On the Inverse Windowed Fourier Transform

Laura Rebollo-Neira and Juan Fernandez-Rubio

Abstract—The inversion problem concerning the windowed Fourier transform is considered. It is shown that, out of the infinite solutions that the problem admits, the windowed Fourier transform is the “optimal” solution according to a maximum-entropy selection criterion.

Index Terms—Gabor transform, inversion problems, maximum entropy, windowed Fourier transform.

I. INTRODUCTION

The use of a generalized Fourier integral to convey simultaneous time and frequency information was first introduced by Gabor (1946). In [2], he defines a windowed Fourier integral, using a Gaussian window. Later, the window was generalized to any function in $L^2(R)$, the space of square integrable functions. The so generalized Gabor transform is mostly referred to as the windowed Fourier transform (WFT).

Restricting the space of signals to $L^2(R)$, the WFT is a mapping from $L^2(R)$ to $L^2(R^2)$ which is not bijective. As a consequence, lack of uniqueness of the inverse problem must be expected. In this contribution, we focus on a statistical analysis of the inversion problem. First the problem is shown to admit an infinite number of solutions. We then work on the space of possible solutions adopting a statistical description as the essential tool. The possible solutions are considered as a stochastic process distributed according to a (to be determined) probability density. The desired solution is estimated as the mean value of the random process. Among all the probability densities capable of yielding admissible mean-value solutions we single out one, adopting the maximum-entropy principle (MEP).

Finally, we show that, from the maximum-entropy (ME) probability density a mean-value solution is inferred which is identical to the WFT. Thereby the WFT is shown to be an “optimal” solution according to an ME selection criterion. This result also holds as a property within the Frame Theory [9].

II. THE WFT INVERSE PROBLEM

Definition: Let $f(x) \in L^2(R)$ be a given signal and $g(x) \in L^2(R)$ be any fixed function in $L^2(R)$. The WFT of $f(x)$ is a function $F(\omega, t) \in L^2(R^2)$ defined by

$$ F(\omega, t) = \left\{ e^{j\omega x} g(x-t) | f(x) \right\} = \int_{R} e^{-j\omega x} g^*(x-t) f(x) dx $$

(1)

where $g^*(x)$ denotes the complex conjugate of $g(x)$.

The signal can be reconstructed from its WFT through the inversion formula [1], [3], [6]

$$ f(x) = \frac{1}{C_g} \int_{R^2} e^{j\omega x} g(x-t) F(\omega, t) d\omega dt $$

(2)

where

$$ C_g = \| g \|^2 = \int_{R} |g(x)|^2 dx. $$

Although the inversion formula (2) allows the recovery of a signal from its WFT, the inversion is not unique. Let us denote $W$ to the image of the WFT, i.e.,

$$ W = \left\{ F(\omega, t); F(\omega, t) = \int_{R} e^{-j\omega x} g^*(x-t) f(x) dx; \right\} $n $f(x) \in L^2(R)\right\}$$

(3)

$W$ is only a closed subspace, not all of $L^2(R^2)$ (not every function $h(\omega, t) \in L^2(R^2)$ belongs to $W$). The next theorem, whose proof is given in [6, p. 56], provides the necessary and sufficient condition for $h(\omega, t) \in W$.

Theorem 1: A function $h(\omega, t)$ belongs to $W$ if and only if it is square integrable and, in addition, satisfies

$$ h(\omega', t) = \frac{1}{C_g} \int_{R^2} K(\omega', t', \omega, t) h(\omega, t) d\omega dt $$

(4)

where

$$ K(\omega', t', \omega, t) = \left\{ e^{j\omega' x} g(x-t') \right\} \left\{ e^{j\omega x} g(x-t) \right\} = \int_{R} e^{-j\omega x} g^*(x-t') e^{j\omega x} g(x-t) dx. $$

(5)
The function $K(\omega', t', \omega, t)$ is called the reproducing kernel determined by the window $g$ and (4) is called the associated consistency condition.

**Theorem 2:** All functions $h^{\perp}(\omega, t)$ belonging to $\mathcal{W}^{\perp}$ (the orthogonal complement of $\mathcal{W}$) satisfy

$$
\int_{R^2} e^{i\omega \cdot x} g(x-t) h^{\perp}(\omega, t) d\omega dt = 0. \quad (6)
$$

**Proof:** Multiplying the right-hand side of (6) by $e^{-i\omega \cdot x'} g^{*}(x-t')$ and integrating over $x$ we have

$$
\int_{R^2} \left( \int_{R} e^{i\omega \cdot x} g(x-t) e^{-i\omega \cdot x'} g^{*}(x-t') dx \right) h^{\perp}(\omega, t) d\omega dt = 0. \quad (7)
$$

From the definition of $\mathcal{W}$ it follows that

$$
\int_{R} e^{i\omega \cdot x} g^{*}(x-t) e^{-i\omega \cdot x'} g(x-t') dx \in \mathcal{W}. 
$$

Consequently,

$$
\int_{R^2} \left( \int_{R} e^{i\omega \cdot x} g(x-t) e^{-i\omega \cdot x'} g^{*}(x-t') dx \right) h^{\perp}(\omega, t) d\omega dt = 0. \quad (8)
$$

because $h^{\perp}(\omega, t) \in \mathcal{W}^{\perp}$ is orthogonal to every function in $\mathcal{W}.$

Notice that (8) can be recast in the form

$$
\langle e^{i\omega \cdot x} g(x-t') | F(x) \rangle = \int_{R} e^{i\omega \cdot x} g^{*}(x-t') F(x) dx = 0 \quad (9)
$$

where

$$
F(x) = \int_{R^2} e^{i\omega \cdot x} g(x-t) h^{\perp}(\omega, t) d\omega dt 
$$

and since

$$
\text{span} \left\{ e^{i\omega \cdot x} g(x-t') \right\}_{\omega, t, t'} \in R^2
$$

is dense in $L^2(R)$

$$
\langle e^{i\omega \cdot x} g(x-t') | F(x) \rangle = 0, \quad \forall (\omega, t')
$$

implies $F(x) \equiv 0,$ whereby the proof is completed. □

The lack of uniqueness of the inverse WFT is an immediate consequence of Theorem 2. Indeed, besides $F(\omega, t),$ for any $h^{\perp}(\omega, t) \in \mathcal{W}^{\perp}$ the function $h(\omega, t) = F(\omega, t) + h^{\perp}(\omega, t)$ also reconstructs the same signal. The inversion formula (2) corresponds to the particular choice $h^{\perp}(\omega, t) = 0$ and obviously gives rise to a solution of the inverse problem which is “optimal” in a minimum norm (MN) sense. The MN requirement may be a reasonable criterion to be adopted in the case of some applications, but, a priori, certainly not in all of them. In this correspondence we address the problem of deciding on an appropriate estimate for the unknown solution $h(\omega, t)$ by recourse to a postulate originally conceived for the purpose of making decisions in indeterminate situations, namely, the MEP [4], [5]. In the next section we show that the WFT is also an “optimal” solution of the inverse problem according to a ME selection criterion, as it turns out to be the mean of the probability density that maximizes the entropy.

### III. The ME Approach

The problem we address now is that of inverting for $h$ the equation

$$
f(x) = \frac{1}{C_D} \int_{R^2} e^{i\omega \cdot x} g(x-t) h(\omega, t) d\omega dt. \quad (11)
$$

We begin by splitting the above complex equation into real and imaginary parts so that it becomes

$$
f(n)(x) = \frac{1}{C_D} \int_{R^2} (g^{\omega}(x) h^{n}(\omega, t) - g^{\omega}(x) h^{n}(\omega, t)) d\omega dt \quad (12)
$$

$$
f(n)(x) = \frac{1}{C_D} \int_{R^2} (g^{\omega}(x) h^{n}(\omega, t) + g^{\omega}(x) h^{n}(\omega, t)) d\omega dt \quad (13)
$$

where $f^{n}(x),$ $f^{n}(x)$ are the real and imaginary parts of $f(x)$ whereas $h^{n}(\omega, t),$ $h^{n}(\omega, t)$ are the real and imaginary parts of $h(\omega, t)$ and $g^{\omega}(x),$ $g^{\omega}(x)$ are the real and imaginary parts of $e^{i\omega \cdot x} g(x-t),$ respectively.

As discussed in the previous section, there exist several functions $h(\omega, t)$ capable of satisfying (12) and (13). Our aim is that of selecting one of those solutions as “optimal” in an ME sense. In order to achieve such a goal, we regard the possible solutions as a stochastic process and estimate the desired solution as its mean value that we denote

$$
h(\omega, t) = h^{n}(\omega, t) + i h^{n}(\omega, t); (\omega, t) \in R^2.
$$

To deal with the stochastic process in a discrete way, we divide $R^2$ into squares of area $\Delta r = \frac{1}{M^2}$, centered at the points $r_j = (\omega_j, t_j)$ and take lim $M \rightarrow \infty$. With this discretization (12) and (13), which provide the constraints to be satisfied by the desired solution, are evaluated as

$$
f^{n}(x) = \lim_{M \rightarrow \infty} \frac{1}{M^2} \sum_{j=1}^{M^2} \left( g^{\omega}(x) h^{n}(r_j) - g^{\omega}(x) h^{n}(r_j) \right) \quad (14)
$$

$$
f^{n}(x) = \lim_{M \rightarrow \infty} \frac{1}{M^2} \sum_{j=1}^{M^2} \left( g^{\omega}(x) h^{n}(r_j) + g^{\omega}(x) h^{n}(r_j) \right) \quad (15)
$$

At a fixed point $r_j$, both $h^{n}(r_j)$ and $h^{n}(r_j)$ are now random variables. To simplify notation let us denote $h^* = h^{n}(r_1), \ldots, h^{n}(r_M)$ and $h^* = h^{n}(r_1), \ldots, h^{n}(r_M)$. Assuming that these $2M$ random variables are distributed according to a probability density $P(h^*, h^*)$, the mean values $h^{n}(r_j)$, $h^{n}(r_j)$ involved in (14) and (15) are calculated as

$$
\frac{h^{n}(r_j)}{M} = \sum_{j=1}^{M^2} P(h^{n}, h^{n}) h^{n}(r_j) d h^{n} \quad (16)
$$

$$
\frac{h^{n}(r_j)}{M} = \sum_{j=1}^{M^2} P(h^{n}, h^{n}) h^{n}(r_j) d h^{n} \quad (17)
$$

where

$$
dh^{n} = dh^{n}(r_1), \ldots, dh^{n}(r_M)
$$

and

$$
dh^{n} = dh^{n}(r_1), \ldots, dh^{n}(r_M).
$$

Since $P(h^*, h^*)$ is a probability density we must require it satisfies the constraint

$$
\int_{R^2} P(h^*, h^*) dh^* dh^* = 1. \quad (18)
$$

In addition, we should set a constraint to ensure $h(\omega, t) \in L^2(R^2)$. This is guaranteed under the requirement that $\|h\|^2$ be finite, which
also ensures that the variance of the probability density is finite. Consequently, we will set the additional constraint

\[ \|h\|^2 = \lim_{M \to \infty} \frac{1}{M} \sum_{j=1}^{M} \int_{R^{2M}} P(h^w, h^v)(h^n(r_j)^2 + h^v(r_j)^2) \, dh^w \, dh^v = C \]

(19)

where \( C \) is an unknown constant.

Constraints (14), (15), (18), and (19), to be satisfied by the probability density we are looking for, are not enough to determine it in a unique way. Among all the \( P(h^w, h^v) \) capable of fulfilling these constraints, we shall select one adopting the MEP. This criterion yields the probability density that, being consistent with the available data, is maximally noncommital with respect to the lack of information (entropy) \[4\], \[5\].

The entropy, or uncertainty, associated with the probability density is given by the generalization of Shannon’s measure \[8\] to continuous-type random variables \[7\], i.e.,

\[ H(h^w, h^v) = -\int_{R^{2M}} P(h^w, h^v) \ln P(h^w, h^v) \, dh^w \, dh^v. \]

(20)

Since we should take \( \lim M \to \infty \) to represent our stochastic process, the appropriate measure to be used is the entropy rate \( \hat{H} \), or entropy per degree of freedom, defined as \[7\]

\[ \hat{H} = \lim_{M \to \infty} \frac{1}{2M} H(h^w, h^v). \]

(21)

We then look for the probability density that maximizes \( \hat{H} \) with constraints (14), (15), (18), and (19). In order to introduce the constraints (14) and (15) into the variational process, we divide the axis \( R \) into intervals of length \( \Delta x = \frac{1}{N} \) centered at the points \( x_i \), and take \( \lim_{N \to \infty} \), at the end. Assuming that \( f^w(x) \) and \( f^v(x) \) are continuous functions we incorporate each constraint (14), evaluated at \( x = x_j \), through a Lagrange multiplier that we write \( \lambda_j^w, \Delta x \) and each constraint (15) through a Lagrange multiplier \( \beta \), respectively. Thus the functional, \( S \), to be maximized is cast

\[ S = -\frac{1}{2M} \int_{R^{2M}} P(h^w, h^v) \]

\[ \left( \ln P(h^w, h^v) + 2\beta \sum_{j=1}^{M} (h^n(r_j)^2 + h^v(r_j)^2) \right) dh^w \, dh^v \]

\[ -\lambda_0 \int_{R^{2M}} P(h^w, h^v) \, dh^w \, dh^v \]

\[ -\frac{1}{N} \sum_{i=1}^{N} \lambda_i^w \frac{1}{MC_a} \sum_{j=1}^{M} \left( g_{ij}^w(x_i) h^n(r_j) - g_{ij}^v(x_i) h^v(r_j) \right) \]

\[ -\frac{1}{N} \sum_{i=1}^{N} \lambda_i^v \frac{1}{MC_a} \sum_{j=1}^{M} \left( g_{ij}^w(x_i) h^n(r_j) + g_{ij}^v(x_i) h^v(r_j) \right) \]

(22)

\( h^n(r_j) \) and \( h^v(r_j) \) are calculated as in (16) and (17).

From the condition \( \frac{\partial S}{\partial \lambda_0} = 0 \) we obtain

\[ P(h^w, h^v) = \exp\left(2M\lambda_0 + 1\right) \exp\left(-\sum_{j=1}^{M} \left(h^n(r_j)\gamma_1(r_j) + h^v(r_j)\gamma_2(r_j)\right) \right) \]

\[ + \beta h^n(r_j)^2 + \beta h^v(r_j)^2 \]

(23)

where

\[ \gamma_1(r_j) = \frac{1}{NC_a} \sum_{i=1}^{N} \left( \lambda_i^w g_{ij}^w(x_i) + \lambda_i^v g_{ij}^v(x_i) \right) \]

(24)

and

\[ \gamma_2(r_j) = \frac{1}{NC_a} \sum_{i=1}^{N} \left( \lambda_i^w g_{ij}^w(x_i) - \lambda_i^v g_{ij}^v(x_i) \right) \]

(25)

Since the entropy (20) is a convex functional \[7\] it takes on its absolute maximum at \( P(h^w, h^v) \) given in (23).

The normalization constraint (18) entails

\[ \exp(2M\lambda_0 + 1) = \int_{R^{2M}} \exp\left(-2\sum_{j=1}^{M} (h^n(r_j)\gamma_1(r_j) + h^v(r_j)\gamma_2(r_j)) \right. \]

\[ + \beta h^n(r_j)^2 + \beta h^v(r_j)^2 \left.) \right) \, dh^w \, dh^v = \left( \frac{\pi}{2\beta} \right)^{M} \prod_{j=1}^{M} \exp\left(\frac{\gamma_1(r_j)^2}{2\beta} \right) \exp\left(\frac{-\gamma_2(r_j)^2}{2\beta} \right) \]

(26)

The remaining Lagrange multipliers, \( \lambda_i^w, \lambda_i^v; i = 1, \ldots, N \), and \( \beta \), should be obtained by using (23) in (14), (15), and (19) and solving the equations. However, as we shall see below, the functional form of \( P(h^w, h^v) \), given in (23), already provides the information that is needed to determine the mean value function \( h(r_j) \) that such probability density will predict. Indeed, by replacing (23) in (16) and (17) and performing the integrals we have

\[ h^n(r_j) = -\frac{\gamma_1(r_j)}{2\beta} = -\frac{1}{2\beta NC_a} \sum_{i=1}^{N} \left( \lambda_i^w g_{ij}^w(x_i) + \lambda_i^v g_{ij}^v(x_i) \right) \]

(27)

\[ h^v(r_j) = -\frac{\gamma_2(r_j)}{2\beta} = -\frac{1}{2\beta NC_a} \sum_{i=1}^{N} \left( \lambda_i^w g_{ij}^w(x_i) - \lambda_i^v g_{ij}^v(x_i) \right) \]

(28)

Taking now \( \lim N \to \infty \), the above equations yield

\[ h^n(r_j) = -\frac{1}{2\beta NC_a} \int_{R} \left( \lambda_i^w g_{ij}^w(x) + \lambda_i^v g_{ij}^v(x) \right) dx \]

(29)

\[ h^v(r_j) = -\frac{1}{2\beta NC_a} \int_{R} \left( \lambda_i^w g_{ij}^w(x) - \lambda_i^v g_{ij}^v(x) \right) dx \]

(30)

or

\[ h(r_j) = \frac{h(\omega, t_j)}{\sqrt{2\pi} \lambda_i^v} = h^n(\omega, t_j) + i h^v(\omega, t_j) \]

\[ = \int_{R} e^{i\omega t_j} g(x - t_j) w(x) \, dx \]

(31)

with

\[ w(x) = \frac{-1}{2\beta NC_a} (\lambda_i^w + i \lambda_i^v). \]

From (31) and definition (3) we gather that

\[ \frac{h(\omega, t_j)}{\sqrt{2\pi} \lambda_i^v} \in \mathbb{W}, \quad (\omega, t_j) \in \mathbb{R}^2 \]

(32)

So that, by Theorem 1, we are in a position to reveal \( h(\omega, t) \). In fact, by using \( h(\omega, t) \) in (11) and performing the inner product of both sides with \( e^{i\omega t_j} g(x - t_j) \) we have

\[ \langle e^{i\omega t_j} g(x - t_j) \mid f(x) \rangle \]

\[ = F(\omega, t_j) \]

\[ = \frac{1}{C_a} \int_{R^2} \langle e^{i\omega t_j} g(x - t_j) \mid e^{i\omega x} g(x - t) \rangle h(\omega, t) \, d\omega \, dt \]

and, since \( h(\omega, t) \in \mathbb{W} \), by Theorem 1 the consistency condition (4) is verified. Hence, from (32) and Theorem 1 we conclude that
\( \tilde{h}(\omega', t') = F(\omega', t') \), which states the WFT as an optimal solution of the inverse problem according to an ME selection criterion. Since such a solution is also optimal in an MN sense, we are led to conclude that the MN requirement works by averaging functions in a maximally noncommittal way. In other words, we give here a new argument supporting the use of the MN solution for the inverse windowed Fourier transform problem, as it has been shown to be the “least biased” assignment one can make on the basis of the available information.

REFERENCES


