Data Compression Using Orthogonal Functions

J. A. Lebo


Follow this and additional works at: [http://commons.erau.edu/space-congress-proceedings](http://commons.erau.edu/space-congress-proceedings)
DATA COMPRESSION USING ORTHOGONAL FUNCTIONS

J. A. Lebo
Autonetics Division of North American Rockwell Corporation
Anaheim, California

Summary

Data Compression using representation of signals by finite series of orthogonal functions has been often discussed in the literature as giving compression ratios of twenty to one or more. Such schemes should be designed, however, based on the accuracy achieved at the receiver, including errors introduced in quantizing and transmitting the coefficient values.

A means for performing such a design is presented, and, assuming particular functions and signal characteristics, design curves are given. This procedure is compared with a sample and hold scheme illustrating significant improvements in data compression ratios with the orthogonal function approach. Experimental results verifying these theoretical relationships are presented and discussed.

Introduction

A great many papers have been written in the last few years describing means of data compression suitable for use in remote sensing systems such as spacecraft. Reference 1, for example, lists 72 representative papers on this subject. Many of these have advanced one or more specific data compression schemes and developed experimental or theoretical compression ratios. Generally overlooked, however, is the fact that the "compressed" data may have different precision requirements than the original "raw" information and hence a simple comparison of number of data values per second is not a valid measure of compression.

One method of approaching this problem is advanced by this paper. Namely, in cases where integral square error is a valid measure of signal accuracy, an optimum number of transmitted coefficients exists for a given compression scheme. Further, assuming that the compression scheme consists of representation of signals by a series of orthogonal functions, determination of this optimum is particularly easy.

This procedure will be illustrated by considering a particular example of expansion function and signal characteristic. With the availability of the modern digital computer similar results for other problems are easily achievable.

Orthogonal Function Data Compression

If \( \{f_n\} \) is a set of functions orthonormal on \((t_1, t_2)\) with respect to weighting function \(w\), then the finite series, \( S_N = \sum_{n=0}^{N} a_n f_n(t) \), that minimizes the mean square error in approximating \( y(t) \) on this interval has expansion coefficients given by \( a_n = \int_{t_1}^{t_2} f_n(t) y(t) w(t) \, dt \) and achieves error \( e^2 = \Delta \int_{t_1}^{t_2} (y(t) - S_N)^2 \, dt = \int_{t_1}^{t_2} w(t) y(t)^2 \, dt - \sum_{n=0}^{N} a_n^2 \),

This well known result can be used to design a data compression system. Such a system might, for example, compute the successive expansion coefficients for a finite data interval by forming the required integrals in real time and transmitting the resulting coefficients during the subsequent intervals. Such a system will represent well behaved data accurately with no signal memory and little buffer storage. Prime candidate orthogonal functions, are, of course, the trigonometric and orthogonal polynomial series.

The final system performance is significantly affected by selection of the appropriate expansion, the expansion interval used, the ability of the system to react to unforeseen events (such as signals outside the anticipated range), and the number of terms employed in the expansion. These factors will be discussed.

Effect of Coefficient Errors

If the expansion coefficients are corrupted by noise, due to integration errors in the original evaluation, quantization noise, or transmission errors, the total error will be increased. If the error in \( a_n \) is denoted by \( e_n \) then the error in the final reconstruction is given by:

\[
e^2 = \int_{-1}^{1} \left[ y(t) - \sum_{n=0}^{N} (a_n + e_n) f_n(t) \right]^2 w(t) \, dt \quad (1)
\]

which expands to:

\[
e^2 = \int_{-1}^{1} y^2(t) w(t) \, dt - 2 \sum_{n=0}^{N} (a_n + e_n) \int_{-1}^{1} y(t) f_n(t) w(t) \, dt + \sum_{n=0}^{N} e_n^2 \quad (2)
\]

Now using the relationships for \( a_n \) and \( e_n^2 \):

\[
e^2 = \int_{-1}^{1} y^2(t) w(t) \, dt - \sum_{n=0}^{N} a_n^2 + \sum_{n=0}^{N} e_n^2 \quad (3)
\]

or

\[
e^2 = e_0^2 + \sum_{n=0}^{N} e_n^2 \quad (4)
\]

where \( e_0^2 \) is the integral square error of the expansion using exact coefficients. This interesting result for orthonormal expansions indicates that errors in transmitted coefficients increase the integral square error of the representation by the sum of the squares of the coefficient errors.

It follows, for example, that the error reduction achieved by transmitting an additional term is equal to the square of its value. This is a useful result if the choice of number of samples is available to the transmitting processor. The error achieved at reconstruction is also available at the transmitter (assuming
errorless transmission) from knowledge of the error introduced by quantization without actual signal reconstruction.

Quantization Noise

If it is assumed that the coefficient errors are due only to quantization, the expected square error can be evaluated. In this case the coefficient error will be a uniformly distributed random variable over the interval \((-R/2M, R/2M)\), where \(R\) is the range of numbers to be represented and \(M\) is the number of quantization levels. In that case

\[
E[e^2] = 0
\]

and

\[
E[e^2] = \int_{-R/2M}^{R/2M} \frac{R}{2M} y^2 \, dy = \frac{R^2}{12M^2}
\]

for all \(n\).

Hence,

\[
E[e^2] = E[e^2_0] + \frac{R^2}{2} (N + 1)
\]

or, assuming \(M = 2^k\) as would normally be the case for a binary channel:

\[
E[e^2] = E[e^2_0] + \frac{R^2}{3} (N + 1) - 2^{-k}(k + 1)
\]

Orthogonal Polynomial Series

The general expression for mean square error with exact coefficients is:

\[
e_0^2 = \int_{-1}^{1} w(t) y^2(t) \, dt
\]

\[
- \sum_{n=0}^{N} \int_{-1}^{1} \int_{-1}^{1} w(t) w(s) f_n(t) f_n(s) y(t) y(s) \, dt \, ds
\]

Since it is assumed that this integral square error is a significant measure of performance, we will consider as a design criterion its expected value:

\[
E[e^2_0] = E[y^2] \left[ \int_{-1}^{1} w(t) \right] - \sum_{n=0}^{N} \int_{-1}^{1} \int_{-1}^{1} w(t) w(s) f_n(t) f_n(s) \zeta(t-s) \, dt \, ds
\]

where

\[
\zeta(T) = \frac{E[y(t) y(t+T)]}{E[y^2]}
\]

is the normalized autocorrelation function of the process.

In order to be more specific a particular set \(\{f_n\}\) and a particular \(\zeta(T)\) must be selected. The approach to be outlined is applicable to any such selection, of course.

Let

\[
f_k(t) = \sum_{m=0}^{n} b_m^n t^m
\]

and \(w(t) = 1\). Then \(\{f_n\}\) is the set of Legendre orthogonal polynomials.

Further let \(\zeta(t) = \exp(-at)\) (exponentially correlated signal).

Then

\[
E[e^2_0] = E[y^2] \left[ \sum_{n=0}^{N} \sum_{m=0}^{N} b_m^n \right] \int_{-1}^{1} \int_{-1}^{1} t^m s^k \exp(-a(t-s)) \, dt \, ds
\]

The integrals can be further evaluated giving a lengthy series, or, more conveniently, the entire expression can be evaluated by a digital computer using numerical integration to avoid the tedious evaluation of the closed form expression. Resulting curves of expected integral square error versus the correlation parameter \(a\) are shown in Figure 1 for several polynomial orders.

Sample and Hold Without Errors

Perhaps the simplest data compression scheme is to represent a finite interval of data with its value at the start of the interval. Such a sample and hold method requires no memory and minimum processing and is therefore an interesting alternative to compare orthogonal expansions against.

If the sampling interval is \(T\) and \(N + 1\) samples are taken we have:

\[
e_0^2 = \int_{n=0}^{N} \int_{nT}^{(n+1)T} [y(nT) - y(t)]^2 \, dt
\]
and

\[ E \left[ e_0^2 \right] = \sum_{n=0}^{N} \int_{nT}^{(n+1)T} \left[ y^2(nT) + y^2(t) - 2y(nT)y(t) \right] dt \]

\[ = 2E[y^2] \left( N + 1 \right) \int_{0}^{T} \left[ 1 - \exp(-aT) \right] dt \]  

(12)
since \( y \) is stationary.

Again selecting \( \zeta(t) = \exp(-at) \) we have;

\[ E \left[ e_0^2 \right] = 2E[y^2] \left( N + 1 \right) \left[ 1 - \exp(-aT) \right] \]  

(13)

In order to compare with the orthogonal function representation the sampling interval must be \( (N + 1)T = 2 \), so that:

\[ E \left[ e_0^2 \right] = 2E[y^2] \left[ 2 - \frac{N + 1}{A} \left[ 1 - \exp \left( \frac{-2a}{N+1} \right) \right] \right] \]  

(14)

or for \( a \ll \frac{N + 1}{2} \)

\[ E \left[ e_0^2 \right] = E[y^2] \frac{4a}{N + 1} \]  

(15)

Sample and Hold with Quantization Noise

As previously let \( \epsilon_n \) be the error in the \( n \)th transmitted value, then

\[ e^2 = \sum_{n=0}^{N} \int_{nT}^{(n+1)T} \left[ y(nT) + \epsilon_n - y(t) \right]^2 dt \]  

(16)

\[ e^2 = e_0^2 + \sum_{n=0}^{N} \int_{nT}^{(n+1)T} \left[ \epsilon_n^2 - 2\epsilon_n y(nT), 32 \right] \]  

(17)

where \( e_0^2 \) is the error-less coefficient value. Taking expected values we will assume that the second term inside the integral has zero expected value. This is essentially true for most processes of interest when \( \epsilon_n \) is quantization noise. The remaining term was shown to have expected value;

\[ E[e_n^2] = \frac{R^2}{12M^2} \]  

(18)

where \( R \) is the range of numbers to be represented and \( M \) is the number of quantization intervals. Again taking \( M = 2^k \) we have

\[ E[e^2] = E[e_0^2] + \frac{R^2}{3} \left( N + 1 \right) 2^{-2(k + 1)} \]  

(19)

or exactly the same as the result for orthogonal functions. That is, the quantization error contributes an amount depending only on the number of terms in an expected integral square error sense, at least in both cases. The result is that, since a sample and hold method requires many samples for equal accuracy, quantization errors become significant sooner requiring many more bits per sample for equal accuracy.

Optimum Compression

The net result of expansion accuracy, which increases with the number of terms, and coefficient quantization precision, which contributes additional error as the number of terms increases, is an optimum expansion order for a given signal. The accompanying examples are interpreted in terms of fixed total bit rates, since this is a not uncommon limitation placed on the designer. The same data can be used to arrive at a total rate for a given accuracy requirement or other criteria, of course.

Since the bit rate is fixed, adding additional coefficients increases the error in each. This is so because fewer bits are available to represent the coefficients, and the number of errors to be summed is increased. The result, as can be seen from Figures 2 and 3, is a relatively sharp minimum in the expected integral square error. With sufficient knowledge of the signal's statistical properties a system can operate at this optimum with significant payoff in accuracy for the given bit rate, or in reduced bit rate for prescribed accuracy. Alternatively, it will be possible to operate on the conservative side of these curves using a few extra bits to guard against unexpected signal variations. In any case, such analysis allows an intelligent choice of the system to meet the chosen criterion.

Figure 2. Polynomial Expansion
Theoretical Performance

Figures 2 and 3 are for exponentially correlated signals with variance normalized to one, and time scale normalized to (-1, 1). The range of quantization, \( R \) was chosen to be six for all examples, corresponding to \( 4\sigma \). Figure 2 shows the predicted performance using Legendre polynomials while Figure 3 shows corresponding results for the sample and hold mechanization.

It can be seen that the polynomial expansion is somewhat more efficient than the sample and hold mechanization. The maximum advantage is achieved at low values of correlation parameter (low correlation), and this difference can approach an
order of magnitude. Of course, the design is penalized for this advantage by increased system complexity.

It can also be seen that the optimum number of coefficients varies only slowly with correlation coefficient, so that precise knowledge of the sample statistics is not essential. If, for example, 64 bits are to be used, eight coefficients would be nearly optimum at a correlation coefficient of 0.01. This would sacrifice a minimum of the achievable performance at one order of magnitude increase and would be less than 3 db above the best error value even at two orders of magnitude increase.

Experimental Results

These results were verified by computer simulation. It was assumed that the signal was actually of the prescribed form (i.e., exponentially correlated with unity variance and known correlation parameter). The signal value was a normal random variable at each time instant. Figures 4 and 5 show corresponding results for one typical sample function (dashed line) and the average of twenty such samples (solid line). Each set of expansion coefficient was represented by the indicated number of bits, again assuming a range of six units for the quantization. The solid line falls almost identically on the theoretical curve (Figure 2). In all twenty samples the optimum number of coefficients was within two of the theoretically best value and the theoretically best number of coefficients produced, on the average for those cases where it was not the optimum choice for a particular sample function, only 5.6 percent greater error than the minimum achievable.

Figure 6 shows the results of applying Legendre polynomial compression to actual telemetry data. The sample correlation function was monotonically decreasing but flat at zero and definitely not exponential. Nevertheless the optimum expansion orders correspond closely to values predicted from Figure 2.
Conclusions

The efficient design of telemetry systems requires effective use of the available channel capacity by matching data transmission precision, data compression method, and required precision of signal reconstruction. A means of improving this match is to make explicit use of the knowledge of noise introduced by quantization to optimize the number of sample values and bits per sample used.

If, as in deep space probes, additional power is prohibitively expensive or impossible to achieve, a very sophisticated processor may be justified, permitting real time determination of the optimum data compression scheme from among a pre-programmed set such as selecting the number of coefficients to transmit. In near-earth and non-space applications such precise optimization of the channel is not likely to be reasonable, but the designer should nevertheless be aware of the tradeoffs to be made between number of values transmitted and precision of each value. This approach of minimizing the expected value of integral square error appears both feasible and profitable, at least in systems where integral square error is itself a meaningful error criteria.

References