Naval Research Laboratory

Washington, DC 20375-5320

NRL/MR/5650--09-9172

Calculation of Entropy and Mutual Information for Sinusoids

Joseph V. Michalowicz

Global Strategies Group
Crofton, Maryland

Jonathan M. Nichols
Frank Bucholtz

Photonics Technology Branch
Optical Sciences Division

January 28, 2009

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
28-01-2009	Memorandum Report	8 August 2008 – 23 January 2009

4. TITLE AND SUBTITLE

Calculation of Entropy and Mutual Information for Sinusoids

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
ONR 61135N

6. AUTHOR(S)

Joseph V. Michalowicz,* Jonathan M. Nichols, and Frank Bucholtz

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
WU EW014-02-41-8955

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

* Global Strategies Group
2200 Defense Highway
Crofton, MD 21114

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/MR/5650--09-9172

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
One Liberty Center
875 North Randolph Street
Arlington, VA 22203-1995

10. SPONSOR / MONITOR'S ACRONYM(S)

ONR

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

In this work, an analytical expression is developed for the differential entropy of a sinusoid. The mutual information between two sinusoids with different amplitudes and phase is then derived using both continuous and discrete entropies.

15. SUBJECT TERMS

Information theory
Differential entropy
Mutual information

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UL	9	Jonathan M. Nichols
Unclassified	Unclassified	Unclassified			**19b. TELEPHONE NUMBER** *(include area code)* (202) 404-5433

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Calculation of Entropy and Mutual Information for Sinusoids

Joseph V. Michalowicz[1], Jonathan M. Nichols[2,⋆], and Frank Bucholtz[2]
[1] *U. S. Naval Research Laboratory, Optical Sciences Division, Washington, D.C. 20375 (Permanent Address: SFA, Inc., Crofton, MD 21114)*
[2] *U. S. Naval Research Laboratory, Optical Sciences Division, Washington, D.C. 20375*

I. EXECUTIVE SUMMARY

In this work, an analytical expression is developed for the differential entropy of a sinusoid. The mutual information between two sinusoids with different amplitudes and phase is then derived using both continuous and discrete entropies.

II. INTRODUCTION

One of the fundamental questions in information theory is the maximum level of data compression. Shannon [1] provided the answer to this question by demonstrating that random processes have an irreducible complexity beyond which no compression is possible. This quantity is called the entropy of the random process, and this concept is currently of great interest in communication theory studies.

The concept of entropy is discussed in several books (see e.g. [2, 3]). Usually its formulation is first given for discrete random variables; that is, if X is a discrete random variable with values $\{x_1, x_2, \cdots, x_m, \cdots\}$ and probability density $p(X)$ defined by $p(x_m) = \text{Prob}\{X = x_m\}$, the entropy is expressed as

$$H(X) = -\sum_m p(x_m)\log[p(x_m)].$$

This log is usually taken to be $\log_2$ and then the entropy is given in units of bits. If the log is taken in the base e, then the entropy is written as $H_e(X)$ and is given in units of "nats". This (discrete) entropy has many interesting properties, one of which is that it is non-negative; it provides a way to quantify the information content of a probability distribution.

This concept is extended to the case of a continuous random variable, and is then called the differential entropy. For a continuous random variable X with probability density function $p(x)$, the differential entropy is given by:

$$h(X) = -\int_S p(x)\log[p(x)]dx$$

where $S = \{x|p(x) > 0\}$ is the support set of X. Again log usually means $\log_2$. If the log is taken to the base e, the notation $h_e(X)$ is utilized. Many properties of (discrete)entropy carry over to differential entropy; however, the differential entropy may take on negative as well as positive values. The differences between discrete and continuous entropy are discussed in [4] where, in fact, an alternative measure of information content is proposed.

In [2] tables of differential entropies are given for various probability distributions. However, in signal detection research the prototypical case is the detection of a sinusoidal signal in background noise. For example, suppose the signal is written as:

$$y = A\sin(\theta).$$

where θ is uniformly distributed on $[-\pi, \pi]$ (A similar discussion applies if $y = A\sin(\omega\theta)$ or $y = A\sin(\theta + \phi)$), where ω and ϕ are constants. Restricting θ to the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$ on which y monotonically increases through its range of

Manuscript approved December 12, 2008.

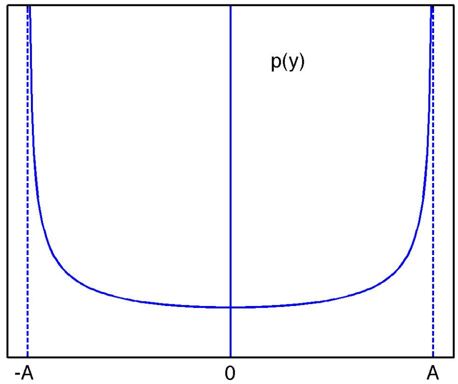

FIG. 1: Probability Density Function of a Sine Wave

values permits the use of the transformation of variables technique to calculate the probability density function of y

$$\begin{aligned} p(y) &= \frac{\frac{1}{\pi}}{|\frac{dy}{d\theta}|} \\ &= \frac{\frac{1}{\pi}}{Acos(\theta)} \\ &= \frac{1}{\pi\sqrt{A^2-y^2}} \qquad -A < y < A \end{aligned} \tag{1}$$

as depicted in Figure 1 This probability distribution has zero mean with variance $\sigma^2 = A^2/2$.

The sine wave differential entropy is important for signal detection applications and does not appear in the literature. Its computation proceeds as follows:

$$\begin{aligned} h_e(y) &= -\int_{-A}^{A} p(y)\ln[p(y)]dy \\ &= -\int_{-A}^{A} \frac{1}{\pi\sqrt{A^2-y^2}}\ln[\frac{1}{\pi\sqrt{A^2-y^2}}]dy \\ &= \frac{2}{\pi}\int_{0}^{A} \frac{1}{\sqrt{A^2-y^2}}\ln[\pi\sqrt{A^2-y^2}]dy. \end{aligned} \tag{2}$$

Making the change of variables: $w = \frac{\sqrt{A^2-y^2}}{A}$, $dw = \frac{-y}{A\sqrt{A^2-y^2}}dy$, and $y = A\sqrt{1-w^2}$ leads to

$$\begin{aligned} h_e(y) &= \frac{2}{\pi}\int_{0}^{1} \ln[\pi A w]\frac{dw}{\sqrt{1-w^2}} \\ &= \frac{2}{\pi}\int_{0}^{1} \frac{\ln(w)dw}{\sqrt{1-w^2}} + \frac{2}{\pi}\ln[\pi A]\int_{0}^{1}\frac{dw}{\sqrt{1-w^2}} \\ &= \frac{2}{\pi}(-\frac{\pi}{2}ln[2]) + \frac{2}{\pi}\ln[\pi A]\sin^{-1}(w)|_0^1 \end{aligned} \tag{3}$$

where the first integral is given in [5] (formula 4.241(7)). The differential entropy for a sine wave is therefore

$$\begin{aligned} h_e(y) &= -\ln[2] + \ln[\pi A] \\ &= \ln[\frac{\pi A}{2}] \end{aligned} \tag{4}$$

or in bits

$$\begin{aligned} h(y) &= \log_2[e]h_e(y) \\ &= \frac{1}{\ln[2]}\ln[\frac{\pi A}{2}] \\ &= \log_2[\frac{\pi A}{2}]. \end{aligned} \tag{5}$$

Note that $h(y)$ can be positive or negative, depending on the value of A.

Due to the ubiquity of analog-to-digital converters for transforming physical measurements into digital format, it is instructive to also examine the discrete entropy of the sine wave. For this purpose we construct a new discrete variable y_Q as follows:

The random variable $y = A\sin(\theta)$ ranges between $-A$ and A; so we cover the interval $[-A, A]$ with $2^N + 1$ bins, each of width $Q = \frac{2A}{2^N} = \frac{A}{2^{N-1}}$. One bin is centered about 0, then there are $1/2(2^N)$ bins for positive values with the last one centered about $+A$ and $1/2(2^N)$ bins for negative values with the last one centered about $-A$. Then we define y_Q to be the value at the center of each bin; that is:

$$y_Q = \begin{cases} i\dfrac{A}{2^{N-1}} = iQ \quad \text{for } iQ - Q/2 < y \le iQ + Q/2 \end{cases} \tag{6}$$

for $i = -2^{N-1}, \cdots, -2, -1, 0, 1, 2, \cdots, 2^{N-1}$. We now define the probability distribution for y_Q so that the probability of each value of y_Q is the same as the integral of the probability density function of the continuous random variable y over that bin. Let $P_i = \text{Prob}\ \{y_Q = iQ\}$. Then for $|i| < 2^{N-1}$,

$$\begin{aligned} P_i &= \int_{iQ-Q/2}^{iQ+Q/2} \frac{1}{\pi\sqrt{A^2 - y^2}}dy \\ &= \frac{1}{\pi}\sin^{-1}(y/A)|_{iQ-Q/2}^{iQ+Q/2} \\ &= \frac{1}{\pi}\Big[\sin^{-1}(\frac{i+1/2}{2^{N-1}}) - \sin^{-1}(\frac{i-1/2}{2^{N-1}})\Big]. \end{aligned} \tag{7}$$

At the extreme bins, i.e. for $|i| = 2^{N-1}$,

$$\begin{aligned} P_i &= \int_{A-Q/2}^{A} \frac{1}{\pi\sqrt{A^2 - y^2}}dy \\ &= \frac{1}{\pi}\sin^{-1}(y/A)|_{A-Q/2}^{A} \\ &= \frac{1}{\pi}\Big[\frac{\pi}{2} - \sin^{-1}(1 - \frac{1}{2^N})\Big]. \end{aligned} \tag{8}$$

Note the symmetry; that is, for $i \neq 0$, $P_i = P_{-i}$. Now the (discrete) entropy for the discrete random variable y_Q is given by

$$\begin{aligned} H(y_Q) &= -\sum_{i=-2^{N-1}}^{2^{N-1}} P_i \log_2[P_i] \\ &= -[P_o \log_2[P_o] + 2\sum_{i=1}^{2^{N-1}} P_i \log_2[P_i]]. \end{aligned} \tag{9}$$

To provide an example of this formula, calculations were made with the parameters of a typical analog-to-digital (A/D) converter. The device is an N-bit digitizer; 2^N bins over the full range from -1 to +1. So the bin width Q is $1/(2^{N-1})$. In this case, $A = 0.99$ represents the A/D converter full scale in order to avoid truncation at the extreme values. Table I shows the results calculated from Eqn. (9) for various values of N.

But the relationship between the discrete entropy $H(y_Q)$ and the differential entropy $h(y)$ is given in [2] as:

$$H(y_Q) + \log_2[Q] \to h(y) \text{ as } Q \to 0.$$

TABLE I: **Theoretical and Estimated Discrete Entropies for Varying Discretizations**

N	Discrete Entropy $H(y_Q)$	$(N-1)+h(y)$
2	1.96	1.637
4	3.86	3.637
6	5.81	5.637
8	7.73	7.637
10	9.66	9.637
12	11.65	11.637
14	13.65	13.637
16	15.64	15.637

TABLE II: Convergence of the discrete entropy to the differential entropy plus $N-1$

In our case, $Q = 1/(2^{N-1})$, so

$$H(y_Q) \rightarrow (N-1) + h(y) \;\text{ as } N \rightarrow \infty. \tag{10}$$

From Eqn. (5), when $A = 0.99$, $h(y) = 0.637$. The right-hand column of Table I gives for comparison the values $(N-1)+0.637$, and the expected convergence is clearly shown. For a typical A/D converter, N would range from 8 to 16.

In this paper we wish to calculate the mutual information between two sinusoids. Again, specific examples of the computation of mutual information for continuous probability distributions do not appear in textbooks. Even for the Gaussian case [2] does not specifically express the mutual information between two Gaussian distributions, although all the necessary terms are given. The following formulas are all taken from [2] for normal variables $X \in N(\mu_x, \sigma_x)$, $Y \in N(\mu_y, \sigma_y)$:

$$\begin{aligned} h(X) &= \frac{1}{2}\log_2[2\pi e \sigma_x^2] \\ h(Y) &= \frac{1}{2}\log_2[2\pi e \sigma_y^2] \\ h(X,Y) &= \frac{1}{2}\log_2[(2\pi e)^2 |K|] \\ &= \frac{1}{2}\log_2[(2\pi e)^2(\sigma_x^2\sigma_y^2 - (E[(X-\mu_x)(Y-\mu_y)])^2)] \end{aligned} \tag{11}$$

where $|K|$ is the covariance matrix. Therefore, we can express the mutual information between X and Y as

$$\begin{aligned} I(X,Y) &= h(X) + h(Y) - h(X,Y) \\ &= \frac{1}{2}\log_2[\sigma_x^2\sigma_y^2] - \frac{1}{2}\log_2[\sigma_x^2\sigma_y^2 - (E[X-\mu_x](Y-\mu_y)])^2] \\ &= \frac{1}{2}\log_2[\sigma_x^2\sigma_y^2] - \frac{1}{2}\log_2[(1-\rho^2)\sigma_x^2\sigma_y^2] \\ &= -\frac{1}{2}\log_2[(1-\rho^2)] \end{aligned} \tag{12}$$

where ρ is the correlation coefficient. This formula has been given explicitly so that we can draw attention to the two extreme special cases:

Case 1: X and Y are independent
$\rho = 0$
$I(X;Y) = 0$

Case 2: X and Y are linearly related so X and Y completely determine one another; that is, for any given X there corresponds one value of Y (so X gives complete information about Y) and vice-versa
$\rho = \pm 1$
$I(X;Y) = +\infty$

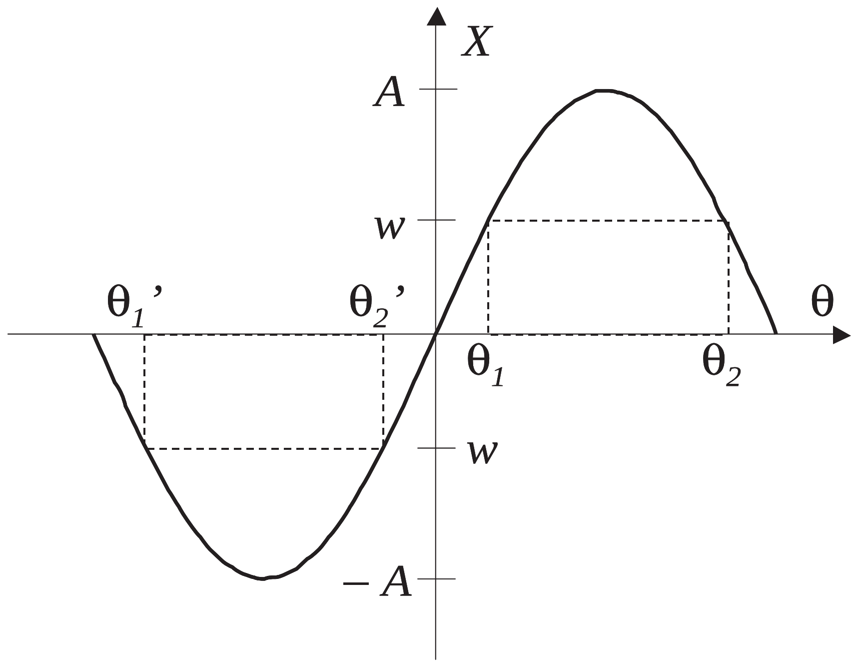

FIG. 2: Determination of two angles for fixed values of X

Now consider the case of interest here, where

$$X = A\sin(\theta)Y = B\sin(\theta+\phi)$$

are two sinusoids with θ uniformly distributed on $[-\pi,\pi]$ and ϕ a fixed phase angle between 0 and 2π. In order to compute the mutual information between X and Y, we first note that a fixed value of X corresponds to two values of Y as depicted in Figure (2). Let w be a specific value of X. If $0 \le w \le A$ then $\theta_1 = \sin^{-1}(w/A)$ and $\theta_2 = \pi - \sin^{-1}(w/A)$ and Y is equal to either $B\sin(\theta_1+\phi)$ or $B\sin(\theta_2+\phi)$. If $-A \le w \le 0$, then $\theta_1' = \sin^{-1}(w/A)$ and $\theta_2' = -\pi - \sin^{-1}(w/A)$ and Y is equal to either $B\sin(\theta_1'+\phi)$ or $B\sin(\theta_2'+\phi)$. Hence the conditional probability distribution $p(Y|X=w)$ is a discrete distribution with only two possible values for the discrete conditional random variable. Since the mutual information of two random variables is the difference between an entropy and a conditional entropy, it can be computed in either of two ways; that is, in terms of differential entropies:

$$I\left(X,Y\right) = h\left(Y\right) - h\left(Y|X\right)$$

or in terms of discretized versions:

$$I\left(X,Y\right) = H\left(Y\right) - H\left(Y|X\right)$$

But the differential entropy of a discrete random variable must be $-\infty$, for the volume of the support set in n-dimensional space that contains most of the probabilities of the variable is 2^{hn}. Since for a discrete random variable, this volume is clearly zero, the entropy h must be $-\infty$. Consequently,$h\left(Y|X\right) = -\infty$ and $I\left(X,Y\right) = +\infty$. Compare this result to the case of the two linearly-related Gaussians that also gave the mutual information to be $+\infty$. In the Gaussian case, fixing X determines a single value of Y. In the sinusoidal case, fixing X determines two values of Y, but reducing the possible values of Y from an infinite number to just two numbers still provides an "infinite amount of information" about Y.

So we have the result that we sought, but it is still informative to examine the discretized calculation of the mutual information. For this, the (discrete) entropy $H\left(Y|X\right)$ is to be computed for the discrete probability distribution $p\left(Y|X\right)$. For a fixed value w of X with $0 \le w \le A$ (the case for $-A \le w \le 0$ yields the very same formulas), the variable Y has values

$$\begin{aligned} y' &= B\sin\left(\sin^{-1}\left(w/A\right)+\phi\right) \\ &= B\left[\tfrac{w}{A}\cos\phi + \cos\left(\sin^{-1}\left(\tfrac{w}{A}\right)\right)\sin\phi\right] \\ &= B\left[\tfrac{w}{A}\cos\phi + \sqrt{1-\left(\tfrac{w}{A}\right)^2}\sin\phi\right] \end{aligned}$$

or

$$\begin{aligned} y'' &= B\sin\left(\pi - \sin^{-1}\left(\tfrac{w}{A}\right) + \phi\right) \\ &= B\left[\tfrac{w}{A}\cos\phi + \cos\left(\pi - \sin^{-1}\left(\tfrac{w}{A}\right)\right)\sin\phi\right] \\ &= B\left[\tfrac{w}{A}\cos\phi - \sqrt{1-\left(\tfrac{w}{A}\right)^2}\sin\phi\right] \end{aligned}$$

.

Now from the probability density function for Y, $p(Y)$, we have

$$\begin{aligned} p\left(y'\right) &= \tfrac{1}{\pi\sqrt{B^2-(y')^2}} \\ &= \tfrac{1}{\pi B\sqrt{\cos^2\phi - \frac{2w}{A}\sqrt{1-\left(\frac{w}{A}\right)^2}\sin\phi\cos\phi - \left(\frac{w}{A}\right)^2\cos 2\phi}} \end{aligned}$$

and likewise

$$\begin{aligned} p\left(y''\right) &= \tfrac{1}{\pi\sqrt{B^2-(y'')^2}} \\ &= \tfrac{1}{\pi B\sqrt{\cos^2\phi + \frac{2w}{A}\sqrt{1-\left(\frac{w}{A}\right)^2}\sin\phi\cos\phi - \left(\frac{w}{A}\right)^2\cos 2\phi}} \end{aligned}$$

.

Normalization then gives the marginal distribution $P\left(Y|X=w\right)$ for each w to be a discrete distribution with just two values of the variable, with probabilities $p_w = \frac{D_1}{D_1)+D_2}$, where

$$D_1 = \sqrt{\cos^2\phi + \frac{2w}{A}\sqrt{1-\left(\frac{w}{A}\right)^2}\sin\phi\cos\phi - \left(\frac{w}{A}\right)^2\cos 2\phi}$$

and

$$D_2 = \sqrt{\cos^2\phi - \frac{2w}{A}\sqrt{1-\left(\frac{w}{A}\right)^2}\sin\phi\cos\phi - \left(\frac{w}{A}\right)^2\cos 2\phi}$$

and $1-p_w$. For each w, the discrete entropy is then given by

$$H\left(p_w\right) = -p_w\log_2(p_w) - (1-p_w)\log_2(1-p_w)$$

and the conditional entropy becomes

$$H\left(Y|X\right) = \int_{-A}^{A}\frac{1}{\pi\sqrt{A^2-w^2}}H\left(p_w\right)dw.$$

Let us call this number H'''. Since $0 \le H(p_w) \le 1$, it is clear that $0 \le H''' \le 1$.

Consequently, in terms of discretized entropies, where the bins covering the range of Y are of width $Q = 1/2^{N-1}$,

$$\begin{aligned} &I\left(X,Y\right) = H\left(Y_Q\right) - H''' \\ &\to (N-1) + \log_2\left(\tfrac{\pi B}{2}\right) - H''' \\ &\to +\infty \mathrm{as} N \to \infty \end{aligned}$$

since $\log_2\left(\frac{\pi B}{2}\right)$ and H''' are finite.

III. CONCLUSIONS

This paper calculates the differential entropy for a sinusoid and compares it to its discrete version for various discretizations. The mutual information between two sinusoids differing in phase is shown to be $+\infty$.

Acknowledgements

The authors acknowledge the Office of Naval Research and the Naval Research Laboratory for providing funding for this work.

[1] C. E. Shannon, A mathematical theory of communication, The Bell System Technical Journal 27 (1948) 379–423, 623–656.
[2] T. M. Cover, J. A. Thomas, Elements of Information Theory, John Wiley and Sons, New Jersey, 2006.
[3] S. Kullback, Information Theory and Statistics, Dover Publications, Inc., New York, 1968.
[4] M. J. Schroeder, An alternative to entropy in the measurement of information, Entropy 6 (2004) 388–412.
[5] I. S. Gradshteyn, I. M. Ryzhik, Table of Integrals, Series and Products, 4th ed., Academic Press, New York, 1965.