Folding: A Method for Semantic Encoding of Error-Tolerant Data

Volkan Rodoplu
vrodoplu@ece.ucsb.edu

Balakrishnan Srinivasan
balakrishnan01@umail.ucsb.edu

Department of Electrical and Computer Engineering
University of California Santa Barbara
Santa Barbara, CA 93106

ABSTRACT

Underwater acoustic sensor networks are characterized by both low link data rates, and very low data generation rates of the sensors. In this regime, Shannon capacity results, which presume long channel codes and an infinitely long information bitstream, are not directly applicable. Further, for scientific data collection, distortion and errors are tolerable at the semantic layer. For this regime, we formulate the problem of sending k successive source symbols using n successive modulation intervals, where $n > k$. We introduce "folding" as a technique to map a k-dimensional source manifold into the n-dimensional modulation space, in order to minimize the average energy consumption per source symbol. We use an Archimedes' spiral, a helix, and Fermat's spiral, as good foldings for low-dimensional mappings, and compute the energy consumption per source symbol under each.

Categories and Subject Descriptors: C.2.0 [Communication/Networking and Information Technology]: Data Communications

General Terms: Design, Performance, Reliability

Keywords: wireless, underwater, network, source-channel coding

1. INTRODUCTION

Underwater acoustic sensor networks are characterized by a limited battery supply of the sensor node, as well as the limited data rate of the acoustic link [1]. MAC protocols for underwater sensors [2] might use sleep schedules to conserve energy; however, the resulting number of bits that can be sent in one cycle of the MAC protocol may turn out to be very small. For example, for a link data rate of 1 kbps, if the node transmits for only 1 second and goes back to sleep, only 1000 bits are transmitted. In this regime, Shannon's channel coding theorem does not directly apply because in order to drive the probability of error to zero, the proof of the theorem needs long channel codes, and 1000 bits do not typically suffice to reach those limits. Hence, in this regime, it no longer makes sense to use the word "capacity" in the Shannon sense; instead, only the number of bits that can be transmitted with a fixed probability of error $\epsilon > 0$ can be specified.

The data generation rates in underwater sensor networks can be very low as well. For example, a temperature sensor may collect only 1 temperature measurement every 15 minutes. If each temperature measurement is represented by 8 bits, there are only 8 bits to send every 15 minutes. Since real-time monitoring of the ocean is an upcoming thrust in the research community, we would like to send this single temperature measurement, through the acoustic network, as soon as the sensor has obtained it. This is doable, but certainly Shannon capacity results are not applicable in this regime, in two respects. First, sending only a small, finite number of source symbols, k, would require infinite energy to achieve an arbitrarily low probability of error on an AWGN channel. Second, the number of bits available for modulation, n, is also very limited (even though $n > k$), as described in the first paragraph. Hence, long channel codes cannot be assumed. A third important distinction is that we are interested in sending these k source symbols, using n modulation slots, with the minimum energy possible; that is, the maximization of the data rate is not the objective, since there are an extremely small number of bits to send in the first place.

The problem setting described above requires a completely new approach to send a small number of source symbols k using a larger yet small number of modulation symbols, n, with the minimum energy possible. This is an important problem in underwater sensor networks in which the link rates are low (i.e. small n), the data generation rates are very low (i.e. small k), and there is a need for real-time monitoring of ocean data. Hence, we need to develop a new understanding of this problem, and seek to derive the optimal transmission of data in this setting.

A second important aspect of underwater sensor networks is that the sensor measurements are usually numerical in nature. Examples are temperature, conductivity, pressure, chlorophyll, oxygen, and nitrogen measurements. In consultation with marine scientists, oceanographers, and limnologists, we have found that the scientists have a tolerance for a certain fraction of errors in their data. For example, if 1 out of 100 temperature measurements is completely off, they are able to discard it either by human intuition, or during statistical analysis of the data, which is carried out after days. Hence, if the communication channel results in errors

WUWNet'07, September 14, 2007, Montréal, Québec, Canada.

1 out of 1000 measurements, for example, it would not be relevant to the scientists at the semantic layer, where these data are analyzed. (At the same time, real-time monitoring of the data, albeit with this error probability, is still to be maintained.) Second, if dense sensor networks (e.g. of simple thermistor strings, which are cheap) are deployed in large-scale in the future, the errors due to the ensemble will average out. This means that the data is "error-tolerant" in sharp contrast to the data carried on the Internet. This again calls for a different approach to the design of underwater sensor data networks. In analogy with cellular phone systems, we may design the communication system such that, 1 part in 1000, we are allowed to make a drastic error. The advantage of this is that it allows us to design the system in a much more relaxed fashion, to lower the average energy of modulation. This bears resemblance to a probability of outage specification in cellular systems: to design the system with 0% outage would require enormous energy. Hence, we relax the worst-case that the system needs to handle, and focus on the BER only in the no-outage case. In our case, we will also call it "outage" to mean that a semantically drastic error has been made.

Outside of this no-outage regime, there is still flexibility in the following sense: If a temperature measurement of 27.7^0 has been mistakenly transmitted as 27.8^0, but if such errors occur a relatively small fraction of the time, this can also be tolerated. The right metric to capture this is "distortion"; that is, some average distortion of the temperature data is allowable in the no-outage regime; that is, the data is distortion-tolerant, again, up to a fraction specified by the scientists at the semantic layer. In the future, a graphical user interface (GUI) can be developed and presented to the scientists at the semantic layer, where they can specify the semantic tolerable distortion and the probability of outage (drastic error) tolerable, dynamically, and for the application at hand. This can dramatically help relax the requirements on the physical layer.

The main contribution of this paper is to formulate and solve the problem of the transmission of a small, finite number k of source symbols using n modulation symbols ($n > k$), with the minimum energy possible, and while satisfying probability of outage and average distortion constraints at the semantic layer. We solve the problem in the simplest setting possible in this paper, and indicate directions for future work.

The rest of the paper is organized as follows: We describe our assumptions in Section 2 and formulate the problem mathematically in Section 3. We state our solution method in Section 4, and present our solutions in Section 5 and 6. We discuss our conclusions in Section 7.

2. PROBLEM SETTING

In this paper, we will pick a simple yet sufficient setting to illustrate the main design principles. Assume that there is a single sensor node that collects measurements of a 1-dimensional source process, at a low rate (e.g. 4 measurements per hour), and aims to send a small number of measurements, k, across a single acoustic link, on n symbol intervals, using amplitude modulation, where $n > k$. In the analysis in this paper, we shall consider the simplest channel model, which is the AWGN channel model.

Since the sensing process itself consumes energy, there is an incentive to make the measurements at intervals that are roughly on the order of the decorrelation time of the underlying process. If this interval is used, then the subsequent measurements can be modeled as uncorrelated. For example, for pressure measurements, assume that there is an average pressure that corresponds to that depth. Then, the deviations of pressure around this average may be modeled as uncorrelated, if taken at large enough intervals. We wish to send the deviations around this average. (We will address the case of correlated measurements in time, in our future work.)

The deviations around such an average, for measurements taken from Nature, usually exhibit a Gaussian distribution. We shall address the Gaussian case in Section 7. In the analysis of this paper, we shall model these deviations as uniformly distributed on $[-c/2, c/2]$. The reason for the choice of the uniform distribution for the source model is that it is the simplest case for which we can present analytical results.

Further, we assume that there is a measurement resolution, δ, due to the measurement device. Hence, all of the measurements have a finite precision. Hence, the uniform distribution on $[-c/2, c/2]$, for the deviations around the average, becomes a discrete uniform distribution, when this finite precision of the sensor is taken into account. We call the discrete set of points on this interval, the "source constellation".

We assume that the modulation is discrete-time, but analog. That is, at each symbol interval, a constant amplitude is used for that interval, but the choice of amplitude is from the set of real numbers (i.e. continuous alphabet).

3. PROBLEM STATEMENT

Based on the assumptions in the previous section, our mathematical problem statement is as follows: We have a source alphabet that has a discrete uniform distribution at equally spaced intervals of length δ over $[-c/2, c/2]$. The source process is i.i.d. and we accumulate k source symbols, where k is finite. We would like to send these k source symbols collected, over an AWGN channel with noise spectral density σ^2, in n time slots, using amplitude modulation. Here, n is also finite, but $n > k$.

We would like to achieve a target probability of "outage" in the sense that, on average, we can tolerate a fraction P_{out} of the source symbols to be completely corrupted. When we are not in outage, we would like to achieve a target average distortion D_{target}.

What is the method that maps the collection of k source symbols to an n-dimensional modulation space, such that the probability of outage, and the average distortion targets are met, and the average modulation energy per source symbol is minimized?

4. SOLUTION METHOD

Our main idea is to map directly from the space of source symbols to the modulation space, such that (1) the nearest neighbor relationships between the source symbols are preserved, and no new nearest neighbors are introduced, (2) the average energy consumption at the output of the modulator, per source symbol, is minimized. The reason for aiming to preserve the nearest neighbor relationships of the source constellation is that we can then send it through the channel without any compression. We aim to design a map from the source space to the modulation space such that the geometry of the source space is preserved in the mapping. Hence, for example, when the AWGN disturbs a modulation symbol and shifts it to its nearest neighbor, the temperature 27.7^0 is shifted to 27.8^0, causing a small distortion.

This falls under the general category of joint source-channel coding, but it is very different from the methods that have been used in the literature [3], which have focused on multimedia applications with very long bitstreams. In that regime, the aim is to maximize the data rate subject to an average distortion constraint, whereas here, the aim is to minimize the average energy consumption under both an outage and distortion specification, for a finite number of source symbols (k) and modulation intervals (n).

Explicitly, our construction will work as follows: Assume that p is any source symbol in the source constellation. Assume, for simplicity, that $n = 2$, i.e. only two subsequent modulation symbols can be used. In the modulation space, we associate the point q with the source symbol p. Then, in order to send the source symbol p, the modulator will send (x_q, y_q) through the channel.

Our main insight to solve this problem is the following: Assume, for simplicity again, that $n = 2$; that is, there are only 2 modulation intervals. Assume also that $k = 1$; that is, a single measurement (as a deviation from an average) is to be sent. Recall that our model for these deviations is discrete uniform, with resolution δ, over a bounded interval $[-c/2, c/2]$. Now, we think of the source constellation on this interval as a string. Our problem is then to fold this string in 2-D such that no new nearest neighbors are introduced, and the string is made as compact as possible around the origin, in order to minimize the average energy consumption.

In the next section, we shall explain this joint source-channel design method for special cases.

5. FOLDING: 1-D TO 2-D

Assume that $k = 1$ and $n = 2$. We propose the following mapping to solve our problem in this case. The mapping is shown in Fig. 1. The interval $[-c/2, c/2]$ has been shaped into a spiral in the 2-D modulation space. The parameter S determines the spacing, in the modulation space, between two adjacent source constellation points. The parameter V determines the spacing between two subsequent layers of the spiral. The spiral is constructed as follows: We call the first revolution in the middle, the "kernel"; it is characterized by the parameter V as shown. At the end of this first revolution, the two ends of the kernel are at a spacing of V. After the kernel, the spiral is drawn so as to maintain a constant distance of V from the previous layer of the spiral. The length of the spiral is L; note that this is a scaled version of c, the original length of the source space. In mathematics, this is known as the Archimedes' spiral whose polar equation is $r = \frac{V}{2\pi}\theta$. To the best of our knowledge, ours is the first joint source-channel coding scheme that has used an Archimedes' spiral.

The idea is to make V large enough to satisfy the P_{out}, and to make S large enough to satisfy the D_{target} constraint. We now determine the values of these two parameters to satisfy these constraints. To this end, let γ denote the curvature of the spiral, which is approximately constant, except at and around the kernel. That is, the distance d_{min} in the modulation space, between two adjacent source points, bears the relationship $d_{min} = \gamma S$, with $\gamma < 1$. Now, solve the following equation $\epsilon \frac{2\sigma}{\gamma} argQ(\epsilon/2) = D_{target}$ for ϵ. It can be

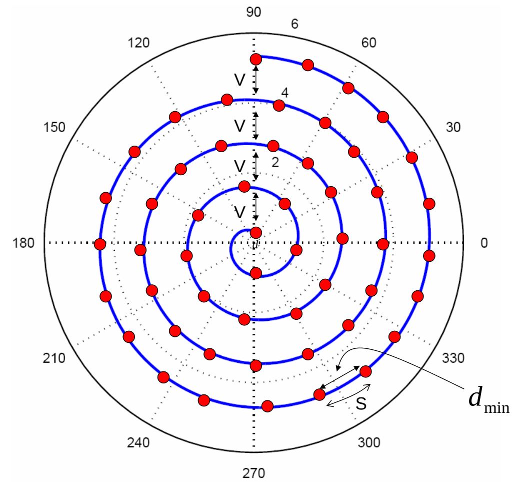

Figure 1: The proposed mapping from the 1-D source space to the 2-D modulation space

shown that there exists a unique ϵ that solves this equation. Fix this ϵ. This ϵ will be the probability of error that we need to satisfy, in order to achieve an average distortion of D_{target} at the end. Now, fix $S = \frac{2\sigma}{\gamma} argQ(\epsilon/2)$.

Then, fix $V = 2\sigma argQ(\min\{P_{out}, \beta\epsilon\})$, where $\beta \ll 1$. The idea here is that we choose the spacing V to create a distance between "global neighbors," which are neighbors on the adjacent strips of the spiral. The above choice of V both satisfies the probability of outage between such neighbors, and controls the distortion via the parameter β. Note that mistaking a source point for a point in the adjacent strip of the spiral would cause a large distortion in the source space; hence, this is controlled via β.

We shall now compute the average L_1 distortion, and the average energy per source symbol, under the above settings. We begin with the average distortion. We shall derive an approximation for large m, where m is the number of half-turns of the spiral. This eliminates the effects of the kernel, where the shape of the spiral is irregular. In addition, each point on the outermost turn of the spiral has only 1 global neighbor, whereas the points on the rest of the spiral have two such global neighbors. Analyzing it for large m makes this irregularity small, as well. Then, the average L_1 distortion is well-approximated by the distortion of a typical point, chosen around the midpoint of the string. By the nearest neighbor union bound, the average distortion at this point is ϵS, since there are only two closest neighbors, one on each side of the source point. It can be shown that the contribution of the sum of the other neighbors to distortion is much smaller, due to the properties of the $Q(x)$ function. Now, plugging in our original choice of ϵ shows that D_{target} is satisfied.

Second, we compute the average energy consumption per source symbol. We approximate each half-turn of the spiral with a semicircle. For large m, we have shown that the average energy per source symbol is given by

$$\bar{E} \approx \frac{1}{2}\left(1 - \frac{1}{m}\right)\left\{\left(\frac{a}{2} + V\left(\frac{m+1}{2}\right)\right)^2 + \left(\frac{a}{2} + V\right)^2\right\}$$

We conjecture that the spiral is an optimal folding for the 1-D to 2D mapping, since it utilizes the volume of the modulation space extremely well while making the structure very compact around the origin.

6. FOLDING: 1-D TO 3-D

Assume that $k = 1$ and $n = 3$. We conjecture that an optimal mapping in this case is one that spins the 1-D string into a spiral, but now in 3-D, in order to make it as compact as possible around the origin. However, this is very difficult to analyze mathematically. Instead, we shall exhibit a good, but suboptimal folding for this case, that can be analyzed mathematically. We propose to fold the 1-D string into a helix in 3-D, such that the height H of the helix satisfies $H = 2R$, where R is the radius of the turns of the helix. This turns the helix into as squarely compact a shape as possible. Then, we shall center the helix around the origin to minimize the average energy consumption per source symbol. The reason that the helix is suboptimal is that it wastes modulation space in the middle of the helix, into which no source symbols have been mapped.

The parameter S is the spacing between the source symbols on the helical string. V is the vertical spacing between the turns of the helix. We fix ϵ exactly as in the previous section, and let $S = \frac{2\sigma}{\gamma} argQ(\epsilon/2)$, where γ is now the curvature of the helix. Further, V is chosen by exactly the same formula as in the previous section, to avoid introducing new nearest neighbors by avoiding to place the turns of the helix too close to each other. Let L denote the total length of the helical string, which is a scaled version of the original source space length c. Then, the optimal radius R^* which makes $H = 2R$ is computed to be

$$R^* = \frac{V}{2\sqrt{2}\pi}\left(\sqrt{1+4\pi^2\left(\frac{L}{V}\right)^2}-1\right)^{1/2} \approx \frac{1}{2}\sqrt{\frac{LV}{\pi}}$$

where the approximation holds for large L.

The average distortion per source symbol is approximately ϵS for large L, and by using the same calculation as in the previous section, can be shown to be less than or equal to D_{target}. The average energy per source symbol, for any fixed R, is computed to be

$$\bar{E} = R^2 + \left(\frac{1}{\left(\frac{2\pi R}{V}\right)^2+1}\right)\frac{S^2}{6}\left(\frac{L}{2S}+1\right)\left(\frac{L}{S}+1\right)$$

and for large L and the optimal choice of R,

$$\bar{E} \approx \frac{LV}{4\pi} + \left(\frac{1}{\frac{\pi L}{V}+1}\right)\frac{S^2}{6}\left(\frac{L}{S}+1\right)\left(\frac{L}{2S}+1\right)$$

7. DISCUSSION AND CONCLUSIONS

Our aim in this paper was to formulate the problem of minimum-energy joint source-channel coding to send a finite number of source symbols k through an AWGN channel using a finite number of modulation intervals n, with $n > k$, subject to probability of outage and average distortion constraints. Our main insight was that it suffices to preserve only the *local geometry* between the points on the source constellation, and we introduced the idea of "folding" the source space as a manifold and fitting it as compactly as possible around the origin in the modulation space, in order to minimize the average energy consumption per source symbol. We presented good foldings for two special cases $(n, k) = (2, 1)$, and $(n, k) = (3, 1)$. We conjecture that the spiral folding for the first case is optimal. We presented a good but suboptimal folding for the second case, namely a helical folding. We conjecture that a spiral folding into 3-D is optimal; however, this is difficult to analyze mathematically. The folding from 2-D to 3-D can be visualized as well, as the folding of a piece of square paper around the origin in 3-D to make it as compact as possible. Turning the paper into an accordion shape appears as a good but suboptimal folding. The most intriguing open question is to find the optimal foldings for general (n, k), that is, the optimal folding from a k-dimensional manifold to $\mathcal{R}^n$.

The implementation of this scheme would require a look-up table, but the size of the look-up table need not be large since k and n are both small in these applications.

In this paper, we considered only the problem of mapping a discrete uniform distribution in the source space, to the modulation space. If, instead, the source distribution is Gaussian, the optimal folding will change since it has to map more of the probability weight in the middle, to the vicinity of the origin in the modulation space, in order to minimize the average energy consumption per source symbol. We propose Fermat's spiral [4], $r = \mp\theta^{1/2}$, as a good folding to map a 1-D Gaussian source distribution to the modulation space. Note that the middle of the string falls near the origin, where most of the probability mass of the Gaussian distribution lies, and the tails appear at the outside. To the best of our knowledge, ours is the first paper that proposes to use a Fermat's spiral for joint source-channel coding of Gaussian sources. We conjecture that a Fermat's spiral is the optimal folding to map a 1-D Gaussian source distribution into a 2-D modulation space.

Finally, after we have completely solved the problem for the AWGN channel, we plan to derive the optimal foldings for the underwater multipath channel. An optimal folding in that case must minimize the average energy per source symbol, while maintaining the geometry of the source space *locally* in a multipath channel. Many replicas of the original manifold will appear in the modulation space, like the petals of a rose, and the problem then turns into how to design the original manifold optimally such that the received set of manifolds can preserve the original geometry of the source locally.

8. REFERENCES

[1] E. Sozer, M. Stojanovic, and J. Proakis, "Underwater acoustic networks," *IEEE J. Oceanic Eng.*, vol. 25(1), pp. 72-83, Jan. 2000.

[2] M. K. Park and V. Rodoplu, "UWAN-MAC: an energy-efficient MAC protocol for underwater wireless acoustic networks", in *IEEE J. Oceanic Engineering*, to appear.

[3] Z. He, J. Cai, C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, issue 6, pp. 511-523, Jun. 2002.

[4] E. W. Weisstein, "Fermat Spiral," http://mathworld.wolfram.com /FermatsSpiral.html at MathWorld.