USING A NEW UNCERTAINTY MEASURE TO DETERMINE OPTIMAL BASES FOR SIGNAL REPRESENTATIONS

Tomasz Przebinda[1], Victor DeBrunner[2], Murad Ozaydin[1]

The University of Oklahoma
[1]Department of Mathematics
202 West Boyd Street, Room 219, Norman, Oklahoma 73019-0631
[2]School of Electrical and Computer Engineering

ABSTRACT

We use a new uncertainty measure, H_p, that predicts the compactness of digital signal representations to determine a good (non-orthogonal) set of basis vectors. The measure uses the entropy of the signal and its Fourier transform in a manner that is similar to the use of the signal and its Fourier transform in the Heisenberg uncertainty principle. The measure explains why the level of discretization of continuous basis signals can be very important to the compactness of representation. Our use of the measure indicates that a mixture of (non-orthogonal) sinusoidal and impulsive or "blocky" basis functions may be best for compactly representing signals.

1. INTRODUCTION

The use of wavelets by the signal processing community over the past decade has been astounding. Wavelet-representations of digital non-stationary signals are used in many compression and enhancement schemes [1]. To date, most research interest has centered on orthogonal or biorthogonal perfect reconstruction filter banks [2]. Recently though, some work has been presented that considers (non-orthogonal) multi-transform representations [3]-[9]. These approaches resulted in the development of greedy "recursive residual projection" algorithms [3]-[5] and multivariate Gauss-Newton searches [6]-[9]. Another approach resulted in the greedy "matching pursuits" algorithm of [10], where discretized Gaussian pulses form the basis. One unanswered question in these papers is the so-called "bast-basis" problem. Ad hoc methods or arbitrary combinations of oft-used basis sets are used in [3]-[9]. A more sophisticated approach that is computationally intensive and introduces latency in coding is provided in [11]. An interesting approach to modeling speech and audio is described in [12]. Our results provided in this paper appear to corroborate the results from that paper. We also provide a different viewpoint for understanding how that method might work.

Specifically, in this paper, we show how the discretization of the basis signals changes our conception of the Heisenberg uncertainty principle. Consequently, we introduce a new measure that we conjecture is more applicable to the discrete signal representation problem. We see that discretized Gaussian pulses are not always optimal. We find that the measure provides a rationale that explains why the multi-transform methods work well.

This paper is organized as follows. First we review the basics of the time/frequency (phase) plane for continuous-time signals. We next extend the discussion to the sample-frequency (phase) plane. We see that the Heisenberg inequality used in the time/frquency (phase) plane is inadequate for the discretized version, and we provide an alternative measure using the notion of entropy. We provide a conjecture that we have ample evidence for, including both computational and mathematical (proof in [13]). From this evidence, we find that we have two research directions possible for choosing a "best basis." These are: multi-transform and non-orthogonal basis sets and optimal orthogonal basis functions. We explore only the former in this paper. We discuss the multi-transform approach to select non-orthogonal basis functions in section 4. The latter possibility is currently under study by the authors. Finally, we conclude.

2. TIME-FREQUENCY PLANE FOR CONTINUOUS-TIME SIGNALS

Recall that $L^2(R)$ is the Hilbert space of functions $u: R \to C$ with norm

$$\|u\|^2 = \int_R |u(t)|^2 dt < \infty$$

The Fourier transform $F: L^2(R) \ni u \to F = \hat{u} \in L^2(R)$ is an invertible isometry defined

$$\hat{u}(f) = \int_R u(t) e^{-j2\pi ft} dt, \left(u \in L^2(R)\right)$$

with inverse

$$u(t) = \int_R \hat{u}(f) e^{j2\pi ft} df$$

The coordinates (t, f) provide a time-frequency description of the continuous-time signal $u(t)$, and the vector space of all (t, f) coincides with R^2 and is called the time-frequency plane.

The accuracy of the time-frequency representation is limited by the Heisenberg inequality [14], which relates the uncertainties of time and frequency. Moreover, the (Heisenberg) product of uncertainties is invariant under translation, dilation, and modulation. We believe (and others do as well, e.g. Coifman, Wickerhauser via oral communication) that humans recognize signals as objects in the time-frequency plane. Consequently, it seems reasonable to assume that representations that are highly concentrated in the time-frequency plane are desirable. Because equality (in the Heisenberg inequality) holds for Gaussian shapes, Mallat and Zhang proposed the use of a dictionary of non-orthogonal Gaussian pulses and a search algorithm that they termed "matching pursuits" [10]. However, as we all can attest to, common practice presents us with digitized signals: digital audio, still images, or video. One might think that as long as the number of data is large that the discretization would not impact these results. We have determined that this intuitive feeling is incorrect. Also, in many real instances, the number of data is small, especially in image processing applications. We concentrate on this understanding next.

3. THE SAMPLE-FREQUENCY PLANE FOR DIGITAL SIGNALS

Fix a finite set of non-negative integers $\{0,1,2,\ldots,N-1\}$. Let H_N be the Hilbert space of sequences $u:\{0,1,2,\ldots,N-1\}\to C$ with the norm

$$\|u\|^2 = \sum_{n=0}^{N-1} |u(n)|^2$$

With the twiddle factor $W_N = e^{-j\frac{2\pi}{N}}$, the discrete Fourier transform (DFT) is defined

$$Fu(n) \equiv \hat{u}(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} u(n) W_N^{nk},\ k = 0,1,2,\ldots,N-1$$

This isometry of H_N has inverse

$$u(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{u}(k) W_N^{-nk},\ n = 0,1,2,\ldots,N-1$$

Here, we need only the translation and modulation operators because dilations in the discrete domain can be defined as compositions of translations. Furthermore, there is no geometric analog of dilations because we cannot "zoom in and out" due to the finite number of data. We should note that no results in this paper are affected by dilations, we just choose not to highlight them because of the lack of physical interpretation.

We now recall some simple facts. If we view the digital signal $u(n)$ as a column vector, then we may write the DFT as the matrix multiplication of the column by

$$F = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W_N & W_N^2 & \cdots & W_N^{N-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & W_N^{(N-1)} & W_N^{2(N-1)} & \cdots & W_N^{(N-1)^2} \end{bmatrix}$$

Hence $F^4 = I_N$ (the N-dimensional identity matrix).

We are unaware of a straightforward adaptation of the Heisenberg inequality for signals in H_N. One problem is that the "position operator" is not well defined. One way to circumvent this problem is to use the information theory notion of entropy

$$y(x) = \begin{cases} -x\ln(x) & \text{if } 0 < x < 1 \\ 0 & \text{if } x = 0 \text{ or } x = 1 \end{cases} \qquad (1)$$

Note that the entropy function is undefined outside the interval $[0,1]$. We use the shorthand $-x\ln(x)$, keeping the definition of (1) in mind. Thus, for $\|u\| = 1,\ u \in H_N$, we have

$$H(u) \equiv -\sum_{n=0}^{N-1} |u(n)|^2 \ln\left(|u(n)|^2\right) \qquad (2)$$

This measures the (lack of) concentration of u in the domain n. The range of the measure is from 0 when the signal is the unit sample sequence to $\ln(N)$ when the signal is a constant value [14]. Entropy can be viewed as a measure of uncertainty then because when all of the signal energy is localized in time, the position is certain and the entropy is zero, and if the signal energy is spread evenly over time then the entropy is maximal. Note that the entropy in (2) is invariant under any permutation of indices.

Since $H(Fu)$ measures the uncertainty of u in frequency, it may seem reasonable to consider the straightforward and obvious analog to the Heisenberg inequality

$$H(u)H(Fu) \geq 0,\ \|u\| = 1 \text{ and } u \in H_N \qquad (3)$$

However, it can be easily shown from our previous discussion that the sequences that minimize this naïve analog of the Heisenberg inequality are functions that have all of their energy concentrated at either one time (e.g. the unit sample sequence) or one frequency (e.g. a sinusoidal function). In these cases, the product in (3) achieves the equality, and so this measure does not usefully represent concentration in both time and frequency. A better procedure (with respect to measuring concentration in both the time and frequency domains) is to combine the entropies by adding them. Since the human perception of a signal may depend differently on its concentration in time and frequency, we introduce the following weighted sum as a measure of concentration in both the time and frequency domains:

$$\begin{aligned} H_p(u) &= pH(u) + (1-p)H(Fu), \\ & u \in H_N, \|u\| = 1, 0 \leq p \leq 1 \end{aligned} \qquad (4)$$

We have used the weighting parameter p to trade-off concentration in time and in frequency. For any allowable p, the function H_p in (4) is invariant under translations and modulations. All our experiments with the signals used to generate Figure 1 and Figure 2, as well as the discrete wavelet transform, several different lapped transforms, and many randomly generated signals indicate the veracity of the following:

Conjecture: For any $u \in H_N$, with $\|u\| = 1$, $H_{\frac{1}{2}}(u) \geq \frac{1}{2}\ln(N)$.

Moreover, the only sequences $u \in H_N$, with $\|u\| = 1$, for which $H_{\frac{1}{2}}(u)$ is minimal are obtained via any composition of translation and modulation operators, the DFT, or multiplication by a complex number of absolute value 1 from the sequence $u = \{1 \quad 0 \quad \cdots \quad 0\}$.

We have numerous computer simulations that support the conjecture, as well as the fact that the conjecture can be proven true for all normalized primitive characters in number theory (proof in [13]).

Where the minimum of H_p exists as u varies will be reached on different sequences at different values of p (e.g. see Figure 1 or Figure 2). In particular, we have noticed that discretized Gaussian pulses with large standard deviation relative to the number of samples N do not yield functions well localized in the sample-frequency plane. Considering that we don't know the proper value of p for any given signal, we see two choices in selecting a set of basis functions:

1. Finding basis sets for which (4) is minimal for all p. This choice leads to non-orthogonal basis, in particular multi-transform selections. This choice is examined in the next section.
2. Finding sequences for which (4) is a function of p with zero derivative everywhere (we call this a "horizontal function of p). This could lead to finding a "best" basis that is either orthogonal or non-orthogonal (as in [10]). This choice is currently under study.

It is important to realize that the entropy measure does not generalize to define a good measure of concentration for a continuous signal $u(t), t \in R$, of norm 1. The point is that $|u(t)|^2$ may well take on values larger than 1 and thus fall outside the defined range given in (1) responsible for the good properties of discrete entropy. Shannon [15] has introduced a notion of entropy for continuous random variables replacing the sum by the integral

$$H(u) = -\int_R |u(t)|^2 \ln\left(|u(t)|^2\right) dt$$

The main drawback of this measure of concentration is that $|u(t)|^2$ may well take values larger than 1 and so give negative entropies, which is contrary to our normal understanding of the measure of concentration. Furthermore, when the energy is concentrated at a finite number of instants in time, the above measure approaches negative infinity. Different (non-standard) measures of discrete entropy have been unified to a corresponding continuous-time entropy measure [16], but these measures have not proven to be useful.

4. NON-ORTHOGONAL BASIS

As we have stated earlier, two potential "good" non-orthogonal solutions exist: the general form and the multi-transform solution. We first consider the general form (with Gaussian waveforms), and then from the explanation and the example, approach the multi-transform solution and see why it might be the "best" non-orthogonal solution. First, the Gaussians.

Consider the Gaussian waveform

$$x(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{t-\mu}{\sqrt{2\sigma^2}}\right)^2}$$

and its continuous-time Fourier transform

$$X(f) = \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{t-\mu}{\sqrt{2\sigma^2}}\right)^2} e^{-j2\pi ft} dt$$
$$= \sqrt{2\pi\sigma^2} e^{-j\mu 2\pi f} e^{-j\left(2\pi\sigma^2\right) f^2}$$

While both the waveform and its Fourier transform are Gaussian in shape, the precise shape is not invariant under the transform except under certain conditions on the mean and variance. In discrete time, this change in shape through the transform depends on the number of samples present in the discretized waveform. For example, if we consider the zero mean Gaussian pulse sampled uniformly with N samples, then we must have the variance σ^2 equal to 2 if we have 32 samples in the discrete waveform for the entropies in the sample and frequency domains to be equal. Further results show that the dependence is inversely related. For instance, when $\sigma^2 = 4$ we must have 16 samples for the two entropies to be equal and when $\sigma^2 = 1$ we must have 64 samples for the two entropies to be equal. This change in shape directly affects the computation of our uncertainty measure.

With the number of discretized samples at N=64 and with the continuous Gaussian pulse having unit variance, Figure 1 shows the H_p vs. p for the Euclidian (Identify) vector, the DCT basis, the Gaussian pulse, and the 0^{th}-order Haar function. The horizontal asymmetric dashed line indicates the conjectured minimum for $H_{1/2}$. Note that the Identity and DCT bases achieve this minimum for $p = 1/2$. In this case the Gaussian pulse localizes well across the entire range of p which we expect from our above discussion. Note, however, that the conjectured minimum is **not** reached. The conjectured (and reached) minimum for this value of N is below the horizontal line connecting the Gaussian pulses. Since H_p is linear in p, the "good" basis function family would have a horizontal function. So the Guassian pulses are "good." However, they do not achieve the known minimum (the dashed line in Figure 1), and so it might be possible to improve on them. This problem is not trivial.

Furthermore, note that when N=256, the Gaussian pulse does not localize uniformly across the range of p, as can be seen in Figure 2. This result should be expected because the oversampling (relative to the variance of the continuous Gaussian pulse) leads to unequal shapes in the sample and frequency domains.

We also note from Figure 1 that the optimal basis set for any p must be either the Euclidian vector (unit sample sequence) when $p \le 1/2$ or the DCT basis set (sinusoidal sequence) if $p \ge 1/2$. This fact is also determined experimentally by minimizing the function H_p for the various p. Thus, the "best" set of basis functions would consist of sinusoidal and impulsive (or blocky) basis functions. This is best seen by noting that in this case, the localization would follow the triangle formed by the DCT (+) functions in the figures for signal portions where the "actual" $p \le 1/2$ and the Identity (-) functions in the figures for signal portions where the "actual $p \ge 1/2$. These results appear to strongly confirm some results given in [12], where a multi-transform technique for speech and audio modeling was presented. The modeling technique used sinusoids and transients (identity transforms in our vernacular). The results may also explain some results given in [3],[4]. In those papers, the authors have noted that the best performance occurs when the DCT basis is used in combination with the Slant basis functions. The results may also explain some of the performance seen in the paper [11].

This paper developed a technique for choosing the transforms in a multi-transform system.

5. CONCLUSIONS

We have shown that a straightforward application of our knowledge of the time-frequency plane (in particular, the Heisenberg inequality) to the discretized sample-frequency plane leads to an incomplete understanding and some erroneous conclusions. Because of an ill-posed question, we are forced to consider the use of entropy in an analog to the Heisenberg product, which we have introduced as the function H_p. This measure accounts for the possibility of different resolutions in the sample-frequency plane. For $p \neq \frac{1}{2}$ the minimizing signals are not discretized Gaussian pulses. We conjecture that signals well located in the sample-frequency plane are signals where H_p vs. p is horizontal (yielding the optimal orthogonal basis). However, if orthogonality is not constrained, then the multi-transform methods employed in [3]-[9] appear to be very good candidates for optimality because of their mixture of sinusoidal and impulsive (blocky) sequences. Furthermore, for varying sampling frequencies, discrete Gaussian pulses obtained from continuous ones are not optimal according to our conjecture.

6. REFERENCES

[1] P. P. Vaidyanathan, "Filter banks with maximum coding gain and energy compaction," *Proc. 29th Annual Asilomar Conference on Signals, Systems, and Computers*, vol. I, pp. 36-40, Pacific Grove, CA, Nov. 1995.

[2] P. P. Vaidyanathan, "On the study of lossless time-varying filter banks," *Pro Proc. 29th* Annual *Asilomar Conference on Signals, Systems, and Computers*, vol. I, pp. 51-55, Pacific Grove, CA, Nov. 1995.

[3] V. E. DeBrunner, W. Lou, and J. Thripuraneni, "Multiple transform algorithms for time-varying signal representation," *IEEE Trans. On Cir. & Sys.: Analog and Digital Signal Processing*, Sep. 1997.

[4] V. E. DeBrunner, L.-X. Chen, and H.-J. Li, "Lapped multiple bases algorithms for still image compression without blocking effect," *IEEE Trans. Image Proc.*, Aug. 1997.

[5] A. A. (Louis) Beex and F. G. Safar, "Multiple bases signal representation, coding, and reconstruction," *Proc. IEEE Int. Symp. On Information Theory*, p. 76, San Diego, CA, January 1990.

[6] A. P. Berg and W. B. Mikhael, "An efficient structure and algorithm for image representation using nonorthogonal basis images," *IEEE Trans. Cir. & Sys. II*, vol 44, no. 10, pp. 818-828, Oct. 1997.

[7] W. B. Mikhael and A. S. Spanias, "Accurate representation of time-varying signals using mixed transforms with application to speech," *IEEE Trans. Cir. & Sys.*, vol. 36, pp. 329-331, Feb. 1989.

[8] W. B. Mikhael and A. S. Spanias, "Efficient modeling of dominant transform components representing time-varying signals," *IEEE Trans. Cir. & Sys.*, vol. 36, pp. 331-334, Feb. 1989.

[9] A. S. Spanias and P. C. Loizou, "Mixed Fourier/Walsh transform coding scheme for speech coding at 4.0 kbits/s," *IEE Proc-I*, vol. 139, pp. 473-478, Oct. 1992.

[10] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Sig.* Proc., vol. 39, pp. 489-492, 1992.

[11] K. Engan, S. O. Aase, and J. H. Husøy, "Designing frames for matching pursuit algorithms," *ICASSP'98*, Seattle, WA, May 1998.

[12] T. S. Verma and T. H. Y. Meng, "An analysis tool for transient signals that allows a flexible sines+transients+noise model for audio," *ICASSP'98*, Seattle, WA, May 1998.

[13] V. E. DeBrunner, T. Przebinda, M. Ozaydin, "Resolution in Time-Frequency," *accepted for publication in IEEE Trans. on Signal Processing*, letter received Aug. 1998.

[14] M. V. Wickerhauser, "Adapted wavelet analysis from theory to software," A K Peters, 1994.

[15] C. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379-656, 1948.

[16] G. Jumarie, Relative information, Springer-Verlag, 1990.

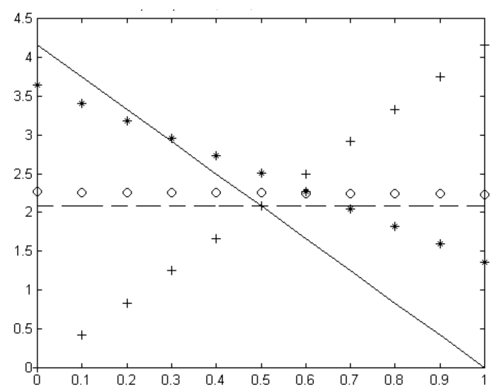

Figure 1. Identity (-), DCT (+), Haar (*), and Gaussian (o) H_p vs. p for N=64

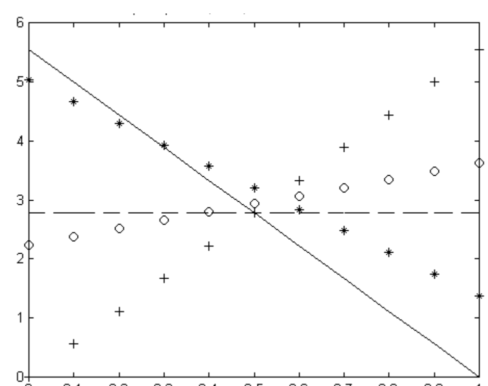

Figure 2. Identity (-), DCT (+), Haar (*), and Gaussian (o) H_p vs. p for N=256