Local fidelity, constrained codes, and the Meru Prastāra

Lav R. Varshney

INFORMATION TRANSMISSION SYSTEMS AND INFORMATION storage systems are fundamental to almost all aspects of life. Whether one is recalling one's name, storing scientific data on a DVD, or flipping through a magazine, there is an underlying information system at work. The fundamental nature of the communication problem solved by these systems is captured in the celebrated Fig. 1 of Claude Shannon's "A Mathematical Theory of Communication." Following Fig. 1, one can define a communication system to have five parts: source, transmitter, channel, receiver, and destination. In statistical characterizations of communication, messages and signals are modeled probabilistically. For consistency with the figure, the received signal is independent of the source message given the transmitted signal, and the destination message is independent of the source message and transmitted signal given the received signal.

Communication systems must meet performance objectives, such as ensuring that the destination message is a reproduction of the source message within specified fidelity. These systems must also operate with limited resources, such as bounded transmission energy. There is usually a tradeoff between the amount of resources consumed, B, and the level of distortion that can be achieved, Δ. Since it is desirable to operate at an optimal tradeoff point between these two fundamental parameters of communication systems, a characterization of optimality is needed. Casting both the fidelity criterion, which is a potentially semantic quality, and the resource criterion, which is a physical quality, in mathematical terms allows the use of mathematical techniques to determine optimality.

Information theory deals precisely with the problem of minimizing average distortion Δ under a constraint on average resource consumption B, over the design of transmitters and receivers. The theory divides communication problems into two sets: those that can be solved and those that cannot. Optimality lies at the boundary. Although Shannon's seminal works in 1948, 1949, and 1959 provided very general expressions for the optimal (Δ, B) tradeoff, evaluating the expressions for specific cases has been a formidable task. The challenge has been particularly stymieing for fidelity criteria that are subjectively significant and for resource criteria that meaningfully quantify the physical costs of signal transmission.

Digital Object Identifier 10.1109/MPOT.2007.913550

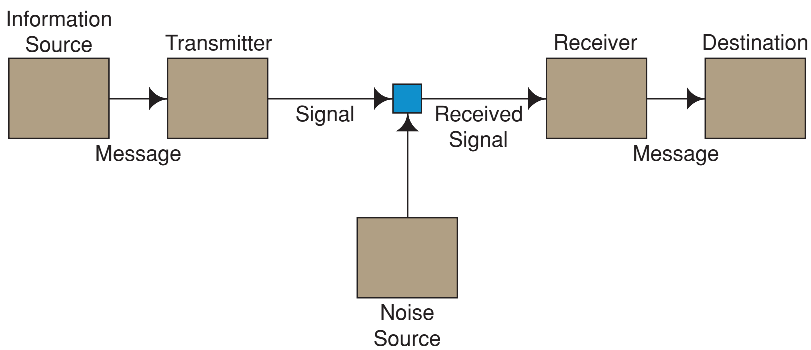

Fig. 1. Schematic diagram of a general communication system, adapted from Shannon's "A Mathematical Theory of Communication," redrawn by author.

In engineering theory, there are often two distinct practices: the development of a theoretical paradigm and the performance of computations within the paradigm. In this article, the basic paradigm of information theory is first reviewed and interpreted. The main focus, however, is on the practice of computing and showing that computation is possible even under meaningful fidelity and resource criteria. Precise computations of optimal tradeoff points for some context-dependent fidelity criteria and resource criteria that arise in optical and magnetic recording are provided; some of these results do not appear to have surfaced explicitly in the literature before. Interestingly, ancient Indian combinatorial results are useful in this setting. Moreover, we point out that the strong relationship between the capacity of constrained codes for recording channels and rate distortion for sources with respect to context-dependent fidelity criteria may be useful in the computation of other rate-distortion functions.

INFORMATION THEORY

Information theory establishes fundamental bounds on the tradeoff between the average distortion and the average cost for a communication problem. Mathematizing the description of a communication system proceeds as follows. Model the information source as an infinite sequence of random variables, $U_1, U_2, \ldots, U_k, \ldots$, which are drawn from a common alphabet according to some distribution, $p_U(u)$. The transmitter converts the source symbols into channel input symbols $X_1, X_2, \ldots, X_n, \ldots$, whose distribution is determined by the source and the transmitter transition probability assignment, $p_{X|U}(x|u)$. The channel provides the physical link between the source and destination and may be restricted to allow only a finite number of states or may introduce noise or both. The mathematical model of the channel is defined as a sequence of transition probability assignments between the channel input space and the channel output space, $p_{Y|X}(y|x)$, where the channel output letters are denoted $Y_1, Y_2, \ldots, Y_n, \ldots$. A cost function, $b(x_1, \ldots, x_n)$, is defined on the channel input alphabet; formally all that we require is that it is a nonnegative function. However, it is desirable that it be based on some fundamental system resource such as energy, bandwidth, volume, or money. The last step before the destination is the receiver. In reversing the operations of the transmitter, the receiver converts the received signal into the form of the original message. It is a transition probability assignment, $p_{V|Y}(v|y)$, that yields the reconstructed message $V_1, V_2, \ldots, V_k, \ldots$. A distortion function $d(u_1, \ldots, u_k; v_1, \ldots, v_k)$ that measures how bad of a reconstruction V is for U is defined for the source-destination pair.

Now returning to our fundamental system parameters, average resource consumption B is defined to be the expected value of b and the average distortion Δ is defined to be the expected value of d, the fidelity criterion. For a given source, fidelity criterion, channel, and resource criterion, and over the choice of the transmitter and receiver, if an information system satisfies the following conditions:

1) Δ cannot be decreased without increasing B, and

2) B cannot be decreased without increasing Δ,

then it is optimal. The optimization over the transmitter and receiver does not impose any further design restrictions on these transducers. A cost distortion curve may be drawn, which gives all system performances (Δ, B) that are optimal. This curve is given by the set of points that simultaneously satisfies the two optimization problems derived from the definition of optimality. Although there is a definition and, in some sense, a formula for an optimal communication system, this description is not at all tractable.

Shannon, however, came to the rescue and provided a much more tractable characterization of optimality. One can think of a transmitter as acting to both compress the source data and to protect it against channel noise; what Shannon showed was that separating the steps of compression and error protection does not reduce the set of achievable (Δ, B). By promulgating what is now called the separation theorem, he suggested an architecture for communication systems where the notion of information rate emerged. This notion of information rate, measured through the mutual information functional in units of bits, has become the primary commodity for communication systems. As Jerry Wiesner of MIT once said, "Before we had the theory, ...we had been dealing with a commodity that we could never see or really define. We were in the situation petroleum engineers would be in if they didn't have a measuring unit like the gallon. We had intuitive feelings about these matters, but we didn't have a clear understanding."

The separation principle splits the problem of optimal transmitter/receiver design into two subproblems: the design of optimal source encoders and decoders for compression and the design of optimal channel encoders and decoders for error protection. Source coding and channel coding are very closely related problems: the first removes redundancy whereas the second adds redundancy back. To take an example from the comics section of the newspaper, one can think about the puzzle Sudoku in both source coding and channel coding terms. As a source code, one thinks of the source message as the fully completed grid of numbers; since a small subset of these numbers uniquely determines the grid, the rest can be erased and not represented, thereby reducing the rate requirement. Alternatively, one can think of the completed grid as a channel code word generated from a small subset of numbers, which passes through an erasure channel to yield the puzzle; then the grid constraints help recover the completed grid from the puzzle, thereby providing error protection. (In fact Sudoku has strong connections to low-density parity check codes and related satisfiability problems in computer science.)

The coding theorems of information theory show that the performance of optimal solutions to the source-coding and channel-coding problems can be expressed in terms of the rate distortion function and the capacity cost function, respectively. The rate distortion function of a source-destination pair represents the minimum information rate of the source message required to reproduce the destination message with average distortion not exceeding Δ. The capacity cost function of a channel represents the maximum information rate of the transmitted signal that can be received across the channel with average cost not exceeding B. To express these two functions, we need to introduce the mutual information functional by which information rate is defined. The mutual information between two random variables W and Z with joint distribution $p_{W,Z}$ is

$$I(W; Z) = \sum_{w} \sum_{z} p_{W,Z}(w, z) \log \frac{p_{W,W}(w, z)}{p_W(w) p_Z(z)},$$

where the sums are taken over the alphabets on which the distributions are defined. The rate distortion function is the minimum information rate between the source and destination messages while still meeting the average distortion constraint. Mathematically,

$$R(\Delta) = \min_{p_{V|U}(v|u) E[d(u;v)] \le \Delta} I(U; V).$$

The capacity cost function is the maximum information between the transmitted and received signals while staying within the average cost constraint. Mathematically,

$$C(B) = \max_{p_X(x) E[b(x)] \le B} I(X; Y).$$

Combining these two functions yields a characterization of the performance of an optimal communication system. A communication system is optimal if and only if $R(\Delta)$ is equal to $C(B)$.

For point-to-point communication, the problem of optimal communication may be reduced to the problem of matching the information rate of the source to the information rate of the channel. Note, however, that such a separation-based scheme implied by the basic result is not required for optimality, and even uncoded transmission may be optimal in certain scenarios. In a separation-based approach, all of the distortion is incurred in the source encoder, whereas in other system architectures, some of the distortion may be incurred in the channel or elsewhere. The notion of information rate, however, is directly tied to a separation-based architecture of communication.

As a consequence of the rate matching interpretation of optimal communication, information may be understood in the same manner as fluid flow. The paradigm of electricity as a fluid has been central to the development of electrical engineering theory and has allowed connections with other branches of engineering theory such as acoustics, mechanics, thermics, and hydraulics. Although the rate matching characterization will be further developed in the sequel, the reader must be warned that there are several communication scenarios where this characterization does not apply. To use the terminology of electricity theory, the field problem cannot always be reduced to a circuit problem, as it is for simple point-to-point communication.

MEANINGFUL FIDELITY AND RESOURCE CRITERIA

Optimal point-to-point communication requires $R(\Delta)$ to equal $C(B)$. The requirement of equality seems to settle the entire question of optimal communication, but there are still several outstanding questions. Rate distortion has been defined in terms of a distortion function d and capacity cost has been defined in terms of a cost function b, but where do these come from? Moreover, once we define the distortion and cost functions, how might one go about performing the mutual information optimizations? We address the first question in this section and the second question in later sections.

As Shannon noted in his 1949 paper, "various different points may represent the same message, insofar as the final destination is concerned. For example, in the case of speech, the ear is insensitive to a certain amount of phase distortion. Messages differing only in the phases of their components (to a limited extent) sound the same. This may have the effect of reducing the number of essential dimensions in the message space. All the points that are equivalent for the destination can be grouped together and treated as one point." This is the basic essence of distortion function design: determining how much the destination cares and using this knowledge to compress.

Gray has listed several desirable properties for distortion functions to possess. These desiderata include tractability, computability, and subjective significance. If one is compressing a database, it seems reasonable not to care about the order of the records, as long as each is preserved without error. Such an unordered distortion function is easy to compute, since it simply involves sorting into some prespecified order and then comparing record by record. Not only is this distortion measure computable, it is quite tractable so an information theory for unordered data follows.

If one is compressing an image, one would want the distortion measure to reflect how objectionable a human observer finds the compressed version as compared to the original version. Making such an evaluation of human perception, however, requires actually testing with humans and is thus not computable. One might approximate human perception with a mathematical function, as several so-called perceptual distortion measures do, to obtain computability. Unfortunately, distortion functions that match perception often do not lead to tractable information theory.

A similar story plays out with respect to cost functions. Computable and tractable functions that also reflect the true costs of signal transmission are desired. In the case of signal transmission, simple functions such as signal power may actually turn out to be the true cost, but often this is not the case. For example, if an amplifier has a maximum value, then a signal power above that level would presumably have very large costs.

To ensure tractability, almost all work in information theory has concentrated on single-letter distortion func-

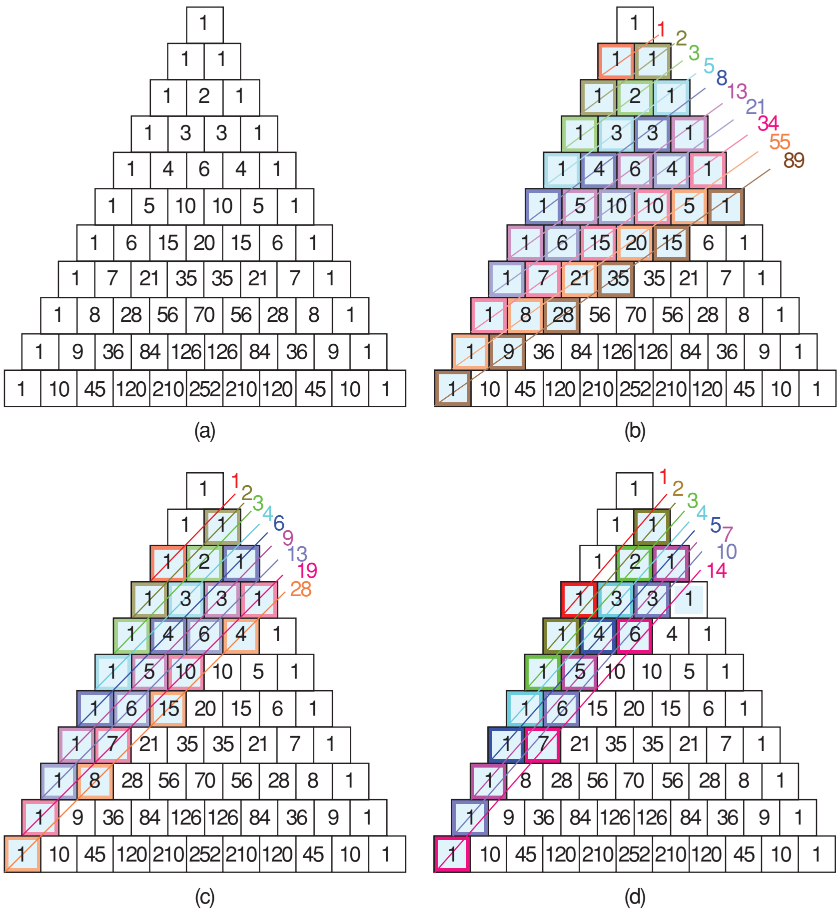

Fig. 2. The Meru sequences are found by summing all the numbers along parallel diagonals of the Meru prastāra of Pingala. Diagonals are shown by outlining boxes with the same color. (a) The Meru prastāra of Pingala, (b) shifted first Meru sequence: 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, . . . (c) second Meru sequence: 1, 2, 3, 4, 6, 9, 13, 19, 28, 41, 60, …, and (d) Fourth Meru sequence: 1, 2, 3, 4, 5, 7, 10, 14, 19, 26, ...

tions and single-letter cost functions. Examples of distortion functions include squared error and absolute error, whereas examples of cost functions include amplitude and energy. Theoretical work using nonsingle-letter cost and distortion functions has been rare, but includes our work on distortion functions where order is irrelevant as well as scenarios described in the sequel.

THE COMBINATORICS OF THE MERU PRASTĀRA

Before embarking on a development of rate distortion and capacity cost theory for meaningful nonsingle-letter functions, we take an aside to present some ancient combinatorial results. As noted by Singh, the basic results were first given in Pingala's book on rhythm and meter in Sanskrit poetry, dated to perhaps 450 B.C. The basic idea is that several recurrent sequences can be formed by summing numbers along parallel diagonals of Pingala's Meru prastāra, as shown in Fig. 2. Meru prastāra may be translated as the staircase of Mount Meru and perhaps less lyrically as the triangle expansion; it is also known as Pascal's triangle.

Taking a shallow set of diagonals we obtain the sequence 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89,..., which is called the first Meru sequence a_m; it is also known as the Fibonacci sequence. Taking a steeper set of diagonals yields the second Meru sequence $b_m = 1, 1, 1, 2, 3, 4, 6, 9, 13, 19, 28, 41, 60, \ldots$. Finally, taking an even steeper set of diagonals yields the fourth Meru sequence $c_m = 1, 1, 1, 1, 2, 3, 4, 5, 7, 10, 14, 19, 26, \ldots$. The Meru sequences are sometimes called generalized Fibonacci sequences.

One can show that these Meru sequences have nice recurrence formulas and limiting values that will be denoted as the Meru constants. These are well known for the first Meru sequence, considering the mystical qualities attached to it in popular works such as *The Da Vinci Code*. The first Meru constant is also known as the golden ratio. The expressions may be less well known for the second and fourth Meru sequences, so we provide them here. A shifted version of the second Meru sequence satisfies the recurrence formula

$$b_m = b_{m-1} + b_{m-3},$$

with initial conditions $b_1 = 1, b_2 = 2, b_3 = 3$. It satisfies the limiting relation $m^{-1} \log b_m \to \log \mathrm{meru}_2$, where

$$\mathrm{meru}_2 = \frac{1}{3} \times \left(\sqrt[3]{\frac{29 + 3\sqrt{93}}{2}} + \sqrt[3]{\frac{29 - 3\sqrt{93}}{2}} + 1 \right) = 1.465571 \ldots .$$

A shifted version of the fourth Meru sequence satisfies the recurrence formula

$$c_m = c_{m-1} + c_{m-4},$$

with initial conditions $c_1 = 1, c_2 = 2, c_3 = 3, c_4 = 4$. It satisfies the limiting relation $m^{-1} \log c_m \to \log \mathrm{meru}_4$, where $\mathrm{meru}_4 = 1.3802775691 \ldots$. These asymptotic values will be useful for computing capacity and rate distortion.

Of course, combinatorics and number theory have developed significantly since the time of Pigala. Several other information theoretically relevant asymptotic formulas in combinatorial analysis are given by Guibas and Odlyzko as well as by Immink.

RATE DISTORTION FOR CONTEXT-DEPENDENT FIDELITY CRITERIA

As first suggested by Shannon in 1959 and further developed by Berger and Yu, local distortion functions, which are defined on sliding windows of letters, may be semantically significant as well as computable and tractable. When storing phone numbers, every digit is equally important, but when storing scientific measurements, a mistake like 3215 → 3219 is probably not as bad as a mistake like 5123 → 9123. Certain kinds of errors must be avoided more stringently in compression than others, much like unequal error protection in error-control coding. Distortion functions that take such considerations into account are not single letter; rather they yield context-dependent fidelity criteria. In general, a local distortion measure of span s is any function $\rho_s : \mathcal{U}^s \times \mathcal{V}^s \to [0, \infty)$ and for block lengths $k > s$, the local distortion measure induces a block distortion function, $d_k : \mathcal{U}^k \times \mathcal{V}^k \to [0, \infty)$ of the form

$$d\left(u_1^k; v_1^k\right) = \sum_{i=1}^{k-s+1} \rho_s\left(u_i^{i-s+1}, v_i^{i-s+1}\right),$$

the sliding sum of ρ_s. Berger and Yu considered in detail the case of binary source and reproduction alphabets with the local distortion function of span 2 given in Table 1 for $d > 0$. This distortion function is derived from the desire to avoid two consecutive mistakes but without worrying about isolated errors; this seems quite reasonable for many applications. For example, in a tracking or navigation application, one might lose track of what is going on when errors burst together too closely.

Table 1. Local distortion measure ρ_2 for binary strings reproduced by binary strings. A penalty is charged only if two errors occur in a row; isolated errors are not penalized.

	(0,0)	(0,1)	(1,0)	(1,1)	(v_1, v_2)
(0,0)	0	0	0	d	
(0,1)	0	0	d	0	
(1,0)	0	d	0	0	
(1,1)	d	0	0	0	
(u_1, u_2)					

Table 2. Local distortion measure ρ_3 for binary strings reproduced by binary strings. A penalty is charged only if two errors occur in a row or for an error pattern error–no error–error.

	(0,0,0)	(0,0,1)	(0,1,0)	(0,1,1)	(1,0,0)	(1,0,1)	(1,1,0)	(1,1,1)	(v_1, v_2, v_3)
(0,0,0)	0	0	0	d	0	d	d	d	
(0,0,1)	0	0	d	0	d	0	d	d	
(0,1,0)	0	d	0	0	d	d	0	d	
(0,1,1)	d	0	0	0	d	d	d	0	
(1,0,0)	0	d	d	d	0	0	0	d	
(1,0,1)	d	0	d	d	0	0	d	0	
(1,1,0)	d	d	0	d	0	d	0	0	
(1,1,1)	d	d	d	0	d	0	0	0	
(u_1, u_2, u_3)									

Using various information theoretic arguments, Berger and Yu showed that determining $R(\Delta = 0)$ reduces to a combinatorics problem; the value $R(0)$ is essentially independent of the source distribution. To find this value, binary strings that avoid the patterns that induce nonzero distortion must be enumerated. Taking the logarithm and normalizing will then directly yield $R(0)$. For the context-dependent fidelity criterion generated by ρ_2, the number of binary strings of length n is exactly $2^{-n}a_n$, where a_n is first Meru sequence. Thus an exact computation of the rate distortion function follows readily from the Meru combinatorics. $R(0)$ is $\log(2/\text{meru}_1)$.

Rather than just caring about two consecutive errors and not worrying about isolated errors, one might want to avoid two mistakes within three source letters but allow other mistakes. This leads to the context-dependent fidelity criterion associated with the local distortion function of span 3, ρ_3, shown in Table 2. Extending the work of Berger and Yu, it can be shown that the zero distortion point $R(0)$ is $\log(2/\text{meru}_2)$. This is because the second Meru sequence corresponds to the enumeration of binary strings that avoid the patterns "11" and "101."

Similarly, if ρ_4 is defined such that there is positive distortion only when there are two errors within four source letters, the enumeration problem is one of binary strings that avoid substrings "11," "101," and "1001." This distortion measure is a natural extension to the previous two. As one might guess, this enumeration problem corresponds to the fourth Meru sequence and so the rate distortion value is $R(0) = \log(2/\text{meru}_4)$. Thus we see there are instances where the rate distortion function can be evaluated for non-single-letter distortion functions.

CONSTRAINED CODES

Just as for rate distortion, there are cases where capacity cost evaluation reduces to a combinatorial problem. This is notably the case for the noiseless channels that Shannon first described in 1948 and is discussed in great depth by Immink in the context of optical and magnetic information storage. Although not the standard description of constraints for recording channels, local cost functions paralleling local distortion functions can be defined so as to impose constraints. A local cost measure of span s is any function $\upsilon_s : \mathcal{X}^s \to [0, \infty)$ and for block lengths $n > s$, the local cost measure induces a block cost function, $b_n : \mathcal{X}^n \to [0, \infty)$ of the form

$$b\left(x_1^n\right) \sum_{j=1}^{n-s+1} \upsilon_s\left(x_j^{j-s+1}\right),$$

the sliding sum of υ_s.

The most common form of constraint imposed on noiseless magnetic or optical recording channels is a run length constraint. Run length-limited sequences are characterized by two parameters that specify the minimum and maximum lengths of runs of zeros or of ones that are allowed. The choice of these parameters implicitly take channel properties such as channel response, jitter, and noise into account; optimal codes for these channels are used to store information on CDs and DVDs. In the traditional notation of run length limited sequences, a d constraint imposes that two ones are separated by a run of consecutive zeros of length at least d. The local cost function υ_2 corresponding to a $(d = 1)$-sequence is given in Table 3. Similarly, the local cost function υ_3 corresponding to a $(d = 2)$-sequence is given in Table 4 and the local cost function υ_4 corresponding to a $(d = 3)$-sequence is given in Table 5.

Table 3. Local cost function υ_2 for binary input channel corresponding to a (d=1) runlength constrained sequence.

(x_1, x_2)	$b\,(x_1, x_2)$
(0,0)	0
(0,1)	0
(1,0)	0
(1,1)	b

To meet the cost constraint $B = 0$, for any $b > 0$, only channel inputs that use zero-cost channel input sequences will be used. Moreover, to maximize the rate, $C(B = 0)$, these will be used equiprobably. The problem reduces to an enumeration of zero-cost strings. As essentially shown by Immink, the enumeration of binary strings with zero-cost for υ_2 is the first Meru sequence, for υ_3 is the second Meru sequence, and for υ_4 is the fourth Meru sequence. Consequently, using the asymptotic properties of the Meru sequences, it follows that the zero-cost capacities of the noiseless channels are $C(0) = \log \text{ meru}_1$, $C(0) = \log \text{meru}_2$, and $C(0) = \log \text{meru}_4$ for υ_2, υ_3, and υ_4, respectively.

For cost constraints $B > 0$, codes that Immink calls weakly constrained run length limited could be used. Just as for the rate distortion problem with context-dependent fidelity criteria that we presented earlier, closed form expressions for these nonsingle-letter, capacity cost points are available.

CONCLUSION

Information theory provides fundamental bounds on the tradeoff between achievable distortion in communication and the amount of resources that are allocated to the communication system. For optimal point-to-point communication, this is a condition on rate matching. Most theoretical computation work has focused on single-letter fidelity and resource criteria due to tractability. However, these may not reflect true system objectives and constraints. So, it is desirable to extend analysis to non-single-letter fidelity and resource criteria. In certain cases, the problem can be reduced to combinatorics and becomes tractable.

Once the problem is reduced to combinatorics, there are all kinds of asymptotic enumeration methods that can be used to compute capacity and rate distortion: Shannon's original spectral methods, the Delést-Schützenberger-Viennot method from automata theory, techniques based on generating functions, or even ancient techniques involving the Meru constants. It is perhaps surprising that the Meru constants make appearances here, as they do in several other communications problems such as group testing and other forms of search, distributed consensus in sensor networks, and many others.

Table 4. Local cost function υ_3 for binary input channel corresponding to a (d=2) runlength constrained sequence.

(x_1, x_2, x_3)	$b\,(x_1, x_2, x_3)$
(0,0,0)	0
(0,0,1)	0
(0,1,0)	0
(0,1,1)	b
(1,0,0)	0
(1,0,1)	b
(1,1,0)	b
(1,1,1)	b

Table 5. Local cost function υ_4 for binary input channel corresponding to a (d=3) runlength constrained sequence.

(x_1, x_2, x_3, x_4)	$b\,(x_1, x_2, x_3, x_4)$
(0,0,0,0)	0
(0,0,0,1)	0
(0,0,1,0)	0
(0,0,1,1)	b
(0,1,0,0)	0
(0,1,0,1)	b
(0,1,1,0)	b
(0,1,1,1)	b
(1,0,0,0)	0
(1,0,0,1)	b
(1,0,1,0)	b
(1,0,1,1)	b
(1,1,0,0)	b
(1,1,0,1)	b
(1,1,1,0)	b
(1,1,1,1)	b

Finally, we saw that there are strong connections between the enumeration of run length constrained sequences and rate distortion problems. There is a duality relation between the problems so that $R(0) + C(0) = \log(2/\text{meru}_1) + \log \text{meru}_1 = 1$ for ρ_2 and υ_2 and similarly for the other local distortion and cost functions we defined. Considering that methods of computing noiseless channel capacity are reasonably well developed, perhaps the results can be used to shed more light on source coding with respect to context-dependent fidelity criteria.

READ MORE ABOUT IT

• T. Berger and W.C. Yu, "Rate-distortion theory for context-dependent fidelity criteria," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 3, pp. 378–384, May 1972.

• T.M. Green, "Recurrent sequences and Pascal's triangle," *Math. Mag.*, vol. 41, no. 1, pp. 13–21, Jan. 1968.

• R.M. Gray, *Entropy and Information Theory*. New York: Springer-Verlag, 1990.

• L.J. Guibas and A.M. Odlyzko, "String overlaps, pattern matching, and nontransitive games," *J. Comb. Theory, Ser. A*, vol. 30, no. 2, pp. 183–208, Mar. 1981.

• K.A.S. Immink, *Codes for Mass Data Storage Systems*. Eindhoven: Shannon, 2004.

• Pingala, *Chandah-sūtra*, c. 450 BC, reprinted as Pinglacarya, *Chandahsastram*, Delhi: Parimal Publications, 2001.

• C.E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, July/Oct. 1948.

• C.E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949.

• C.E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *IRE National Convention Rec*, Part 4, pp. 142–163, 1959.

• A.N. Singh, "On the use of series in Hindu mathematics," *Osiris*, vol. 1, pp. 606–628, Jan. 1936.

• L.R. Varshney, "Optimal information storage: nonsequential sources and neural channels," S.M. thesis, Dept Electrical Engineering and Computer Science, Massachusetts Institute of Technology, June 2006.

ABOUT THE AUTHOR

Lav R. Varshney (lrv@mit.edu) earned a B.S. degree with honors in electrical and computer engineering (magna cum laude) from Cornell University, Ithaca, New York, and an S.M. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (M.I.T.), in Cambridge, where he is currently pursuing a doctoral degree. He is a National Science Foundation graduate research fellow and a research assistant in the Laboratory for Information and Decision Systems and the Research Laboratory of Electronics at M.I.T.