HAL open science

Two signal processing problems involving coherent states methods

Bruno Torrésani

▶ To cite this version:

Bruno Torrésani. Two signal processing problems involving coherent states methods. Reports on Mathematical Physics, 1999, Coherent states, differential and quantum geometry, 43 (1-2), pp.341-356. 10.1016/S0034-4877(99)80042-1 . hal-01304149

HAL Id: hal-01304149

https://hal.science/hal-01304149v1

Submitted on 19 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TWO SIGNAL PROCESSING PROBLEMS INVOLVING COHERENT STATES METHODS

Bruno Torrésani[1,2]

[1] CPT, CNRS-Luminy, case 907, 13288 Marseille Cedex 09;
[2] LATP, Université de Provence, 39 rue Joliot Curie,
13453 Marseille Cedex 13,
France

Abstract

We discuss the relevance of coherent states based methods in signal processing applications, and illustrate with a couple of practical examples, namely problems of waveform detection in a noisy environment, and problems of spectral analysis for non stationary time series.

1. INTRODUCTION

Coherent states methods have already quite an long history in mathematical physics, and have been used in several contexts (quantization, optics,...). More recently, coherent states and similar tools have been used thoroughly in more applied domains, including in particular signal processing. The main features of coherent states decompositions which are used in that context are of algebraic nature (covariance properties) as well as of analytic nature (coherent states provide simple approximations). All together, it may be said that coherent states have a great simplifying power.

In this paper, we wish to illustrate the potential of such methods with a few examples of practical situations in which it is necessary to consider simultaneously time and frequency variable. We will essentially focus on two problems, namely problems of time-frequency detection of chirp-like signals, and problems of modeling and estimation of non-stationary time series. In both cases, we aim at emphasizing two essential features of coherent states decompositions: the covariance with respect to the underlying group, and the fact that for a large class of problems, coherent states transforms yield sparse representations of functions or signals, which makes easier further tasks (such as detection, parameter estimation, denoising,...).

The use of methods similar to coherent states in signal processing goes back to Gabor, who proposed to use basis functions similar to classical coherent states for providing alternate representations of signal exhibiting simultaneously time and frequency contents. Gabor's approach was later on identified with canonical coherent states by Grossmann, Morlet and Paul [9] and extended to more general situations. The coherent states we shall consider here are based upon the following algebraic construction. Let G be a locally compact Lie group, and let π be a continuous unitary square-integrable representation of G on some Hilbert space $\mathcal{H}$. A famous theorem by Duflo and Moore [7]

states that π is unitarily equivalent to a subrepresentation of the left[1] regular representation of G; in addition, given $\psi \in \mathcal{H}$ such that the function $\langle \pi(g)\psi, \psi \rangle \in L^2(G)$, the Shur's coefficients

$$\phi \in \mathcal{H} \hookrightarrow T_\phi(g) = \langle \phi, \pi(g)\psi \rangle \tag{1.1}$$

serve as intertwiners. In other words, the above mapping $T : \mathcal{H} \to L^2(G)$ is multiple of an isometry, and may therefore be inverted by its adjoint, as follows

$$\phi = \frac{1}{c_\psi} \int T_\phi(g)\, \pi(g)\psi \, d\mu(g) \tag{1.2}$$

where μ is the left Haar measure on G, and the number c_ψ is the formal dimension of the representation π. In the last equation, equality is understood in the weak sense on $\mathcal{H}$. Equation (1.2) is called a *resolution of the identity*, and ψ is called an *admissible vector*.

Grossmann, Morlet and Paul recognized in Gabor's representation a (discretized form of) the transform T for the case of the Weyl-Heisenberg group (in which case $\mathcal{H} = L^2(\mathbb{R})$). They suggested to generalize such constructions to different groups, and to use the so-defined function $T_\phi \in L^2(G)$ as an alternative representation for $\phi \in \mathcal{H}$. The groups of interest in signal analysis are generally those which may be identified with some *time-frequency space*, leading to representations in which time and frequency variables appear simultaneously. The most interesting example is provided by the Weyl-Heisenberg group (Gabor transform), the affine group of the real line, which led to the theory of wavelets, and motion groups in higher dimensions (i.e. semidirect products of translation groups $\mathbb{R}^n$ with subgroups of $GL(n, \mathbb{R})$, see e.g. [2]). Further developments involve the treatment of continuous representations which are not square-integrable on G, but which become square-integrable when restricted to an appropriate quotient $X = G/H$ of G. We shall not go into more details on this point[2], and refer to [1] and references therein for a detailed treatment.

2. CHIRPS AND WAVELET-BASED DETECTION

We start with a classical problem of signal processing, namely a detection problem. Essentially, given a parametric form for a signal, one wants to detect it in a situation where it is embedded into noise, and estimate the parameters. We will see that situation is simpler when it is possible to use a group structure on the parameter space. To illustrate this point, we consider the case of the affine group of the line, and the corresponding wavelet transforms.

We first show how the optimal filtering theory may be reformulated in such a context. We then describe coherent states transform as an alternate approach from which detection algorithms may be designed.

2.1. Abstract Detection Theory

For the sake of simplicity, we limit our discussion to the one-dimensional case, i.e. we address signal processing problems with signals in $L^2(\mathbb{R})$ or related spaces.

Let us start with a simple problem, namely the problem of detecting a given waveform $f_g(t)$, embedded into noise. The noise is generally modelled as a second order

[1]The right regular representation would work as well.

[2]Notice however that the quotient space $X = G/H$ to be considered is often closely related to the *phase space* associated with the considered representation. We find this point very appealing, in connection with the concept of time-frequency representations, by now widely used in signal processing.

stochastic process X_t of zero mean, and covariance $C(t,s) = \mathbb{E}\left\{X_t \overline{X}_s\right\}$ (here, the symbol $\mathbb{E}\{Y\}$ denotes the expectation of the random variable Y). The observed signal assumes the form

$$f(t) = A f_g(t) + X_t \ . \tag{2.1}$$

A is an (unknown) amplitude, g denotes a set of (unknown) parameters, specifying the waveform, and which have to be estimated from the data. We are interested in situations where the waveform $f_g(t)$ is generated in a simple way from a group action, i.e.

$$f_g(t) = \pi(g) f_0(t) \ , \tag{2.2}$$

where π is a continuous unitary representation of some underlying group G on the Hilbert space $\mathcal{H}$, and $f_0 \in \mathcal{H}$.

The classical *matched filter theory* applies to the case $G = \mathbb{R}$, the group of translations of the real line. We now discuss the general case. We consider the detection problem as a least square optimization problem, namely

$$g_{opt} = \min_{g \in G} ||f - f_g||^2 \ . \tag{2.3}$$

It follows directly from the unitarity of π that this problem is equivalent to

$$g_{opt} = \max_{g \in G} \langle f, f_g \rangle \ . \tag{2.4}$$

The purpose of optimal filtering is to construct a linear transform $\mathcal{L} : G \times \mathcal{H} \to \mathbb{C}$, such that $\mathcal{L}(f_g, h)$ attains its maximum for $h = g$, and such that the maximum is as "sharp" as possible. More precisely, let $\varphi \in \mathcal{H}$, and consider the following function

$$\mathcal{O}(h) = \mathcal{L}(f, h) = \langle f, \pi(h)\varphi \rangle \ . \tag{2.5}$$

According to the model (2.1), we have that

$$\mathcal{O}(h) = S(h) + N(h) = \langle \pi(g) f, \pi(h)\varphi \rangle + \langle X, \pi(h)\varphi \rangle \ . \tag{2.6}$$

$S \in L^2(G)$ corresponds to the signal to be detected, and is termed *output signal*, and the random process N on G is the *output noise*. A simple computation shows that $\mathbb{E}\{N(h)\} = 0$ for all $h \in G$, and that $\mathbb{E}\left\{N(h)\overline{N(h')}\right\} = \langle \mathcal{C}\pi(h')\varphi, \pi(h)\varphi \rangle$, where $\mathcal{C}$ is the covariance operator of the process, defined by its kernel $C(t,s) = \mathbb{E}\left\{X_t \overline{X_s}\right\}$. The *output signal to noise ratio (SNR)* is defined as the quotient

$$\rho^2 = \frac{|S(g)|^2}{\mathbb{E}\{|N(g)|^2\}} \ . \tag{2.7}$$

Detection is possible when ρ^2 is large enough; clearly, the higher ρ^2, the easier the detection.

Definition 1. *The* optimal filter *is the function $\varphi \in \mathcal{H}$ which maximizes the SNR ρ^2.*

Then we have the following result:

Proposition 1. *Assume that $\mathcal{C}$ is self adjoint and invertible on $\mathcal{H}$. Then the optimal filter is given by*

$$\varphi_{opt} = K\, \pi(g)^{-1} \mathcal{C}^{-1/2} \pi(g) f_0 \ , \tag{2.8}$$

for some nonzero finite constant K.

The proof is easy. By Cauchy-Schwarz inequality, the unitarity of π and the self-adjointness of $\mathcal{C}$, we have that $|S(g)| = |\langle f_0, \varphi\rangle| \leq ||\mathcal{C}^{-1/2}\pi(g)f_0||\,||\mathcal{C}^{1/2}\pi(g)\varphi||$, and equality is attained if and only if $\mathcal{C}^{-1/2}\pi(g)f_0$ and $\mathcal{C}^{1/2}\pi(g)\varphi$ are proportional. In other words, if and only if φ is given as in (2.8). This proves the proposition. In such a case, the optimal value of the SNR is

$$\rho_{opt}^2 = K\,||\mathcal{C}^{-1/2}\pi(g)f||^2 \ . \tag{2.9}$$

Remark 1. Notice that in general, the optimal function φ depends explicitly upon the unknown parameter g. This implies that such an approach is of poor practical interest, unless $\pi(g)$ and $\mathcal{C}$ commute (which is the case in particular when X is a white noise). Unfortunately, this is not the case for in general, even when the noise is supposed to be stationary, i.e. $\mathcal{C}$ is a convolution operator. This means that for such problems, alternative approaches have to be seeked. Coherent states provide one such approach, which we shall analyze in Section 2.2.

Such an abstract formulation was motivated by practical detection problems, in particular problems of gravitational waves detection at interferometric detectors. In such a case, an idealized model for gravitational waves generated by coalescing massive binary neutron stars describes the corresponding signal $f(t)$ by

$$f_0(t) = \Theta(-t)|t|^\alpha \cos\left(\Phi - |t|^{\beta+1}\right) \ , \tag{2.10}$$

where $\Theta(t)$ is Heaviside's step function, and Φ, α and β are some fixed parameters[3]. The group G is in that case the affine group of the real line, i.e. expected signals are modeled as translated and rescaled copies of the waveform[4] in (2.10). Then, if $g = (b, a) \in G_{aff}$, and we have $f_{b,a}(t) = a^{-r/2} f_0\left(\frac{t-b}{a}\right)$. Then the problem is to estimate from observations the parameters $(b, a) \in G_{aff}$, directly related to physical parameters (masses of the stars, distance,...). Here, r is a fixed real parameter which specifies the representation (see Section 2.2 below).

2.2. Continuous Wavelet Transform and weighted coherent states

The continuous wavelet transform is naturally associated with the coherent states of the affine group $G_{aff} \sim \mathbb{R}^+ \times \mathbb{R}$ of the real line, with group operation: $(b, a)(b', a') = (b + ab', aa')$. Consider the family of generalized Hardy spaces

$$\mathcal{H}_r^\pm = \left\{ f, \hat{f}(\pm\omega) = 0\,\forall\omega \leq 0 \text{ and } ||f||_r^2 := \frac{1}{2\pi}\int |\hat{f}(\omega)|^2 \frac{d\omega}{|\omega|^{r-1}} < \infty \right\} \tag{2.11}$$

(where $\hat{f}(\omega) = \int f(t)e^{-i\omega t}dt$ stands for the Fourier transform of f). $\mathcal{H}_r^\pm$ is equipped with a Hilbert space structure thanks to the scalar product $\langle .,. \rangle_r$ naturally associated with the norm in (2.11). One sometimes needs to introduce the "two-sided" version $\mathcal{H}_r = \mathcal{H}_r^+ \oplus \mathcal{H}_r^-$. Then the representation $\pi = \pi_r$ defined by

$$\pi_r(b, a)f(t) = a^{-r/2} f\left(\frac{t-b}{a}\right), t \in \mathbb{R} \tag{2.12}$$

is a continuous unitary representation of G_{aff} on $\mathcal{H}_r$, $\mathcal{H}_r^+$ and $\mathcal{H}_r^-$, and is square-integrable on both $\mathcal{H}_r^+$ and $\mathcal{H}_r^-$. Specializing the Duflo-Moore theorem to this context, and specializing to the $\mathcal{H}_r^+$ situation for the sake of simplicity, we obtain

[3]This corresponds to a first order post-Newtonian approximation.

[4]Of course, such an oversimplified model is far from accurate, and higher order approximations are used in practice. For the sake of simplicity, we limit our description to this simple case.

Theorem 1. *Let $\psi \in \mathcal{H}_r^+$ be such that $c_\psi = \int_0^\infty |\hat{\psi}(\omega)|^2 \frac{d\omega}{\omega^r} < \infty$. Then $\forall f \in \mathcal{H}_r^+$, we have*

$$f = \frac{1}{c_\psi} \int T_f(b,a) \psi_{(b,a)} \frac{da\, db}{a^2} \ , \tag{2.13}$$

where $\psi_{(b,a)} = \pi_r(b,a)\psi$ and $T_f(b,a) = \langle f, \psi_{(b,a)} \rangle_r$, and the equality holds weakly in $\mathcal{H}_r^+$.

Remark 2. In fact, it may be proved that equality (2.13) holds also strongly in $\mathcal{H}_r^+$, as well as in more general Banach spaces (under some appropriate assumptions on ψ). We refer to [15] for a more detailed discussion.

Remark 3. Equation (2.13) holds also in $\mathcal{H}_r$, provided that ψ is such that the constant $c_\psi = \int_0^\infty |\hat{\psi}(\omega)|^2 \frac{d\omega}{\omega} = \int_0^\infty |\hat{\psi}(-\omega)|^2 \frac{d\omega}{\omega}$ is finite and nonzero.

Let us now examine the wavelet transform of models of the form

$$f(t) = A\, \pi_r(g) f_0(t) + X_t \ , \tag{2.14}$$

where $f \in \mathcal{H}_r^+$, $g \in G_{aff}$ A is a constant and X_t is as before a zero mean weakly stationary second order process, with covariance (convolution) operator $\mathcal{C}$ (defined by its matrix elements $\langle \mathcal{C} f, g \rangle_r = \mathbb{E}\left\{ \langle X, g \rangle_r \overline{\langle X, f \rangle_r} \right\}$). The unitarity of π_r implies that $T_f(h) = A\, T_{f_0}(g^{-1}h) + T_X(h)$, again in the form "signal + noise", and

$$\begin{aligned} \mathbb{E}\left\{ T_f(h) \right\} &= A\, T_{f_0}(g^{-1}h) \ , & (2.15) \\ \mathbb{E}\left\{ |T_f(h)|^2 \right\} &= |A|^2\, |T_{f_0}(g^{-1}h)|^2 + \langle \mathcal{C} \pi_r(h)\psi, \pi_r(h)\psi \rangle_r \ . & (2.16) \end{aligned}$$

Notice that because of the presence of the covariance $\mathcal{C}$, the contribution of the noise is not uniform in G. More precisely, some values of the process $\{T_f(h)\,, h \in G\}$ are more perturbed by the presence of $\{X_t, t \in \mathbb{R}\}$ than others. To overcome such a shortcoming, it is worth introducing the following *weighted coherent states*:

Definition 2. *With the same notations as above, and assuming that $\mathcal{C}$ is self adjoint and invertible on $\mathcal{H}_r^+$, the weighted coherent states transform of $f \in \mathcal{H}_r^+$ is defined by*

$$W_f(h) = \langle \mathcal{C}^{-1/2} f, \pi_r(h)\psi \rangle_r \ . \tag{2.17}$$

Then we have $W_f(h) = A\, W_{\pi_r(g) f_0}(h) + W_X(h)$, and

$$\begin{aligned} \mathbb{E}\left\{ W_f(h) \right\} &= A\, W_{\pi_r(g) f_0}(h) \ , & (2.18) \\ \mathbb{E}\left\{ |W_f(h)|^2 \right\} &= |A|^2\, |W_{\pi_r(g) f_0}(h)|^2 + ||\psi||_r^2 \ . & (2.19) \end{aligned}$$

Now, everything goes as if the noise were a white noise (i.e. $\mathcal{C} = 1$). Weighted coherent states may also be termed *prewhitening coherent states* (or *prewhitening wavelet transform* in [10]).

Remark 4. Notice that by doing so, we lose the covariance of the transform with respect to the group action. But this was already the case in the matched filter approach. To design algorithms for detection, one has to rely on another aspect of coherent states decompositions, which is that they provide efficient (i.e. localized) representations for some signals. This is in particular the case for "chirp-like" signals (which are used as models for gravitational waves generated by coalescing binary stars), which we examine in Section 2.3 below.

2.3. Time-Frequency Analysis of Chirps

Let us now be more specific concerning the waveforms to be detected. For the sake of simplicity, we limit ourselves to $r = 1$ (the other spaces may be obtained by a simple change of normalization).

In the signal processing literature, chirps are generally defined as signals whose amplitude and frequency (whatever it means) vary continuously as functions of time. A simple generic model for such chirps in $\mathcal{H}_r^+$ is

$$f(t) = A(t)e^{i\phi(t)} \ , \tag{2.20}$$

where both $A(t)$ and $\phi'(t)$ are supposed slowly varying (see [16] for more precise conditions). Then, assuming $A \in C^1(\mathbb{R})$ and $\phi \in C^2(\mathbb{R})$, simple arguments show that the wavelet coefficients of such a function satisfy

$$T_f(b,a) = \sqrt{a}A(b)e^{i\phi(b)}\overline{\hat{\psi}(a\phi'(b))} + r(b,a) \ , \tag{2.21}$$

where $|r(b,a)| \leq C\sqrt{a}(\sup|A'| + \sup|\phi''|)$. See [16] for more precise estimates. Therefore, assuming that the wavelet $\psi(t)$ has a "peak" in the frequency domain at a frequency $\omega = \omega_0$, we see that the wavelet transform is "concentrated" near a curve (the so-called "ridge") of the form $a = \omega_0/\phi'(b)$. In addition, the restriction of the transform $T_f(b,a)$ to the ridge is very close (up to a normalization) to the signal itself.

The above approximation gives a simple form for the behavior of the wavelet transform near a ridge. For more specific signals, estimates describing the decay of the wavelet transform away from the ridges may be obtained. A more specific definition of chirps has been given by Meyer (see for example [12]):

Definition 3. *Let $\lambda \in \mathbb{R}$, $\mu > 0$ be two real numbers. A* trigonometric chirp *of type (λ, μ) is a function of the form*

$$h_{(\lambda,\mu)}(t) = |t|^\lambda g\left(t^{-\mu}\right) \ , \tag{2.22}$$

where $g\left(t^{-\mu}\right)$ means $g_+(t^{-\mu})$ if $t > 0$ and $g_-(|t|^{-\mu})$ if $t < 0$, for two T-periodic functions $g_\pm$ of class C^r, such that $\int_0^T g_\pm(t)dt = 0$.

Then it may be shown that if ψ is a wavelet with enough vanishing moments (i.e $\int t^n\psi(t)dt = 0$, $n = 0, \ldots N$ for some N), the corresponding wavelet transform has a tendancy to localize near a curve $a = C|b|^{\mu+1}$ where C is a constant which only depends on ψ. More precisely, one may prove a "local version" of the localization (see [12] for a proof):

Theorem 2 (Meyer)**.** *Let $h_{(\lambda,\mu)}$ be a trigonometric chirp as defined above. Let $\chi \in C_0^\infty(\mathbb{R})$ be a compactly supported function, equal to 1 in a neighborhood of the origin. Let $\psi \in \mathcal{S}(\mathbb{R})$ be a wavelet with all moments vanishing. Let $T^{(\ell)}(b,a) = \langle \chi h_{(\lambda,\mu)}, \pi(b,a)\psi\rangle$ be a local version of the wavelet transform. Then there exists $\delta > 0$ such that*

1. *if $0 < a \leq C|b|^{\mu+1} \leq \delta$, $|T^{(\ell)}(b,a)| \leq C'\sqrt{a}|b|^\lambda \left(\frac{a}{|b|^{\mu+1}}\right)^r$ for some constant C'.*
2. *if $C|b|^{\mu+1} \leq a < \delta$, $\forall N$, $\exists C_N$ such that $|T^{(\ell)}(b,a)| \leq C_N\sqrt{a}|b|^\lambda \left(\frac{a}{|b|^{\mu+1}}\right)^{-N}$.*

Note that such a form is very close to the model (2.10). Such localization properties yield a tremendous simplification of the (b,a) plane. We now show how to take advantage of such a simplification for detection purpose.

2.4. Radon transforms

We have illustrated in the last section the localization properties of the wavelet transform. Let us now come back to the detection problem, and show how such considerations may lead to simplified detection algorithms.

As we have seen, the wavelet transform of chirp-like signals may be sharply localized in the neighborhood of curves, named ridges of the transform. This suggests to base detection methods upon such localization properties, and to set the detection problem as a curve estimation problem. The following approach was suggested in [3]. Let us consider a curve on G, assumed to be a parametric curve of the form $s \in [0,1] \to \varphi(s) \in G$ for the sake of simplicity. Associate with it the following quantity (*line integral*, or *Radon transform* of $|T|^2$):

$$\Psi_f[\varphi] = \int_0^1 |T_f(\varphi(s))|^2 ds - \epsilon \int_0^1 |\varphi(s)^{-1} d\varphi(s)|^2 \, ds \ . \tag{2.23}$$

The idea is to look for maxima of such an expression. The second term enforces smoothness of the ridge, the first one enforces localization near the maxima of the wavelet transform, and ϵ is a real parameter which balances the influence of the two terms. For suitable values of ϵ, such a functional to attains its maximum when the curve $s \to \varphi(s)$ matches the ridge of the wavelet transform. This suggests to use $\Psi_f[\varphi]$ within an optimization procedure to estimate the ridge. This is a difficult optimization problem in infinite dimensional spaces, which require a careful treatment. Such an approach (using Gaussian measures on the space of curves $s \to \varphi(s)$ and stochastic relaxation methods) has been advocated in [3].

We are interested here in a somewhat simpler approach. Indeed, let us assume that the signal we are trying to detect is obtained from a reference signal f_0 by the action of an element $g \in G$; then we have

Proposition 2. *Let $s \to \varphi_0(s)$ minimize the functional $\Psi_{f_0}[\varphi]$, and let $g \in G$. Then $s \to g\varphi_0(s)$ minimizes $\Psi_{\pi_r(g)f_0}[\varphi]$.*

The proposition follows from the covariance of the wavelet transform, and the invariance of the Maurer-Cartan form $\int_0^1 |\varphi(s)^{-1} d\varphi(s)|^2 \, ds$ under the transformation $\varphi \to g\varphi$.

Therefore, we replace the maximization of $\Psi_f[\varphi]$ in (2.23) with respect to generic curves $s \to \varphi(s) \in G$ by the following simpler one (in this parametric situation, the Maurer-Cartan form is not needed any more): solve

$$\max_{h \in G} \Phi_f[h\varphi_0] := \max_{h \in G} \int_0^1 |T_f(h\varphi_0(s))|^2 \, ds \ . \tag{2.24}$$

According to the model in (2.1), we have

$$\mathbb{E}\left\{\Phi_f[h\varphi_0]\right\} = |A|^2 \int_0^1 |T_{f_0}(g^{-1}h\varphi(s))|^2 \, ds + \int_0^1 \langle \mathcal{C}\pi_r(g^{-1}h\varphi(s))\psi, \psi\rangle_r \, ds \ .$$

Such an approach has been proposed and applied to the gravitational waves detection problem in [4, 10, 11]

Remark 5. As before, such an approach has to be modified to take into account the contribution of noise. Indeed, using arguments similar to the above ones, one easily

verifies that the contribution of the noise may be "whitened" by considering weighted coherent states. In such a case, the problem in (2.24) may be replaced with

$$\max_{h\in G} \Phi'_f[h\varphi_0] := \max_{h\in G} \int_0^1 |W_f(h\varphi_0(s))|^2 \, ds \ . \tag{2.25}$$

and it is easily verified that according to (2.1) we have

$$\mathbb{E}\left\{\Phi'_f[h\varphi_0]\right\} = |A|^2 \int_0^1 |W_{\pi_r(g)f_0}(h\varphi_0(s))|^2 \, ds + ||\psi||_r^2 \ .$$

However, in this last case, we do not have $W_{\pi_r(g)f_0}(h\varphi(s)) = W_{f_0}(g^{-1}h\varphi(s))$ as in the standard case, and the deviation to this ideal situation has to be estimated. The numerical simulations in [11] indicate that for the gravitational waves detection problem, the error remains small.

2.5 Back to matched filtering

Let us now come back to our original interest, namely the detection of "chirp-like" waveforms labeled by elements $g \in G$. We have shown in subsection 2.4 how coherent states transforms may be used to construct functions on G which may serve as estimators for the parameter g: to estimate the parameters A and g in (2.1), solve (2.24) or (2.25) instead of (2.4). However, the performances of such algorithms are generally not optimal, contrary to those of matched filters. Then, why should one prefer them to optimal filters ? The answer is robustness. Optimal filters methods turn out to be extremely sensitive to perturbations. Therefore, for the chirp detection problem we have been considering, optimal filters strategies require using a huge number of "templates" φ_{opt} in (2.8), which results in a dramatic computational load. Coherent states approximations have been shown to be more robust.

3. NON-STATIONARY TIME-SERIES

We now address a problem of similar nature (essentially an estimation problem, forgetting about the detection part of the previous Section), in a non-deterministic context. This is also an important issue in signal processing, for many phenomena have to be modeled as random phenomena. It is well known that Fourier methods are well suited for analyzing stationary time series[5], and we shall show how coherent states approaches allow one to go beyond stationary situations.

3.1. Generalities

For the sake of simplicity, we limit our analysis to the case of second order stochastic processes $\{X_t, t \in \mathbb{R}\}$ with zero mean $\mathbb{E}\{X_t\} = 0$. The main object under study is the so-called covariance operator $\mathcal{C}$ of the process, defined by its matrix elements

$$\langle \mathcal{C} f, g\rangle = \mathbb{E}\left\{\langle f, X\rangle \overline{\langle g, X\rangle}\right\} \ , \tag{3.1}$$

for any functions f, g such that such an expression makes sense. For the sake of simplicity, we limit ourselves to the case of processes whose covariance is Hilbert-Schmidt

[5]Stationarity is a weak form of translation invariance, which explains why Fourier methods are so natural.

(which we write $\mathcal{C} \in \mathcal{L}^2$), and we denote by $C \in L^2(\mathbb{R}^2)$ the kernel of $\mathcal{C}$. Formally we have that $C(t,s) = \mathbb{E}\left\{X_t \overline{X_s}\right\}$.

If the process is weakly stationary, $C(t,s) = C(t-s)$ is a function of $t-s$ only by definition and $\mathcal{C}$ is a convolution operator. If not, it is interesting to study processes which are "almost stationary", i.e. which may be considered stationary upon short time intervals. In such a case, the standard practice is to truncate the process into pieces within which it may reasonably considered stationary, and treat it as such (i.e. associate with each "piece" a spectral density for example). An alternative is to look for a "continuously varying" spectral density. This suggests to rewrite the kernel $C(t,s)$ so as to separate the "stationary part", as $C(t,s) = C_0(\frac{t+s}{2}, t-s)$. Assuming that the process is almost stationary amounts to consider that the function $C_0(u,v)$ is slowly varying as a function of the first variable u, and localized as a function of the variable v. Consider now its Fourier transform with respect to the second variable $t-s$:

$$\mathcal{E}_0(b,\omega) = \int C_0(b,\tau)\, e^{-i\omega\tau}\, d\tau = \mathbb{E}\left\{\int X_{b+\tau/2}\overline{X}_{b-\tau/2}\, e^{-i\omega\tau}\, d\tau\right\} . \tag{3.2}$$

$\mathcal{E}_0$ may be thought of as a candidate for *local spectral density* of the process. Notice that the function $\mathcal{E}_0(b,\omega)$ is nothing but a generalization of the celebrated Wigner function to the random situation. Also of interest is the so-called *ambiguity function* (introduced by Woodward in a radar context, see e.g. [3, 5]), defined as

$$\mathcal{A}_0(\tau,\xi) = \frac{1}{2\pi}\int C_0(b,\tau)\, e^{i\xi b}\, db = \mathbb{E}\left\{\int X_{b+\tau/2}\overline{X}_{b-\tau/2}\, e^{i\xi b}\, db\right\} . \tag{3.3}$$

The ambiguity function essentially measures the "spreading" of the process in the time-frequency plane (and is sometimes termed "spreading function" in signal processing). It follows from $\mathcal{C} \in \mathcal{L}^2$ that the Wigner function and the ambiguity function are square-integrable. In addition we have

Lemma 1. *The Wigner function and the ambiguity function are symplectic Fourier transforms of each other.*

In practice, one generally has to work with one (or in rare cases a finite number of) realization(s) of the process. The practical questions are the following:

1. How may one estimate a covariance $\mathcal{C}$, or a local spectral density $\mathcal{E}_0(b,\omega)$, from one (or several) realization(s) of the process X_t ?

2. How may then one mimic the spectral decomposition of the process (find eigenfunctions, or singular functions if $\mathcal{C}$ is not self adjoint)?

The most immediate solution to the first problem is the sample estimate[6], obtained by dropping the expectations in (3.2) and (3.3): for example

$$\tilde{\mathcal{E}}_0(b,\omega) = \int X_{b+\tau/2}\overline{X}_{b-\tau/2}\, e^{-i\omega\tau}\, d\tau .$$

However, the latter does not have good statistical properties (the sample estimate is unbiased, i.e. $\mathbb{E}\left\{\tilde{\mathcal{E}}_0(b,\omega)\right\} = \mathcal{E}_0(b,\omega)$ for all $b,\omega \in \mathbb{R}$), but generally turns out to have a large variance, and is generally not regular even when $\mathcal{E}_0$ is smooth.) In addition, a spectral decomposition obtained from such estimates is often difficult to manipulate numerically (see [3] for a detailed discussion of these points). Therefore, there is room for alternatives.

[6] We use the notation $\tilde{\theta}$ to denote an estimator for the parameter θ.

3.2. Coherent States Approximations

A simple alternative is sometimes provided by time-frequency coherent states. Indeed, if the process is almost stationary as indicated before, simple (formal) arguments show that Gabor functions or wavelets may be considered as almost eigenfunctions of $\mathcal{C}$, therefore yielding approximate spectral decompositions.

Given the process $\{X_t, t \in \mathbb{R}\}$ considered above, its corresponding coherent states transform is a new stochastic process

$$T_X(g) = \langle X, \pi(g)\psi \rangle \ . \tag{3.4}$$

T_X is clearly of zero mean, and its covariance structure is given by the matrix elements of $\mathcal{C}$ between coherent states:

$$C(g, g') = \mathbb{E}\left\{T_X(g)\overline{T_X}(g')\right\} = \langle \mathcal{C}\pi(g')\psi, \pi(g)\psi \rangle \ . \tag{3.5}$$

Let us now introduce the *coherent states spectrum*

$$\mathcal{E}_G(g) = \frac{1}{||\psi||^2}\ \mathbb{E}\left\{|T_X(g)|^2\right\} \ , \tag{3.6}$$

and the corresponding sample estimator

$$\tilde{\mathcal{E}}_G(g) = \frac{1}{||\psi||^2}\ |T_X(g)|^2 \ ,$$

which is easily seen to be unbiased (i.e. $\mathbb{E}\left\{\tilde{\mathcal{E}}_G(g)\right\} = \mathcal{E}_G(g)$ for all $g \in G$.)

Remark 6. When the group G may be identified with a (generalized) "time-frequency" space associated with the representation space $\mathcal{H}$ (for example when G is the Weyl-Heisenberg group or the affine group, or higher-dimensional generalizations), the coherent states spectrum in (3.6) provides an alternative to the Wigner spectrum given in (3.2). However, it should be noticed that while the latter is "intrinsic" to the process, the former depends on an analysis tool, namely the admissible vector $\psi \in \mathcal{H}$. We shall see later on how to deal with such an apparent shortcoming.

The Wigner function and the coherent states spectrum are related by the following

Proposition 3. *Let $\mathcal{C}$ be the covariance operator of the process $\{X_t, t \in \mathbb{R}\}$, and assume $\mathcal{C} \in \mathcal{L}^2$. Then, denoting generically by Π_f the rank one projector onto $f \in \mathcal{H}$, we have*

$$\mathcal{E}_G(g) = \langle \mathcal{C}, \Pi_{\pi(g)\psi} \rangle = \langle \mathcal{C}, ad(\pi(g)) \cdot \Pi_\psi \rangle_{\mathcal{L}^2} \ . \tag{3.7}$$

Here, we denote as usual by ad the adjoint action, i.e. $ad(\pi(g))\Pi = \pi(g)\Pi\pi(g)^{-1}$, and $\langle ., . \rangle_{\mathcal{L}^2}$ is the Hilbert-Schmidt scalar product. The proposition is readily verified by an easy calculation.

Let us now turn to the second aspect of the program, namely the spectral decomposition of $\{X_t, t \in \mathbb{R}\}$. The classical spectral analysis of (possibly non stationary) processes is based upon the spectral analysis of the corresponding covariance operator $\mathcal{C}$. The (possibly generalized) eigenfunctions serve as a basis for expanding the process. The objective is to study the relationship between the considered coherent states and the generalized eigenfunctions[7] of $\mathcal{C}$. To this end, let us set

$$\mathcal{C}\pi(g)\psi = \lambda(g)\pi(g)\psi + r_g \ , \tag{3.8}$$

[7]Let us recall here that as a Hilbert-Schmidt operator, $\mathcal{C}$ may be decomposed as an infinite series of the form $\mathcal{C} = \sum_{k=1}^{\infty} \lambda_k P_k$, where $\lambda_1 \geq \lambda_2 \geq \ldots$ are the singular values of $\mathcal{C}$, and the P_k are rank one projections onto the corresponding eigenfunctions. Such a decomposition may in principle be performed directly once $\mathcal{C}$ is known. However, this problem turns out to become a difficult practical (i.e. numerical) problem. Therefore, it is worth trying to approximate the above singular value decomposition with a decomposition with respect to a suitable coherent states system.

for some number $\lambda(g)$, and where r_g is a remainder, such that $r_g \perp \pi(g)\psi$. This last property implies clearly that $\lambda(g) = \mathcal{E}_G(g)$. Moreover, we have $||r_g||^2 = ||\mathcal{C}\pi(g)\psi||^2 - |\mathcal{E}_G(g)|^2$, and it follows from the square-integrability of π that $\int ||\mathcal{C}\pi(g)\psi||^2 d\mu(g) = c_\psi ||\mathcal{C}||^2_{\mathcal{L}^2}$. Therefore, we obtain

$$\int ||r(g)||^2 d\mu(g) = c_\psi ||\mathcal{C}||^2_{\mathcal{L}^2} - \int |\mathcal{E}_G(g)|^2 \, d\mu(g) \tag{3.9}$$

The practical problem now is to make such a remainder as small as possible. An elegant solution to this problem (which addresses at the same time the problem of dependence with respect to the admissible vector ψ) may be derived via an optimization procedure.

Let us suppose that we consider a family of admissible vectors $\psi \in \mathcal{H}$, normalized so that $||\psi|| = 1$. We will say that the optimal vector is the vector ψ such that $\int ||r(g)||^2 d\mu(g)$ is minimal. By the above argument, we see that this minimization problem is equivalent to

$$\max_{\psi, ||\psi||=1} \frac{1}{c_\psi} \int |\mathcal{E}_G(g)|^2 \, d\mu(g) \ . \tag{3.10}$$

Heuristically, the latter problem amounts to find the admissible vector $\psi \in \mathcal{H}$ which makes the covariance $\mathcal{C}$ "as diagonal as possible" in the coherent states representation. Such a problem may be solved numerically, once a family of admissible vectors $\psi \in \mathcal{H}$ has been chosen. We shall see in Section 3.3 below the case of Gabor coherent states, and how this connects to ambiguity functions.

Remark 7. Let us stress that the complete spectral decomposition problem presents an extra level of difficulty, since one looks for a *discrete* family of coherent states, as close as possible to a basis of $\mathcal{H}$, and such that the corresponding representation of $\mathcal{C}$ is as close as possible to diagonal. We shall not go that far. Let us just notice that this problem has been addressed in the Weyl-Heisenberg case in [14].

3.3. The case of Gabor coherent states

After these abstract considerations, let us come back to practical problems and focus on the case of Gabor functions, i.e. Weyl-Heisenberg coherent states. This case was studied by several authors, see in particular [13]. Let $G_{WH} \sim \mathbb{R}^2 \times S^1$ be the Weyl-Heisenberg group, with group law $(b, \omega, \varphi)(b', \omega', \varphi') = (b + b', \omega + \omega', \varphi + \varphi' + \omega b')$. We consider the canonical representation π of G_{WH} on $\mathcal{H} = L^2(\mathbb{R})$, defined by $\pi(b, \omega, \varphi)f(t) = e^{i(\varphi + \omega(t-b))} f(t-b)$, $f \in L^2(\mathbb{R})$. Let $\psi \in L^2(\mathbb{R})$ be such that $||\psi|| \neq 0$, and set

$$\psi_{(b,\omega)}(t) = \pi(b, \omega, 0)\psi(t) = e^{i\omega(t-b)}\psi(t-b) \ . \tag{3.11}$$

A simple computation shows that ψ is an admissible vector, and $c_\psi = 2\pi ||\psi||^2$. Therefore the Duflo-Moore theorem applies and we consider the corresponding coherent states[8].

Given the process $\{X_t, t \in \mathbb{R}\}$ considered above, its corresponding coherent states transform is a new stochastic process

$$T_X(b, \omega) = \langle X, \psi_{(b,\omega)} \rangle \tag{3.12}$$

[8]In fact, we consider a slight extension of Duflo-Moore's theorem, in which the factor S^1 is taken out. Since this is a compact factor, this does not change anything to square-integrability, and just introduces a "twist" in the intertwining of π and the regular representation. The associated coherent states transform is only covariant modulo a phase.

with zero mean and covariance

$$\mathbb{E}\left\{T_X(b,\omega)\overline{T_X(b',\omega')}\right\} = \langle \mathcal{C}\psi_{(b',\omega')}, \psi_{(b,\omega)}\rangle \ . \tag{3.13}$$

The following gives an alternative to Wigner's function as a candidate for local spectral density of the process $\{X_t, t \in \mathbb{R}\}$.

Definition 4. *With the same notations as before, the* Gabor spectrum *of the process* $\{X_t, t \in \mathbb{R}\}$ *is the function of two variables*

$$\mathcal{E}_{G_{WH}}(b,\omega) = \frac{1}{||\psi||^2}\,\mathbb{E}\left\{|T_X(b,\omega)|^2\right\} \ . \tag{3.14}$$

Then we have the following

Proposition 4. $\mathcal{E}_{G_{WH}} \in L^2(\mathbb{R}^2)$, *and*

$$\mathcal{E}_{G_{WH}}(b,\omega) = \int \mathcal{E}_0(x,k)\overline{\mathcal{E}_\psi}(x-b,\omega-k)dbdk \ , \tag{3.15}$$

where $\mathcal{E}_\psi(b,\omega)$ *is the (deterministic) Wigner function of the function* $\psi \in L^2$*:*

$$\mathcal{E}_\psi(b,\omega) = \int \psi(b+\tau/2)\overline{\psi}(b-\tau/2)\, e^{-i\omega\tau}\, d\tau \ . \tag{3.16}$$

The proof makes use of the ambiguity function of the window

$$\mathcal{A}_\psi(\tau,\xi) = \int \psi(b+\tau/2)\overline{\psi}(b-\tau/2)\, e^{i\xi b}\, db \ . \tag{3.17}$$

It may be proved that $\mathcal{A}_\psi \in L^\infty(\mathbb{R}^2)$ (see below). Then, a simple calculation shows that

$$\mathcal{E}_{G_{WH}}(b,\omega) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \mathcal{A}_0(\tau,\xi)\mathcal{A}_\psi(\tau,\xi)e^{i(\xi b-\omega\tau)}\, d\xi\, d\tau \ ,$$

which proves the result.

The boundedness of $\mathcal{A}_\psi(\tau,\xi)$ (which may also be verified directly) is a particular case of estimates due to E. Lieb: If $\psi \in L^2(\mathbb{R})$, then $\mathcal{A}_\psi \in L^p(\mathbb{R}^2)$ for $1 \le p \le \infty$; in particular, if $2 \le p \le \infty$, $||\mathcal{A}_\psi||_p \le \left(\frac{4\pi}{p}\right)^{1/p} ||\psi||^2$.

In that context, the question of the dependence of $\mathcal{E}_{G_{WH}}(b,\omega)$ with respect to the admissible vector $\psi \in L^2(\mathbb{R})$ may be understood as follows. From Lemma 1, Equation (3.10) may be rewritten as

$$\max_{\psi,||\psi||=1} \frac{1}{c_\psi}\int |\mathcal{E}_G(g)|^2\, d\mu(g) = \max_{\psi,||\psi||=1} \frac{1}{2\pi}\int_{\mathbb{R}^2} |\mathcal{A}_0(\tau,\xi)|^2|\mathcal{A}_\psi(\tau,\xi)|^2\, d\tau d\xi \ , \tag{3.18}$$

which makes sense since $\mathcal{A}_\psi \in L^\infty(\mathbb{R}^2)$. Therefore, the optimal window is the L^2 function whose spreading in the time-frequency plane best matches that of the process.

In practice, one has to limit oneself to finite families of windows to solve the optimization problem. Those families may often be modified by simple transformations of a unique one. Examples adapted to problems of speech coding have been given in [13].

Remark 8. We have described here a simple criterion for selecting an admissible vector. Several alternatives have been proposed (see [13] for example). However, most of them rely on the same kind of ideas, namely to find a vector whose ambiguity function best matches that of the process, in some appropriate sense.

Remark 9. We have investigated here the simpler case of Weyl-Heisenberg coherent states and their relationship to the classical Wigner function. Similar methods may be developed in the case of other groups, in particular the affine group we discussed in Section 2, or more general semidirect products in higher dimensions. It is important to stress that different types of coherent states are adapted to different types of processes. In particular, the Weyl-Heisenberg coherent states we just used are well adapted to "locally stationary processes", which present spectral characteristic which vary slowly with time. On the other hand, affine coherent states are more suited for processes in which dilations play a non trivial role (such as the ordinary or the fractional Brownian motions).

4. CONCLUSIONS

We have described a couple of problems of interest in signal processing and statistics, for which coherent states and related methods provide accurate and efficient solutions. In particular, we have tried to show explicitely the simplifying power of coherent states, which relies on two different kinds of arguments: algebraic arguments (namely, the choice of the covariance group G), and analytic arguments (approximations). We have essentially focused on detection (and parameter estimation), and estimation problems, which are current important problems in signal processing and statistics, with the goal of giving an overview of the techniques employed.

Such methods have enjoyed significant success during the last ten years (see e.g. [3, 6, 8] and references therein for reviews. Besides their adaptation to concrete problems, they are also remarkably simple to implement numerically. However, there is still a lot to be done, especially in the construction of new coherent states families adapted to different situations. We refer to [18] for several examples (for examples, constructions on spheres, or associated to aperiodic tilings...).

Another point which is still to be understood is that of the selection of the decomposition to be used for a given problem. On the practical side, solutions have been proposed, which generally rely on numerical optimization algorithms. But very little has been done on a more theoretical basis.

To conclude, it is fair to say that several of the recent advances in time-frequency signal theory go out of the scope of the coherent states constructions we have used in this paper. In many instances, coherent states have been generalized to get rid of obstructions imposed by algebraic constraints (the most significant examples are presumably the wavelet bases, which have not received any group theoretical interpretation so far, and the Balian-Low phenomenon which prevents from constructing bases of smooth and localized Weyl-Heisenberg coherent states; see e.g [6] for a review). However, it is worth mentioning that in all these cases, the "coherent states philosophy", namely the time-frequency localization, is still present. The deviation from algebraic constructions is therefore minor.

References

[1] S.T. Ali, J.P. Antoine, J.P. Gazeau and U.A. Müller (1995): Coherent States and their Generalizations: an Overview. *Rev. Math. Phys.* **7**, 1013-1104.

[2] D. Bernier and M. Taylor (1996): Wavelets and Square Integrable Representations. *SIAM J. Math. Anal.* **27** pp 594-608

[3] R. Carmona, W.L. Hwang and B. Torrésani (1998): *Practical Time-Frequency Analysis: Wavelet and Gabor transforms, with an implementation in S*, Academic Press, to appear.

[4] E. Chassande-Mottin and P. Flandrin (1997): On the time-frequency detection of chirps, Preprint, ENS-Lyon.

[5] L. Cohen (1995): *Time–frequency Analysis.* Prentice Hall, Englewood Cliffs (NJ).

[6] I. Daubechies (1992): *Ten Lectures on Wavelets.* CBMS-NFS Regional Series in Applied Mathematics **61**.

[7] M. Duflo and C. C. Moore (1976): On the regular representation of a non-unimodular locally compact group. *J. Funct. An.* **21**, 209-243.

[8] P. Flandrin (1993): *Temps-Fréquence.* Traité des Nouvelles Technologies, série Traitement du Signal, Hermes.

[9] A. Grossmann, J. Morlet, T. Paul (1985): Transforms associated with square integrable group representations I. *J. Math. Phys.* **27**, p.2473. Transforms associated with square integrable group representations II. *Ann. Inst. H. Poincaré* **45**, p.293.

[10] J.M. Innocent and B. Torrésani (1996): Wavelet Transforms and Binary Collapse Detection, Banach Center Publications **41**, *Mathematics of Gravitation II*, pp. 179-208.

[11] J.M. Innocent and B. Torrésani (1997): Wavelets and Binary Coalescence Detection, *Appl. and Comp. Harm. Anal.* **4**, pp. 113-116.

[12] S. Jaffard and Y. Meyer (1995): Wavelet Methods for Pointwise Regularity and Local Oscillations of Functions. *Memoirs of the AMS* **123**, n. 587.

[13] W. Kozek (1996): *Spectral Estimation in Non-Stationary Environments.* PhD Thesis, Vienna.

[14] S. Mallat, G. Papanicolaou and Z. Zhang (1998): Adaptive Covariance Estimation of Locally Stationary Processes, *Advances in Statistics* **26**, to appear.

[15] Y. Meyer (1995): *Wavelet and Operators I: Wavelets*; Cambridge University Press (English translation of *Ondelettes et opérateurs*, Herman, Paris (1989)).

[16] Ph. Tchamitchian, B. Torrésani (1991): Ridge and Skeleton extraction from wavelet transform, in *Wavelets and their Applications*, M.B. Ruskai & Al Eds, Jones&Bartlett, Boston.

[17] B. Torrésani (1996): *Analyse Continue par Ondelettes.* Inter Editions/Editions du CNRS (in French). *Continuous Wavelet Analysis*, English translation to appear at SIAM (translated and revised by R.D. Ryan).

[18] B. Torrésani Ed. (1998): *Wavelet and time-frequency analysis*, special issue of the *J. Math. Phys.*, August 1998.