Arithmetic coding for constrained noiseless communication channels.

Project Outline and Objectives

At the most general level, the project is concerned with the problem of error-free retrieval of information passed through a communication channel. Examples of such channels are given by Earth atmosphere serving as a guide for sound and radio waves, a telephone cable, a data storage device (e. g. hard disk drive or flash memory stick), to name just a few.

Given the communication channel, the engineer's task is to design the optimal (error-resistant, low-complexity, energy-efficient, etc.) communication system. In this project we will restrict our attention to *digital* communication systems: digital information (strings of zeros and ones) is converted to an analogue wave which is passed to the channel by the transmitter and converted to back to binary by the receiver.

A common feature of virtually all communication channels is the presence of a *transparency window*: some signals can be passed through the channel with less degradation than others. For example, low frequency signals transmitted over a long distance wired communication system suffer much more degradation than high frequency signals. Very high frequency signals written on a magnetic disk are difficult to read reliably due to inter-symbol interference, very low frequency signals written on the disk degrade due to the loss of synchronisation in the read channel of the drive.

An elegant way to communicate through a channel with a known transparency window is by applying *spectrum shaping* coding: instead of using all possible binary string to communicate information, one communicates only with strings whose analogue conversion 'fits' into the transparency window of the channel. If for example, the channel is not transparent for low frequencies, one communicates with binary strings such that the corresponding analogue wave form is DC-free.

The mathematical abstraction for channels transparent only for some types of signals is given by *constrained noiseless channels*: these are the definitions of binary strings which can be used to communicate through the channel. The simplest example of a constraint channel is zero-discrepancy DC-free channel: it only lets through strings of length N such that the number of ones in the string is equal to the number of zeros.

How can one communicate through such a channel? For example, we can start with information string of length $N/2$ and then insert 1 after each information bit 0 and 0 after each information bit 1. That means that we have to transmit one auxiliary bit per each information bit. The *rate* of information transfer is therefore just just $1/2$ information bits per channel use.

Can we do better than that? In other words, can we find a map of information strings to constraint-satisfying strings (codewords) such that the ratio of information string length to the codeword length is as close to one as possible? The maximum rate of information transfer through the channel is called the *capacity* of the channel. As it turns out, the capacity of zero-discrepancy DC-free channel is approximately $1 - log_2(N)/(2N)$, i. e. it can be made as close to one as required by increasing N.

Spectrum shaping codes are truly ubiquitous in modern communication devices: for example, they are used to write data onto CD's, DVD's, HDD's, see [1]. If it wasn't for these codes, information capacity of modern magnetic hard drives would probably be measured in hundreds of megabytes rather than gigabytes.

The present mini-project is concerned with the calculation of capacity for various constrained channels and performance analysis of a very interesting class of capacity-achieving codes - arithmetic codes, [3], [4]. Arithmetic codes are based on a fantastically rich idea that information should be encoded using subintervals of unit interval of length approximately equal to the information content of the string, see [2] for a pedagogical introduction. The aim of the project is to design arithmetic codes for some practically important constrained channels and analyse the complexity of encoding-decoding operations in terms of speed and the number of operation required to encode/decode a bit of information.

Required Background and Methodology

The project is concerned with the analysis of RLLDC-free constrained channels which are of paramount importance for modern HDD's which employ the perpendicular magnetic recording technology. The objective is to design an arithmetic coding algorithm for these channels, implement this algorithm in MATLAB or C and estimate the complexity of this algorithm as the function of the code block size.

The project requires a good working knowledge of basic probability theory, combinatorics and programming in C or MATLAB. The background knowledge of Markov chains, graph theory and basic information theory is a definite plus.

The student will have a chance to work on the project in the mixed academic-industrial environment collaborating with scientists designing spectrum shaping codes for the next generation of magnetic storage devices.

Research Outcomes and Outlook

By the end of the project the student will have learned about the applications of mathematics to some real life industrial problems. He/she will discover how the tools routinely applied to the analysis of complex systems (probability theory, statistical mechanics, information theory,

Mathematics Institute, University of Warwick

combinatorics on graphs and networks) can be also used to design efficient communication codes which can eventually find their way into a commercial product. In particular, he/she will acquire a good understanding of arithmetic codes used to communicate over constrained channels and explore the issues related to the implementation complexity for these codes.

References

[1] Immink, K. A. S. , *Coding Techniques for Digital Recorders*, Shannon Foundation Publishers, Netherlands (1999);

[2] D. Mackay, *Information Theory, Inference, and Learning Algorithms*, CUP 2003;

[3] G. Langdon, *An introduction to arithmetic coding*, IBM J. Res. Dev., vol. 27, No. 2, p. 107 (1983);

[4] S. Todd, G. Langdon, G. Martin, *A general fixed rate arithmetic coding method for constrained channels*, IBM J. Res. Dev., vol. 27, No. 2, p. 107 (1983);

Contact Information: Oleg Zaboronski – Mathematics Institute, University of Warwick. Phone: 22676, E-mail: Oleg.Zaboronski@warwick.ac.uk.

Mathematics Institute, University of Warwick