The Missed Message of the Cutoff Rate

Richard E. Blahut
Department of Electrical and Computer Engineering
University of Illinois

Abstract

The history of the cutoff rate is reviewed. It is argued that the cutoff rate is a fundamental constant associated with each discrete memoryless channel. It is the largest rate for which maximum-likelihood (or minimum-distance) decoding is tractable for that channel. This statement is the appropriate backdrop against which to see recent progress in coding.

1 Introduction

The cutoff rate was introduced in a long series of papers [2,3,5,6,8,9] dealing with specific algorithms for the sequential decoding. These papers were concerned specifically with algorithms for sequential decoding of convolutional codes. These early papers were concerned with the analysis of specific algorithms rather than developing a unifying information-theoretic principle. However, the cutoff rate has a universal meaning.

The cutoff rate for the discrete memoryless channel $Q$ is defined by

$$R_0 = \max_p \left[ -\log \sum_k \left[ \sum_j p_j Q_{k|j}^{1/2} \right]^2 \right]$$

in contrast to the capacity, which is

$$C = \max_p \sum_j \sum_k p_j Q_{k|j} \log \frac{Q_{k|j}}{\sum_j p_j Q_{k|j}}.$$

The cutoff rate was shown to be the maximum rate at which various known sequential-decoding algorithms for convolutional codes are computationally tractable. At rates above $R_0$, the expected computational work of these algorithms is unbounded, and (block) maximum-likelihood decoding is intractable. We argue that the cutoff rate has a universal meaning. Minimum-distance decoding, or maximum-likelihood decoding, is not tractable at code rates larger than the cutoff rate. This means that the cutoff rate is the appropriate boundary between low-rate codes and high-rate codes. Accordingly, the development of modern information codes can be seen as a means of moving beyond $R_0$.

2 Background

It is the goal of this paper to propose that the recent work on channel coding can be better understood by referring this work to the cutoff rate. Thus the scope of coding theory can be divided into two regimes, rates below $R_0$ and rates above $R_0$. A rates below $R_0$, algebraic codes and minimum-distance decoding are suitable. Above $R_0$, graph-based codes, or information codes,
(to coin a general term encompassing turbo codes, LDPC codes, polar codes, etc), and iterative decoding are suitable, and minimum bit-error decoding is the appropriate criterion.

We revisit the concept of cutoff rate, arguing that it applies not only to sequential decoding of convolutional codes, but to any algorithm for decoding any code. We will discuss the cutoff rate in formal terms as the maximum tractable rate of any code for channel Q if maximum-likelihood decoding (or its special case of minimum-distance decoding) is to be used. In particular, we can conclude that algebraic-decoding, as such, with minimum-distance decoding is not tractable beyond rate $R_0$. Thus the recent progress in coding theory, including turbo codes [1], LDPC codes, polar codes, and so on, can be seen as a result of the need to abandon maximum-likelihood and minimum-distance decoding in order to communicate at rates in the range between the cutoff rate $R_0$ and the capacity $C$. This need has also led to the notions of iterative decoding and channel polarization, and has changed the decoding performance criteria from probability of message error to probability of bit error.

3 Probability of Error

It is well-known that the probability of block decoding error for a sequence codes of rate $R_0$ larger than $R_{crit}$ is asymptotically given by

$$p_e \sim e^{-nE_t(R)}$$

where $E_t(R)$ is the well-known random-coding exponent, developed by Gallager [4]. This exponent is nonzero for $R$ smaller than the channel capacity. Notice that a sequence of codes at fixed slope rate $R$ is not a sequence of codes whose rate approaches capacity. Because it is now common to consider sequences of codes whose rate approaches capacity, it is appropriate to convert this asymptotic statement to an alternative form.

Consider a sequence of codes whose rate $R_n$ converge to $C$. Recall that $E_t(R)$ is zero at $R = C$ and has slop zero at $R = C$. Thus near $C$, $E_t(R) = A(C - R)^2$ for some $A$. Thus

$$p_e \approx e^{-nA(C-R_n)^2}.$$  

But if $p_e$ goes to zero and $C - R_n$ goes to zero, $p_e$ must behave as

$$p_e \approx e^{-n\beta}$$

where $\beta < 1/2$. Because it has been observed that polar codes satisfy this statement [7], the expression for polar codes is tight.

4 List Decoding

It follows from Forney’s list decoding exponent that any maximum-likelihood decoder has a probability of error distribution bounded by a Pareto distribution. The Pareto distribution has infinite mean if $R$ exceeds $R_0$. In particular, Forney’s bound on list size is

$$N \leq e^{-n(E(R')-R+R')}$$

where $R'$ is the conjugate rate of $R$. By arguing that $n$ must be increased until the random variable $N$ is less than one, the waiting time $n$ is an exponential random variable, and the work depends exponentially on the waiting time. Thus the computational work is a Pareto random variable. It is easy then to show that the Pareto random variable emerges with infinite-mean computational work if $R$ is larger than $R_0$.
References


