The R_0-Parameter for the Gaussian Channel

R. J. McEliece and E. R. Rodemich
Communications Systems Research Section

We first define, and then compute, the cutoff parameter R_0 for the additive white Gaussian channel. This important channel parameter seems not to have been previously computed for this important channel model, except in the case when the input is restricted to be binary.

I. Introduction

The computational cutoff parameter R_0 has lately begun to assume an important significance in communication systems. It appears in many situations to measure a given channel's "quality" in a way that is superior, from a practical standpoint, even to the capacity of the channel. It is our object in this paper to compute the R_0 parameter for the important additive white Gaussian channel (AWGC), which is, for example, the appropriate channel model for deep-space communication. This parameter is well known, when the channel input is restricted to two levels:

$$R_0 = \log_2 \frac{2}{1 + e^{-E/N_0}} ,$$

where E/N_0 is the signal-to-noise ratio (Ref. 7, Eq. 5-56). However, R_0 seems not to have computed for the AWGC when there has been no restriction on the number of channel inputs.

In the next section we shall give what we feel is the correct definition of R_0 for the AWGC, but also discuss the merits of another candidate, the quantity R_0^* discussed by Shannon (Ref. 6). In Section III we shall prove that the input distribution achieving R_0 is always concentrated at a finite number of points. Finally in Section IV we will give some numerical values of R_0.

II. A Definition of R_0 for the Gaussian Channel

The additive white Gaussian channel (AWGC) can be described as follows (Ref. 4, Chapter 4): If $\ldots, X_{-1}, X_0, X_1, X_2, \ldots$ denotes the input sequence and $\ldots, Y_{-1}, Y_0, Y_1, Y_2, \ldots$ the output sequence, we have $Y_k = X_k + Z_k$, where $\{Z_k\}$ is a sequence of independent, identically distributed (iid), mean zero, variance $N_0/2$ random variables. The input sequence is constrained in "average energy" by requiring that $\{X_k\}$ be iid, and $E(X_k^2) = A$. It is convenient, and involves no real loss in generality, to use the normalization $N_0/2 = 1$, and we shall do so. Our goal in this section is to give a defensible definition for the cutoff parameter R_0 for this channel.

We first assume that the input distribution is given, viz., that $F(x)$ is the cumulative distribution function for each of the random variables X_k. Then according to Eq. (5.53b) in Ref. 7, the cutoff parameter *with respect to F* is given by

$$R_0(F) = -\log_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x-y)^2/8} \, dF(x) \, dF(y) \tag{1}$$

Since the input distribution must satisfy the energy constraint

$$\int_{-\infty}^{\infty} x^2 \, dF(x) = A, \tag{2}$$

it is reasonable to define R_0 for the Gaussian channel by

$$R_0 = \sup \{R_0(F) : F \text{ satisfies } (2)\} \tag{3}$$

Indeed this is the definition we take, and Sections III and IV describe the solution to this mathematical optimization problem. However, one cannot assert that the quantity so defined is "R_0" for the AWGC, without discussing a competing number discussed by Shannon (Ref. 6) in 1959. This we now do.

There are numerous conjectures about the practical significance of the cutoff parameter for a given channel (see for example, Ref. 3 or 5), but there are also two provable theorems about R_0. The first theorem is that for any rate R less than R_0, then there exists a code of length n for which the error probability is bounded by $P[\mathscr{E}] \leqslant 2^{-n[R_0-R]}$ (see for example, Ref. 2, Chapter 5 for the discrete channel case of this theorem.) It is true, and not hard to prove, that this theorem is true for the AWGC with R_0 defined as in (3). However, in the paper cited above (Ref. 6), Shannon proved that this theorem remains true for what turns out to be a larger number; viz.,

$$R_0^* = \frac{\log_2 e}{2}\left[1 + \frac{A}{2} - \sqrt{1 + \frac{A^2}{4}}\right] + \frac{1}{2}\log_2\left[\frac{1}{2}\left(1 + \sqrt{1 + \frac{A^2}{4}}\right)\right] \tag{4}$$

Thus although our definition (3) is perhaps plausible, if one defines R_0 to be the largest possible intercept of a line of slope -1 which supports the reliability exponent $E(R)$ for the given channel, then the quantity (4) is the correct definition. However, there is another possible definition, which is derived from a theorem with communications significance, which favors our definition. The theorem deals with the expected number of computations needed for a sequential decoding algorithm.

In a celebrated paper on sequential decoding, Berlekamp and Jacobs (Ref. 1) showed that there exists a certain rate, called R_{comp}, which represents the supremum of all rates R such that the average number of computations made by a sequential decoder operating on a code of rate R remains bounded. They showed that $E_0(1) \leqslant R_{\text{comp}} \leqslant \hat{E}_0(1)$, where $E_0(p)$ is a certain function which depends on the channel statistics, and $\hat{E}_0(p)$ is the convex $\cap$ hull of $E_0(p)$. For the Gaussian channel, if the code being used must satisfy the average energy constraint $E(X^2) \leqslant A$, it is easy to show that the parameter $E_0(1)$ is precisely our definition (3) of R_0. Now for "ordinary" channels, the function E_0 is already convex, and so $R_{\text{comp}} = E_0(1)$. And we conjecture that this holds for the Gaussian channel too, but have not yet been able to prove it. If our conjecture proves to be correct, then our definition (3) will have been proved to be the value of "R_{comp}" for the AWGC, and incidentally will have been shown to be *strictly less* than the "R_0" for this channel.

In the next two sections, we will discuss the computation of R_0, as defined by (3).

III. A Characterization of R_0

We recall that R_0 is defined as follows. If Q is defined as the value of the program:

minimize: $$Q = \iint K(x,y) \, dF(x) \, dF(y),$$

$$\left(K(x,y) = e^{-\frac{1}{8}(x-y)^2}\right)$$

subject to: $$\int dF(x) = 1, \quad \int x^2 \, dF(x) = A,$$

where F is a distribution function. Then $R_0 = -\log_2 Q$. In this section we will show that the optimizing distribution F is discrete, i.e., has mass at only a finite number of points.

We shall use the calculus of variations to find necessary conditions that must be satisfied by an external distribution F.

If we use Lagrange multipliers μ and λ for the two side conditions, and apply a variation δF to F, the variation of the Lagrangian function

$$L = \frac{1}{2} Q + \mu \int x^2 \, dF - \lambda \int dF$$

is given by

$$\delta L = \frac{1}{2} \delta Q + \mu\delta \int x^2 \, dF - \lambda\delta \int dF$$

$$= \int \left\{ \int K(x,y) \, dF + \mu x^2 - \lambda \right\} d(\delta F)$$

Since we are looking for a minimum, δL must be $\geqslant 0$ for all admissible variations δF. Because the Lagrange multipliers account for the two integral side conditions, the only restriction on δF is that $F + \delta F$ must be an increasing function. Hence $d(\delta F)$ can be concentrated at or near one point, and must be nonnegative if this point is not in support of $dF(x)$. Hence we must have

$$\phi(x) = \int K(x,y) \, dF(y) + \mu x^2 - \lambda \geqslant 0, \text{ all } x, \qquad (5)$$

$$\phi(x) = 0 \qquad (6)$$

at all points of support of dF.

We note that $0 \leqslant \int K(x,y) \; dF(y) \leqslant 1$ for all x, and this integral approaches zero as $x \to \pm\infty$. Thus if we divide (5) by x^2 and let $x \to \infty$, we see that $\mu \geqslant 0$. Also, if x is in the support of dF the integral in (5) is positive, and by (6), $\phi(x) = 0$. Hence $\lambda > 0$. If $\mu = 0$, then as $x \to \infty$, we would have $\phi(x) \to -\lambda$, which contradicts (5). Hence in (5) we must have

$$\mu > 0, \lambda > 0 \qquad (7)$$

Now for any x in the support of dF, we see from (6) that

$$\mu x^2 = \lambda - \int K(x,y) \, dF(y) \leqslant \lambda,$$

$$x^2 \leqslant \lambda/\mu$$

Thus the mass of the distribution F all lies in a bounded interval. Thus $\phi(x)$ is analytic, and so can have only a finite number of zeroes on a bonded interval 0. But by (6) this means that F has only a finite number of support points. This is what we set out to prove.

In the next section we shall give some numerical values of the function R_0.

IV. Some Numerical Results

Once it is known that the optimizing distribution F is concentrated at a finite number of points it is possible to program a computer to calculate R_0. The number of points needed is an increasing function of the parameter A, which we denote by $n(A)$. It turns out, for example, that $n(A) = 2$ for $A \leqslant 2.38586$.

In the table below we list for $k = 2, 3, 4$ the largest value of A for which a k-point distribution is optimal:

k	$A(k)$
2	2.386
3	5.292
4	8.6913

In the next table we list the actual value for R_0, as a function of A. For reference we also tabulate Shannon's function R_0^* as given by (4).

A	R_0	R_0^*
0.0	0	0
0.5	0.1691	0.1692
1.0	0.3161	0.3169
1.5	0.4419	0.4456
2.0	0.5481	0.5583
2.5	0.6367	0.6578
3.0	0.7149	0.7464
3.5	0.7861	0.8260
4.0	0.8512	0.8982
4.5	0.9110	0.9461
5.0	0.9659	1.0247
5.5	1.0165	1.0808

A	R_0	R_0^*
6.0	1.0638	1.1330
6.5	1.1082	1.1817
7.0	1.1502	1.2274
7.5	1.1898	1.2704
8.0	1.2274	1.3111
8.5	1.2630	1.3496
9.0	1.2969	1.3861
9.5	1.3292	1.4210
10.0	1.3602	1.4542
10.5	1.3899	1.4860

These numbers should be compared to Fig. (5.18) in Ref. 7, where "R_0" is computed using a set of equally spaced points, each less in absolute value than $\sqrt{A}$. For the optimal distribution the points are not in general equally spaced, nor are they all less than $\sqrt{A}$.

References

1. Berlekamp, E. R., and Jacobs, I. M., "A Lower Bound to the Distribution of Computation for Sequential Decoding," *IEEE Trans. Inform. Theory*, IT-13, pp. 167-174, 1967.

2. Gallager, R. G., *Information Theory and Reliable Communication*, McGraw-Hill, New York, 1968.

3. Massey, J. L., "Capacity, Cutoff Rate, and Coding for a Direct Detection Optical Channel," *TDA Progress Report 42-60*, pp. 68-76, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 5, 1980.

4. McEliece, R. J., *The Theory of Information Coding*, Addison-Wesley, Reading, Mass., 1977.

5. McEliece, R. J., Rodemich, E. R., and Rubin, A. L., "The Practical Limits of Photon Communications, *DSN Progress Report 42-55*, pp. 63-67, Jet Propulsion Laboratory, Pasadena, Calif., Feb. 15, 1980.

6. Shannon, C. E., "Probability of Error for Optimal Codes in a Guassian Channel," *Bell System Tech., J.*, 38, pp. 611-656, 1959.

7. Wozencraft, J. M., and Jacobs, I. M., *Principles of Communication Engineering*, Wiley, New York, 1965.