bounded measurable subsets of the real line. Given $x \in L^2(\mathbb{R})$ such that $Supp(\hat{x}) \subset F$, assume observations of the form

$$y(t) = \begin{cases} x(t) + n(t) & \text{if } t \notin T \\ n(t) & \text{otherwise} \end{cases}$$

where n is some noise, simply assumed to be bounded.

Denote by P_T the orthogonal projection onto L^2 signals supported by T in the time domain, and by P_F the corresponding projection in the

frequency domain. If $|F| \cdot |T| < 1$, then $\|P_T P_F\| < 1$ and x is stably recovered by solving

$$\tilde{x} = (1 - P_T P_F)^{-1} y \,,$$

where stability means $\|x - \tilde{x}\| \leq C \|n\|$.

The same paper by Donoho and Stark provides several other versions of the uncertainty principle, in view of different applications.

In a similar spirit, Benedicks theorem states that every pair of sets of finite measure (T, F) is *strongly annihilating*, i.e. there exists a constant $C(T, F)$ such that for all $f \in L^2(\mathbb{R})$,

$$\|f\|^2_{L^2(\mathbb{R}\setminus T)} + \|\hat{f}\|^2_{L^2(\mathbb{R}\setminus F)} \geq \|f\|^2 / C(T, F) \,. \tag{14}$$

We refer to [20] for more details, together with generalizations to higher dimensions as well as explicit estimates for the constants $C(T, F)$.

3.4 Non-separable dispersion measures

Traditional uncertainty principles bound joint concentration in two different representation spaces. In some situations, it is possible to define a joint representation space (phase space) and derive corresponding uncertainty principles. This is in particular the case for time-frequency uncertainty. The quantities of interest are then functions defined directly on the time-frequency plane, such as the short time Fourier transform and the ambiguity function. Given $f, g \in L^2(\mathbb{R})$, the STFT (Short time Fourier transform) of f with window g and the ambiguity function of f are respectively the functions $\mathcal{V}_g f, \mathcal{A}_f \in L^2(\mathbb{R}^2)$ defined by

$$\mathcal{V}_g f(b, \nu) = \int_{-\infty}^{\infty} f(t) \overline{g}(t - b) e^{-2i\pi \nu t} \, dt \,, \quad \mathcal{A}_f = \mathcal{V}_f f \,. \tag{15}$$

Concentration properties of such functions have been shown to be relevant in various contexts, including radar theory (see [23]) or time-frequency operator approximation theory [6]. We highlight a few relevant criteria and results.

3.4.1 L^p-norm of the ambiguity function: Lieb's inequality

E. Lieb (see [3] for example) gives bounds on the concentration of the Ambiguity function (resp. STFT). Contrary to Heisenberg type uncertainty inequalities, which privilege a coordinate system in the phase space (i.e. choose a time and a frequency axis), bounds on the ambiguity function don't. Here, concentration is measured by L^p norms, and the bounds are as follows

Theorem 4 *For all $f, g \in L^2(\mathbb{R})$,*

$$\begin{cases} \|\mathcal{A}_f\|_p \geq B_p \|f\|_2^2 & \text{for } p < 2 \\ \|\mathcal{A}_f\|_p \leq B_p \|f\|_2^2 & \text{for } p > 2 \\ \|\mathcal{A}_f\|_2 = \|f\|_2^2 \end{cases} , \quad \begin{cases} \|\mathcal{V}_g f\|_p \geq B_p \|g\|_2 \|f\|_2 & \text{for } p < 2 \\ \|\mathcal{V}_g f\|_p \leq B_p \|g\|_2 \|f\|_2 & \text{for } p > 2 \\ \|\mathcal{V}_g f\|_2 = \|g\|_2 \|f\|_2 \end{cases} \tag{16}$$

where $B_p = (2/p)^{1/p}$ is related to the Beckner-Babenko constants.

The norm $\| \cdot \|_p$ can be regarded as a diversity, or spreading measure for $p < 2$ and as a sparsity, or concentration measure for $p > 2$ (see Section 4.2). Again, the optimum is attained for Gaussian functions. It is worth noticing that as opposed to the measures on subsets, these concentration estimates are strongly influenced by the tail of the Gabor transform or the ambiguity function. It is not clear at all that the latter is actually relevant in practical applications.

3.4.2 Time-frequency concentration on compact sets

As a consequence of Lieb's inequalities, one can show (see [17] for a detailed account) the following concentration properties for ambiguity functions and STFTs: let $\Omega \subset \mathbb{R}^2$, measurable, and $\epsilon > 0$ be such that

$$\int_\Omega |\mathcal{V}_g f(b, \nu)|^2 \, db \, d\nu \geq (1 - \epsilon) \|g\|^2 \|f\|^2 \,, \qquad (17)$$

then $\forall p > 2$, $|\Omega| \geq (1 - \epsilon)^{p/(p-2)} (p/2)^{2/(p-2)}$. In particular, for $p = 4$, this yields

$$|\Omega| \geq 2(1 - \epsilon)^2 \,. \qquad (18)$$

Remark 5 It would actually be worth investigating possible corollaries of such estimates, in the sense of Gerchberg-Papoulis. For instance, assume a measurable region Ω of a STFT has been discarded, under which assumptions can one expect to be able to reconstruct stably the region? Also, when T is large, one can probably not expect much stability for the reconstruction, however what would be reasonable regularizations for solving such a time-frequency inpainting problem?

3.4.3 Peakyness of ambiguity function

Concentration properties of the ambiguity function actually play a central role in radar detection theory (see e.g. [31]). However, the key desired property of ambiguity functions, namely *peakyness*, otherwise stated the existence of a sharp peak at the origin, is hardly accounted for by L^p norms, entropies or concentration on compact sets as discussed above.

Ambiguity function peakyness optimization can be formulated in a discrete setting as follows. Suppose one is given a sampling lattice $\Lambda = b_0 \mathbb{Z} \times \nu_0 \mathbb{Z}$ in the time-frequency domain, peakyness of $\mathcal{A}_g$ can be optimized by maximizing (with respect to g) the quantity

$$\mu(g) = \max_{(m,n) \neq (0,0)} \mathcal{A}_g(mb_0, n\nu_0) \,. \qquad (19)$$

Two examples of waveforms with different concentration properties are given in Fig. 2. The gaussian function (left) has well known concentration properties, while the ambiguity function of a high order hermite function (right) is much more *peaky*, even though the function itself is poorly time localized and poorly frequency localized.
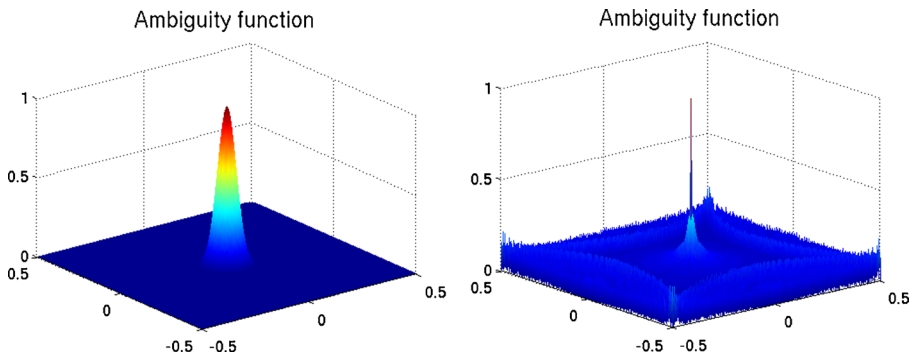
Fig. 2 3D plots of the ambiguity functions of a standard Gaussian (*left*) and a Hermite function of hight order (*right*)

The quantity in Eq. 19 is actually closely connected (as remarked in [29]) to the so-called *coherence*, or *self-coherence* of the Gabor family $\mathcal{D} = \{g_{m,n}, \ m, n \in \mathbb{Z}\}$ generated by time-frequency shifts $g_{mn}(t) = e^{2i\pi n v_0 t} g(t - mb_0)$ of g on the lattice Λ (see [17] for a detailed account), as

$$\mu = \max_{(m',n') \neq (m,n)} |\langle g_{mn}, g_{m'n'} \rangle| \ .$$

Hence, optimizing the peakyness of the ambiguity function is closely connected to minimizing the coherence of the corresponding Gabor family, a property which has been often advocated in the sparse coding literature.

Remark 6 As mentioned earlier, sparsity requirements lead to minimize the joint coherence in the case of separable uncertainty principles, and the self-coherence in the case of non-separable uncertainty principles.

4 Discrete inequalities

4.1 Introduction

The uncertainty principle in the discrete setting has gained increasing interest during the last years due to its connection with sparse analysis and compressive sensing. Sparsity has been shown to be an instrumental concept in various applications, such as signal compression (obviously), signal denoising, blind signal separation,... We first review here the main sparsity/diversity measures that have been used in the signal processing literature, show that they are closely connected and present several versions of the uncertainty principle. Then we present a few examples of their adaptation to phase space concentration problems.

In the discrete finite-dimensional setting, we shall use the Hilbert space $\mathcal{H} = \mathbb{C}^L$ as a model signal space. In terms of signal representations, we consider finite frames $\mathcal{U} = \{u_\lambda \in \mathcal{H}, \ \lambda \in \Lambda\}$ (see Section 2.1 for motivations and definitions) in $\mathcal{H}$, and

denote by $U : x \in \mathcal{H} \to \{\langle x, u_\lambda \rangle, \lambda \in \Lambda\}$ the corresponding *analysis operator*, and by its adjoint U^* the *synthesis operator*.

The time-frequency frames offer a convenient and well established framework for developing ideas and concepts, and most of the approaches described below have been developed using Gabor frames. For the sake of completeness, we give here the basic notations that will be used in the sequel. Given a reference vector $\psi \in \mathcal{H}$ (called the mother waveform, or the window), a corresponding Gabor system associates with ψ a family of time-frequency translates

$$\psi_{mn}(t) = e^{2i\pi m \nu_0 t} \psi(t - nb_0) , \quad m \in \mathbb{Z}_M, \ n \in \mathbb{Z}_N, \ t \in \mathbb{Z}_L ,$$

where $b_0 = a/L$ and $\nu_0 = b/L$ (with a, b integers that divide L) are constants that define a time-frequency lattice Λ. The corresponding transform $\mathcal{V}_\psi$ associates with any $x \in \mathcal{H}$ a function $(m, n) \in \Lambda \to \mathcal{V}_\psi x(m, n) = \langle x, \psi_{mn} \rangle$. When $a = b = 1$, the corresponding transform is called the Short-time Fourier transform (STFT).

The ambiguity function of the window ψ is the function $\mathcal{A}_\psi$ defined as the STFT of the waveform ψ using ψ as mother waveform, in other words $\mathcal{A}_\psi = \mathcal{V}_\psi \psi$.

4.2 Sparsity measures

As mentioned earlier, the variance as a measure of spreading is problematic in the finite setting both with its definition and the inequalities it yields (see nevertheless [26] for an analysis of the connection between continuous and finite variance inequalities). More adapted measures have been proposed in the literature, among which the celebrated ℓ^1-norm used in optimization problems, entropy used by physicists and in information theory and support measures favored for sparsity related problems.

4.2.1 ℓ^p-norms and support measure

Given a finite-dimensional vector $x \in \mathbb{C}^L$, it is customary in signal processing applications to use ℓ^p (quasi-) norms of x to measure the sparsity ($p > 2$) or diversity ($p < 2$) of the vector x:

$$\|x\|_p = \sqrt[p]{\sum_{\ell=1}^{L} |x_\ell|^p} . \tag{20}$$

These quantities (except for $p = 0$) do not fully qualify as sparsity or diversity measures since they depend on the ℓ^2-norm of x. To circumvent this problem, normalized ℓ^p-norms are also considered:

$$\gamma_p(x) = \frac{\|x\|_p}{\|x\|_2} = \|\widetilde{x}\|_p , \ \text{with} \ \widetilde{x} = x/\|x\|_2 . \tag{21}$$

The normalized quantity $|\widetilde{x}|^2$ may be seen as a probability distribution function.

The special case $p = 0$ gives the support measure (number of non-zero coefficients) also denoted ℓ^0. This is not a norm but is obviously a sparsity measure.

4.2.2 Rényi entropies

Entropy is a notion of disorder or spreading for physicists and a well-established notion for estimating the amount of information in information theory. Given $\alpha \in \mathbb{R}^+$ and a vector $x \in \mathbb{C}^L$, the corresponding Rényi entropy [27] $R_\alpha(x)$ is defined as

$$R_\alpha(x) = \frac{2\alpha}{1-\alpha} \ln(\gamma_{2\alpha}(x)) \quad \alpha \neq 1 . \tag{22}$$

Rényi entropies provide diversity measures, i.e. sparsity is obtained by minimizing the entropies. The limit $\alpha \to 1$ is not singular, and yields the Shannon entropy

$$S(x) = -\sum_{\ell=1}^{L} \frac{|x_\ell|^2}{\|x\|_2^2} \ln\left(\frac{|x_\ell|^2}{\|x\|_2^2}\right) . \tag{23}$$

These notions have been proven useful for measuring energy concentration in signal processing, especially in the time-frequency framework [1] and [19].

4.2.3 Relations between sparsity measures

Equation 22 shows that minimizing the ℓ^p-norm with $p < 2$ is equivalent to minimizing the Rényi entropy for $\alpha = p/2$. Note also that for $p = 2\alpha > 2$, $1/(1 - \alpha)$ is negative and minimizing the α-entropy leads to the same results as for *maximizing* the ℓ^p-norm. The limit $\alpha \to 1$ gives the Shannon entropy. Note also that the limit $\alpha = 0$ is not singular and gives the logarithm of the support size. Hence, all these measures a related through Eq. 22 and belong to the same family.

So far, the focus has been put on the Rényi entropies and their limit, the Shannon entropy; Tsallis entropies $T_\alpha(x) = -\left(\gamma_{2\alpha}^{2\alpha}(x) - 1\right)/(\alpha - 1)$, initially introduced in statistical physics, may be seen as some first order approximations of Rényi entropies and can also be used along the same lines. Comparison between these measures could be an interesting issue.

4.3 Sparsity related uncertainties in finite dimensional settings

Discrete uncertainty inequalities have received significant attention in many domains of mathematics, physics and engineering. We focus here on the aspects that have been mostly used in signal processing.

4.3.1 Support uncertainty principles

The core idea is that in finite dimensional settings, two orthonormal bases provide two different representations of the same object, and that the same object cannot be represented sparsely in two "very different bases". In the original work by Donoho and Huo [8], the finite-dimensional Kronecker and Fourier bases were used, and Elad and Bruckstein [10] extended the result to arbitrary orthonormal bases. The ℓ^0 quasi-norm is used to measure diversity.

Theorem 5 *Let $\Phi = \{\varphi_n, n \in \mathbb{Z}_N\}$ and $\Psi = \{\psi_n, n \in \mathbb{Z}_N\}$ denote two orthonormal bases of $\mathbb{C}^N$. For all $x \in \mathbb{C}^N$, denote by $\alpha \in \mathbb{C}^N$ and $\beta \in \mathbb{C}^N$ the coefficients of the expansion of x on Φ and Ψ respectively. Then if $x \neq 0$*

$$\|\alpha\|_0 \cdot \|\beta\|_0 \geq \frac{1}{\mu^2} \quad and \quad \|\alpha\|_0 + \|\beta\|_0 \geq \frac{2}{\mu} , \tag{24}$$

where $\mu = \mu(\Phi, \Psi)$ is the mutual coherence of Φ and Ψ (see Eq. (2)*).*

Remark 7 The *Welch bound* states that the mutual coherence of the union of two orthonormal bases of $\mathbb{C}^N$ cannot be smaller than $1/\sqrt{N}$; the bound is sharp, equality being attained in the case of the Kronecker and Fourier bases.

Remark 8 The result was extended later on by Donoho and Elad [7] to arbitrary frames, using the notion of Kruskal's rank (or spark): the Kruskal rank of a family of vectors $\mathcal{D} = \{\varphi_0, \ldots \varphi_{N-1}\}$ in a finite-dimensional space is the smallest number r_K such that there exists a family of r_K linearly dependent vectors. Assume that $x \in \mathbb{C}^N$, $x \neq 0$ has two different representations in $\mathcal{D}$:

$$\text{if} \quad x = \sum_{n=0}^{N-1} \alpha_n \varphi_n = \sum_{n=0}^{N-1} \beta_n \varphi_n , \quad \text{then} \quad \|\alpha\|_0 + \|\beta\|_0 \geq r_K .$$

Bounds describing the relationship between the Kruskal rank and coherence have also been given in [7].

Let us also mention at this point the discrete versions of the concentration inequality (Eq. 14), obtained in [16]. Given two bases in $\mathbb{C}^N$, let T, F be two subsets of the two index sets $\{0, 1 \cdots, N-1\}$ and assume $|T| \cdot |F| < 1/\mu^2$. Then for all x,

$$\|\alpha\|_{\ell^2(\mathbb{Z}_N \setminus T)} + \|\beta\|_{\ell^2(\mathbb{Z}_N \setminus F)} \geq \left(1 + \frac{1}{1 - \mu\sqrt{|T| \cdot |F|}}\right)^{-1} \|x\|_2 . \tag{25}$$

Remark 9 As expected, all these inequalities imply a strictly positive lower bound and the coherence μ.

In a recent study [28] the support inequalities have been extended from basis representations to frame ones. More precisely, for any vector $x \in \mathcal{H}$, bounds of the following form have been obtained

Theorem 6 *Let $\mathcal{U} = \{\mathcal{U}^{(1)}, \ldots \mathcal{U}^{(n)}\}$ denote a set of K frames in a Hilbert space $\mathcal{H}$. Then for any $x \in \mathcal{H}$*

$$\sum_{k=1}^{n} \|U^{(k)} x\|_0 \geq \frac{n}{\mu_\star} , \tag{26}$$

where $\mu_\star$ is a generalized coherence, defined as follows:

$$\mu_\star = \inf_{\widetilde{\mathcal{U}}} \inf_{1 \leq r \leq 2} \sqrt[n]{\mu_r \left(\widetilde{\mathcal{U}}^{(1)}, \mathcal{U}^{(2)}\right) \ldots \mu_r \left(\widetilde{\mathcal{U}}^{(n-1)}, \mathcal{U}^{(n)}\right) \mu_r \left(\widetilde{\mathcal{U}}^{(n)}, \mathcal{U}^{(1)}\right)} , \tag{27}$$

where the infimum over $\widetilde{\mathcal{U}}$ is taken over the family of all possible dual frames $\widetilde{\mathcal{U}} = \{\widetilde{\mathcal{U}}^{(1)}, \ldots \widetilde{\mathcal{U}}^{(n)}\}$ of the elements of $\mathcal{U}$, and the r-coherences μ_r are defined as

$$\mu_r(\mathcal{U}, \mathcal{V}) = \sup_{v \in \mathcal{V}} \left(\sum_{u \in \mathcal{U}} |\langle u, v \rangle|^{r'} \right)^{r/r'} , \quad \frac{1}{r} + \frac{1}{r'} = 1 . \tag{28}$$

Therefore, the control parameter here is the generalized coherence $\mu_\star$. If the canonical dual frame is chosen, μ_r is often smaller than μ which shows an improvement. This suggests new definitions for the coherence which may improve further the inequality bound.

4.3.2 Entropic uncertainty

In Section 4.2 we introduced the entropy as a measure of concentration and we also stated earlier an entropic version of the uncertainty principle for the continuous case (Hirschman-Beckner). It turns out that the latter can be extended to more general situations than simply time-frequency uncertainty. For example, in a discrete setting, given two orthonormal bases it was proven by Maassen and Uffink [25] and Dembo et al. [5] independently that for any x, the coefficient sequences α, β of the two corresponding representations of x satisfy

$$S(\alpha) + S(\beta) \geq -2 \ln \mu , \tag{29}$$

with μ the mutual coherence of the two bases. In the particular case of Fourier-Kronecker bases, $\mu = 1/\sqrt{N}$, which leads to the similar result given in Prop. 2 for ambiguity function; the picket fences are the minimizers (see next section).

These results were generalized recently in [28], where entropic inequality for frame analysis coefficients were obtained.

Theorem 7 *Let $\mathcal{H}$ be a separable Hilbert space, let $\mathcal{U}$ and $\mathcal{V}$ be two frames of $\mathcal{H}$, with bounds $A_{\mathcal{U}}$, $B_{\mathcal{U}}$ and $A_{\mathcal{V}}$, $B_{\mathcal{V}}$. Let $\tilde{\mathcal{U}}$ and $\tilde{\mathcal{V}}$ denote corresponding dual frames, and set*

$$\rho(\mathcal{U}, \mathcal{V}) = \sqrt{\frac{B_{\mathcal{V}}}{A_{\mathcal{U}}}}, \quad \sigma(\mathcal{U}, \mathcal{V}) = \sqrt{\frac{B_{\mathcal{U}} B_{\mathcal{V}}}{A_{\mathcal{U}} A_{\mathcal{V}}}} \geq 1 , \quad \nu_r(\mathcal{U}, \widetilde{\mathcal{U}}, \mathcal{V}) = \frac{\mu_r(\widetilde{\mathcal{U}}, \mathcal{V})}{\rho(\mathcal{U}, \mathcal{V})^r} . \tag{30}$$

Let $r \in [1, 2)$. For all $\alpha \in [r/2, 1]$, let $\beta = \alpha(r - 2)/(r - 2\alpha) \in [1, \infty]$. For $x \in \mathcal{H}$, denote by a and b the sequences of analysis coefficient of x with respect to $\mathcal{U}$ and $\mathcal{V}$. Then

1. *The Rényi entropies satisfy the following bound:*

$$(2 - r) R_\alpha(a) + r R_\beta(b) \geq -2 \ln(\nu_r(\mathcal{U}, \tilde{\mathcal{U}}, \mathcal{V})) - \frac{2r\beta}{\beta - 1} \ln(\sigma(\mathcal{U}, \mathcal{V})) \tag{31}$$

2. *If $\mathcal{U}$ and $\mathcal{V}$ are tight frames, the bound becomes*

$$(2 - r) R_\alpha(a) + r R_\beta(b) \geq -2 \ln(\nu_r(\mathcal{U}, \widetilde{\mathcal{U}}, \mathcal{V})) . \tag{32}$$

3. *In this case, the following inequalities between Shannon entropies hold true:*

$$S(a) + S(b) \geq -2 \ln \left(\mu_\star(\mathcal{U}, \widetilde{\mathcal{U}}, \mathcal{V}, \tilde{\mathcal{V}}) \right), \tag{33}$$

where $\mu_\star$ is defined in Eq. 27.

The proof is both a refinement and a frame generalization of the proof in [5, 25]. A main result of [28] is the fact that these (significantly more complex) bounds indeed provide stronger estimates than the Maassen-Uffink inequalities, even in the case of orthonormal bases. They are however presumably sub-optimal for non tight frames, as they yield in some specific limit support inequalities that turn out to be weaker than the ones presented above.

4.3.3 Phase space uncertainty and localization

Again, as in the continuous case, uncertainty inequalities defined directly in phase space can be proven. For example, in the joint time-frequency case, finite-dimensional analogues of Lieb's inequalities have been proven in [13].

Proposition 2 *Let $\psi \in \mathbb{C}^N$ be such that $\|\psi\|_2 = 1$. Then, assuming $p < 2$,*

$$\|\mathcal{A}_\psi\|_p \geq N^{\frac{1}{p} - \frac{1}{2}}, \qquad \text{and} \quad S(\mathcal{A}_\psi) \geq \log(N). \tag{34}$$

The inequality is an equality for the family of "picket fence" signals, translated and modulated copies of the following periodic series of Kronecker deltas:

$$\omega(t) = \frac{1}{\sqrt{b}} \sum_{n=1}^{b} \delta(t - an), \quad ab = N.$$

Hence, the result is now completely different from the result obtained in the continuous case: The optimum is not the Gaussian function (which by the way is not well defined in finite-dimensional situations) any more, and is now a completely different object, as examplified in Fig. 3, where a picket fence and a periodized Gaussian window are displayed. This is mainly due to the choice of underlying model signal spaces (generally $L^2(\mathbb{R})$), which impose some decay at infinity.

Remark 10 It is worth noticing that the above diversity measures (norms or entropy of the ambiguity function) are non-convex functionals of the window sequence.
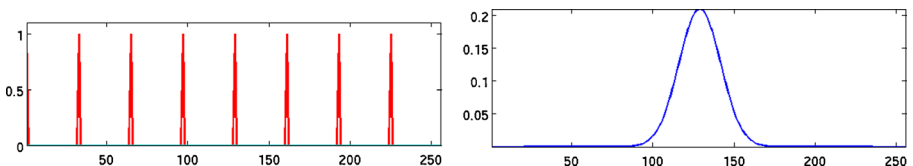

Fig. 3 Picket fence (*left*) vs periodized gaussian (*right*)

For example, if N is a prime number, there are (up to normalization) $2N$ window vectors (picket fences) whose ambiguity function is optimally concentrated (in terms of entropy). When N is not prime, the degeneracy is even higher.

4.4 Two signal processing applications

The uncertainty principle and its consequences have long been considered as constraints and barriers to access precise knowledge and measurement. The innovative idea behind compressive sensing where the uncertainty principle is turned into an advantage for retrieving information promise many exciting developments. In this section we present two prototype applications which involve the uncertainty principle. In the same spirit as compressive sensing, the first one shows how the uncertainty principle can be used for the separation of signals. The second application is a more classical one which provides time-frequency windows with minimum uncertainty, under some additional constraints.

4.4.1 Sparsity-based signal separation problem

The signal separation problem is an extremely ill-defined signal processing problem, which is also important in many engineering problems. In a nutshell, it consists in splitting a signal x into a sum of *components* x_k, or *parts*, of different nature:

$$x = x_1 + x_2 + \cdots + x_n .$$

While this notion of *different nature* often makes sense in applied domains, it is generally extremely difficult to formalize mathematically. Sparsity (see [22] for an introduction in the data separation context) offers a convenient framework for approaching such a notion, according to the following paradigm:

Signals of different nature are sparsely represented in different waveform systems.
    Given a union of several frames (or frames of subspaces) $\mathcal{U}^{(1)}, \mathcal{U}^{(2)}, \ldots \mathcal{U}^{(n)}$ in a reference Hilbert space $\mathcal{H}$, the separation problem can be given various formulations, among which the so-called analysis and synthesis formulations.

– In the *synthesis formulation*, each component x_k will be synthesized using the k-th frame in the form $\sum_j \alpha_j^{(k)} u_j^{(k)}$, and the synthesis coefficients α will be sparsity constrained. The problem is then settled as

$$\min \sum_{k=1}^{n} \|\alpha^{(k)}\|_0 , \quad \text{under constraint} \quad x = \sum_{k=1}^{n} \sum_j \alpha_j^{(k)} u_j^{(k)} .$$

– In *analysis formulations* the splitting of x is sought directly as the solution of

$$\min_{x_1,\ldots x_n \in \mathcal{H}} \sum_{k=1}^{n} \|U^{(k)} x_k\|_0 , \quad \text{under constraint} \quad x = x_1 + x_2 + \ldots x_n ,$$

where $U^{(k)}$ denotes the analysis operator of frame k.

In the case of two frames, it may be proven that if one is given a splitting $x = x_1 + x_2$, obtained via any algorithm, if $\|U^{(1)}x_1\|_0 + \|U^{(2)}x_2\|_0$ is small enough, then this splitting is necessarily optimal. More precisely [28]

Corollary 3 *Let $\mathcal{U}^{(1)}$ and $\mathcal{U}^{(2)}$ denote two frames in $\mathcal{H}$. For any $x \in \mathcal{H}$, let $x = x_1 + x_2$ denote a splitting such that*

$$\left\|U^{(1)}x_1\right\|_0 + \left\|U^{(2)}x_2\right\|_0 < \frac{1}{\mu_\star} .$$

Then this splitting minimizes $\left\|U^{(1)}x_1\right\|_0 + \left\|U^{(2)}x_2\right\|_0$.

Hence, the performances of the analysis-based signal separation problems rely heavily on the value of this generalized coherence function.

The extension to splittings involving more than two parts is more cumbersome. It can be attacked recursively, but this involves combinatorial problems which are likely to be difficult to solve.

4.4.2 Sparsity-based algorithms for window optimization in time-frequency analysis

Proposition 2 shows that the finite dimensional waveforms that optimize standard sparsity measures in the ambiguity domain are not localized, neither in time nor in frequency. This was also confirmed by numerical experiments reported in [13], where numerical schemes for ambiguity function optimization were proposed. This approach has so far been developed mainly with time-frequency representations, but is generic enough to be adapted to various situations.

More precisely, the problem addressed by these algorithms is the following: solve

$$\psi_{\text{opt}} = \underset{\psi : \|\psi\|=1}{\arg\max} \sum_z F\left(|\mathcal{A}_\psi(z)|, z\right) \left|\mathcal{A}_\psi(z)\right|^2, \tag{35}$$

for some *density function* $F : \mathbb{R}^+ \times \Lambda \to \mathbb{R}^+$, chosen so as to enforce some specific localization or sparsity properties. A simple approach, based upon quadratic approximations of the target functional, reduces the problem to iterative diagonalizations of Gabor multipliers.

Two specific situations were considered and analyzed, namely:

– the optimization of the ambiguity function sparsity through the maximization of some ℓ^p norm (with $p > 2$), which naturally leads to choose $F(|\mathcal{A}_\psi(z)|, z) = |\mathcal{A}_\psi(z)|^{p-2}$. The functional to optimize is non-convex, and the outcome of the algorithm depends on the initialization. In agreement with the theory, numerical experiments can converge to picket fence signals (Dirac combs) as limit windows. In addition, for some choices of the initial input window, a Gaussian-like function (the Gaussian is the sparsest window in the continuous case) may also be obtained (local minimum).
– the optimization of the concentration within specific regions, through choices such as $F(|\mathcal{A}_\psi(z)|, z) = F_0(z)$, for some non-negative function F_0 satisfying symmetry constraints, due to the particular properties of the ambiguity function
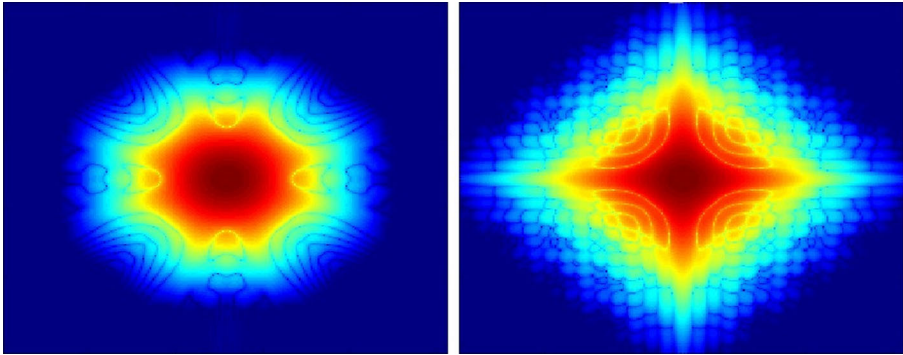
Fig. 4 Logarithm of modulus of optimal ambiguity functions with mask $F(|\mathcal{A}_\psi(z)|, z) = F(0, z)$. *Left*: Optimal function obtained for F the indicator of a disk. *Right*: Optimal function obtained for F the indicator of a diamond

$(\mathcal{A}(0, 0) = 1, \mathcal{A}(z) = \mathcal{A}(-z))$. The algorithm were shown to converge to optimal windows matching the shape of F in the ambiguity plane. That is to say this window is sharply concentrated *and* satisfy the shape constraint provided by F. However, the convergence is not guaranteed for all F and convergence issues should be treated in more details in future works. The algorithm has been shown to converge for simple shapes such as discs, ellipses or rectangles in the ambiguity plane. Numerical illustrations can be found in Fig. 4 (disc shape and rectangular/diamond shape). Since the Ambiguity plane is discrete, the masks are polygons rather that perfect circle and diamonds, and this implies the amazing shape of the ambiguity function, with interferences. For some more complex shapes (such as stars for examples), the algorithm was found not to converge; convergence problems are important issues, currently under study.

Such approaches are actually fairly generic, and there is hope that they can be generalized so as to be able to generate waveforms that are optimal with respect to large classes of criteria.

5 Conclusions

We have reviewed in this paper a number of instances of uncertainty inequalities, in both continuous and discrete situations. Through these particular examples we have focused on specific properties and connections between these different instances. Indeed, from its first statement in quantum mechanics to its newest developments in signal processing, the uncertainty principle has encountered many parallel evolutions and generalizations in different domains. It was not a smooth and straightforward progress, as different situations call for adapted spreading measures, yield different inequalities, bounds and different minimizers (if any), and involve different proof techniques. A main point we have tried to make in this paper is that several classical

approaches, developed in the continuous setting, do not go through in more general situations, such as discrete settings. For example, the very notions of mean and variance do not necessarily make sense in general. In such situations other, more generic, spreading measures such as the (Rényi) entropies and ℓ^p-norms can be used. We attempted in this paper to point out the close connection between these quantities and suggest other candidates for further research.

Signal representations were first understood as the function itself and its Fourier transform. It was then generalized to any projection on orthonormal bases and now any set of frame coefficients. These latter representations play an important role in signal processing and bring some new insight on the uncertainty bounds. The introduction of the mutual coherence measuring how close two representations can be, as well as the phase space coherence that measures the redundancy of a corresponding waveform system, lead to new corresponding bounds. A careful choice for this quantity is needed for obtaining the sharpest bound possible. We showed how this notion of coherence can be extended and generalized, using ℓ^p-norms with $p \neq \infty$.

Concerning the uncertainty optimizers, i.e. waveforms that optimize an uncertainty inequality, they are of very different nature in the discrete and continuous cases. In a few words, in the continuous situations, some underlying choice of functional space implies localization as a consequence of concentration (as measured by the chosen spreading criterion). This is no longer the case in the discrete world where localization and concentration have different meanings.

Therefore, the transition from continuous to discrete spaces is far more complex than simply replacing integrals by sums and a more thorough analysis of the connections between them is clearly needed.

References

1. Baraniuk, R., Flandrin, P., Janssen, A.J., Michel, O.: Measuring time frequency information content using the renyi entropies. IEEE Trans. Inf. Theory **47**(4), 1391–1409 (2001)
2. Beckner, W.: Inequalities in Fourier analysis. Ann. Math. **102**(1), 159–182 (1975)
3. Brascamp, H.J., Lieb, E.H.: Best constants in Young's inequality, its converse, and its generalization to more than three functions. Adv. Math. **20**, 151–173 (1976)
4. Breitenberger, E.: Uncertainty measures and uncertainty relations for angle observables. Found. Phys. **15**(3), 353–364 (1985)
5. Dembo, A., Cover, T.M., Thomas, J.A.: Information theoretic inequalities. IEEE Trans. Inf. Theory **37**, 1501–1518 (1991)
6. Doerfler, M., Torrésani, B.: Representation of operators by sampling in the time-frequency domain. Sampl. Theory Sign Image Process. **10**(1–2), 171–190 (2011)
7. Donoho, D.L., Elad, M.: Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. Proc. Natl. Acad. Sci. **100**, 2197–2202 (2003)
8. Donoho, D.L., Huo, X.: Uncertainty principles and ideal atomic decomposition. IEEE Trans. Inf. Theory **47**, 2845–2862 (2001)
9. Donoho, D.L., Stark, P.B.: Uncertainty principles and signal recovery. SIAM J. Appl. Math. **49**(3), 906–931 (1989)
10. Elad, M., Bruckstein, A.: A generalized uncertainty principle and sparse representation in pairs of bases. IEEE Trans. Inf. Theory **48**, 2558–2567 (2002)
11. Erb, W.: Uncertainty Principles on Riemannian Manifolds. Logos Berlin (2011)
12. Everett, H.I.: The Many-Worlds Interpretation of Quantum Mechanics: The Theory of the Universal Wave Function. Mathematics, Princeton University, Princeton (1957)

13. Feichtinger, H.G., Onchis-Moaca, D., Ricaud, B., Torrésani, B., Wiesmeyr, C.: A method for optimizing the ambiguity function concentration. In: Proceedings of Eusipco 2012 (2012)
14. Flandrin, P.: Inequalities in mellin-fourier analysis. In: Debnath, L. (ed.) Wavelet Transforms and Time-Frequency Signal Analysis, chap. 10, pp. 289–319. Birkhaüser, Cambridge (2001)
15. Folland, G.B., Sitaram, A.: The uncertainty principle: a mathematical survey. J. Fourier Anal. Appl. **3**(3), 207–238 (1997)
16. Ghobber, S., Jaming, P.: On uncertainty principles in the finite dimensional setting. Linear Algebra Appl. **435**, 751–768 (2011)
17. Gröchenig, K.: Foundations of time-frequency analysis. In: Applied and Numerical Harmonic Analysis. Birkhäuser Boston Inc, Boston (2001)
18. Hirschman, I.: A note on entropy. Am. J. Math. **79**, 152–156 (1957)
19. Jaillet, F., Torrésani, B.: Time-frequency jigsaw puzzle: adaptive multiwindow and multilayered gabor expansions. Int. J. Wavelets Multiresolution Inf. Process. **5**(2), 293–315 (2007)
20. Jaming, P.: Nazarov's uncertainty principle in higher dimension. J. Approx. Theory **149**, 611–630 (2007)
21. Judge, D.: On the uncertainty relation for angle variables. Il Nuovo Cimento Ser. **10**(31), 332–340 (1964)
22. Kutyniok, G.: Data separation by sparse representations. In: Eldar, Y. (ed.) Compressed Sensing, Theory and Applications, pp. 485–514. Cambridge University Press, Cambridge (2012)
23. Lieb, E.H.: Integral bounds for radar ambiguity functions and wigner distributions. J. Math. Phys. **31**, 594–599 (1990)
24. Maass, P., Sagiv, C., Sochen, N., Stark, H.-G.: Do uncertainty minimizers attain minimal uncertainty? J. Fourier Anal. Appl. **16**(3), 448–469 (2010)
25. Maassen, H., Uffink, J.: Generalized entropic uncertainty relations. Phys. Rev. Lett. **60**(12), 1103–1106 (1988)
26. Nam, S.: An uncertainty principle for discrete signals. In: Proceedings of SAMPTA' 13. Technical report, LATP, Aix-Marseille Université, Marseille. to appear (2013)
27. Rényi, A.: On measures of information and entropy. In: Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, pp. 547–561 (1960)
28. Ricaud, B., Torrésani, B.: Refined support and entropic uncertainty inequalities. (2012) arXiv:1210.7711
29. Song, X., Zhou, S., Willett, P.: The role of the ambiguity function in compressed sensing. In: IEEE (ed) 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP) (2010)
30. Wehner, S., Winter, A.: Entropic uncertainty relations—a survey. New J. Phys. **12**(2), 025009 (2010)
31. Woodward, P.: Probability and Information Theory with Applications to Radar. Artech House, Norwood (1980)