Capacity of the generalized PPM channel

Jon Hamkins, Matt Klimesh, Bob McEliece, Bruce Moision

December 2, 2003

Abstract

We show the capacity of a generalized pulse-position-modulation (PPM) channel, where the input vectors may be any set that allows a transitive group of coordinate permutations, is achieved by a uniform input distribution. We derive a simple expression in terms of the Kullback-Leibler distance for the binary case, and find the asymptote in the PPM order. We prove a sub-additivity result for the PPM channel and use it to show PPM capacity is monotonic in the order.

1 Introduction

NASA is currently developing the first operational deep space optical communications link for launch on the Mars Telesat Orbiter in 2009. The deep space optical channel is well modeled as memoryless and operates efficiently at large peak to average power ratios, which may be efficiently implemented with pulse-position-modulation (PPM)[1, 2], in which each channel symbol is a unit vector. PPM satisfies the property that each symbol may be obtained as a permutation of the coordinates of another. We consider a generalization of this, where the input vectors may be any set that allows a transitive group of coordinate permutations. We derive an expression for the capacity of this generalized PPM channel in the binary case, and examine the behavior of the capacity of the PPM channel as a function of the PPM order.

In Section 2 we show that for a memoryless generalized PPM channel, capacity is achieved with equiprobable inputs. We show that a simple expression for the capacity follows for the binary case, and illustrate the asymptotic behavior in large n for the PPM channel. In Section 3 we prove a sub-additivity result for the PPM channel and show that certain monotonic behavior follows.

2 Capacity of generalized PPM

We use X, Y to denote random variables and x, y their realizations. Similarly, we let $\mathbf{X}, \mathbf{Y}$ and $\mathbf{x}, \mathbf{y}$ denote n-vectors of random variables and their realizations. Let $p_{Y|X}(y|x)$ be the conditional density

(or probability mass) function of a memoryless channel, and $p_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}|\mathbf{x})$ its n-th extension. When it's clear from the context, we simply write $p(y|x)$ or $p(\mathbf{y}|\mathbf{x})$ for $p_{Y|X}(y|x)$ and $p_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}|\mathbf{x})$.

Let $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_s\}$ be a set of length n vectors and $p_{\mathbf{X}}(\cdot)$ a probability distribution on S. The S-capacity of the channel is defined as

$$C_S = \max_{p_{\mathbf{X}}} I(\mathbf{X}; \mathbf{Y}),$$

i.e., the capacity with inputs restricted to S. Let G be a group of coordinate permutations that fix S, i.e., such that for each $g \in G$, $gS = S$. If, in addition, G acts transitively on S then we call S a transitive set. (G acts transitively on S if for each $\mathbf{x}_i, \mathbf{x}_j \in S$ there exists $g \in G$ such that $\mathbf{x}_i = \sigma_g(\mathbf{x}_j)$, where σ_g is the mapping imposed by g).

The capacity of channel whose input is a transitive set follows from the well-known fact that $I(\mathbf{X}, \mathbf{Y})$ is convex-$\cap$ in the input distribution p [3, Theorem 4.4.2].

Theorem 1 *If S is a transitive set, then C_S if achieved by a uniform distribution on S.*

Proof: Let p be a distribution on S and for $g \in G$ let p^g be given by $p^g(\mathbf{x}_i) = p(\sigma_g(\mathbf{x}_i))$. Clearly any p^g produces the same mutual information as p. Thus by Jensen's inequality, $(1/|G|)\sum_{g\in G} p^g$ yields mutual information greater than or equal to that yielded by p. But this new distribution is simply the uniform distribution: for any $\mathbf{x}_i$ we have

$$\frac{1}{|G|}\sum_{g\in G} p^g(\mathbf{x}_i) = \frac{1}{|G|}\sum_{g\in G} p(\sigma_g(\mathbf{x}_i)),$$

and as g ranges over G, $\sigma_g(\mathbf{x}_i)$ ranges over each element of S the same number of times (by the Orbit Stabilizer Theorem, e.g., [4, Theorem 8.2]) thus the above quantity is equal to $1/s$. □

2.1 Binary inputs

With binary inputs, C_S reduces to a simple expression. Let the input alphabet be $\{0, 1\}$, and $\mathcal{I}_0(\mathbf{x})$ and $\mathcal{I}_1(\mathbf{x})$ the collection of indices of the 0's and 1's in $\mathbf{x}$, respectively. For example, $\mathcal{I}_0(101) = 2$, $\mathcal{I}_1(101) = \{1, 3\}$. Let $N_1 = |\mathcal{I}_1(\mathbf{x})|$, a constant for each $\mathbf{x} \in S$, and $D(\cdot||\cdot)$ be the Kullback-Leibler distance.

Theorem 2 *On a binary input channel with $p(y|1)/p(y|0) < \infty$,*

$$C_S = N_1 D(p(y|1)||p(y|0)) - D(p(\mathbf{y})||p(\mathbf{y}|\mathbf{0})).$$

Proof: With equiprobable inputs from Theorem 1 we have

$$\begin{aligned}
C_S &= \int p(\mathbf{y}|\mathbf{x}_1) \log\left(\frac{p(\mathbf{y}|\mathbf{x}_1)}{p(\mathbf{y})}\right) d\mathbf{y} \\
&= \int p(\mathbf{y}|\mathbf{x}_1) \log\left(\frac{\prod_{i\in\mathcal{I}_1(\mathbf{x}_1)} p(y_i|1) \prod_{i\in\mathcal{I}_0(\mathbf{x}_1)} p(y_i|0)}{p(\mathbf{y})}\right) d\mathbf{y} \\
&= \int p(\mathbf{y}|\mathbf{x}_1) \log\left(\frac{p(\mathbf{y}|\mathbf{0}) \prod_{i\in\mathcal{I}_1(\mathbf{x}_1)} \frac{p(y_i|1)}{p(y_i|0)}}{p(\mathbf{y})}\right) d\mathbf{y} \\
&= \int p(\mathbf{y}|\mathbf{x}_1) \log\left(\prod_{i\in\mathcal{I}_1(\mathbf{x}_1)} \frac{p(y_i|1)}{p(y_i|0)}\right) d\mathbf{y} - \int p(\mathbf{y}) \log\left(\frac{p(\mathbf{y})}{p(\mathbf{y}|\mathbf{0})}\right) d\mathbf{y} \\
&= N_1 D(p(y|1)||p(y|0)) - D(p(\mathbf{y})||p(\mathbf{y}|\mathbf{0})).
\end{aligned}$$

□

2.2 Pulse position modulation

In the remainder we investigate the behavior of C_S as a function of n for the PPM channel. To that end, let $I(n)$ be the capacity of a memoryless channel with PPM inputs of length n. We first treat the case in Theorem 2 where $p(y|1)/p(y|0)$ is unbounded.

Let U and A be the collections of unambiguous and ambiguous outputs when $x = 1$ is transmitted,

$$\begin{aligned}
U &= \{y|p(y|1) \neq 0, p(y|0) = 0\} \\
A &= \{y|p(y|1) \geq 0, p(y|0) \neq 0\}.
\end{aligned}$$

If any coordinate of $\mathbf{y}$ belongs to U, the input will be known with certainty. In order to treat the ambiguous and unambiguous outputs separately, define a *reduced channel* $p^*(y|x)$ with output $y \in A$ as follows

$$\begin{aligned}
p^*(y|0) &= p(y|0) \\
p^*(y|1) &= \frac{p(y|1)}{p(A|1)},
\end{aligned}$$

where $p(A|1) = \int_A p(y|1)dy$. Let $I^*(\mathbf{X};\mathbf{Y})$ and $I^*(n)$ be the mutual information and capacity of the reduced channel.

Lemma 1

$$I(n) = p(U|1)\log n + p(A|1)I^*(n)$$

Proof: Let $U_n(\mathbf{x}_j)$ and $A_n(\mathbf{x}_j)$ be the collections of unambiguous and ambiguous outputs when

$\mathbf{x}_j$ is transmitted,

$$U_n(\mathbf{x}_j) = \{\mathbf{y}|p(\mathbf{y}|\mathbf{x}_j) \neq 0, p(\mathbf{y}|\mathbf{x}_i) = 0, \forall i \neq j\}$$
$$A_n(\mathbf{x}_j) = \{\mathbf{y}|p(\mathbf{y}|\mathbf{x}_j) \neq 0\} \backslash U(\mathbf{x}_j)$$

Let $\tilde{U} = \bigcup_{x \in S} U_n(\mathbf{x})$, $\tilde{A} = \bigcup_{x \in S} A_n(\mathbf{x})$. Introduce a binary random variable Z as follows:

$$Z = \begin{cases} 0, & \mathbf{Y} \in \tilde{U} \\ 1, & \mathbf{Y} \in \tilde{A} \end{cases}$$

Since each $\mathbf{x}$ contains exactly one nonzero entry, $P(Z = 0) = p(U|1)$ and $P(Z = 1) = p(A|1)$. Since the channel is memoryless and S is transitive, $p(U_n(\mathbf{x}_j)|\mathbf{x}_j)$ and $p(A_n(\mathbf{x}_j)|\mathbf{x}_j)$ do not depend on j. Therefore $H(\mathbf{X}|Z) = H(\mathbf{X})$ and

$$\begin{aligned} I(\mathbf{X};\mathbf{Y}) &= I(\mathbf{X};\mathbf{Y}|Z) \\ &= I(\mathbf{X};Z) + I(\mathbf{X};\mathbf{Y}|Z) \\ &= P(Z=0)I(\mathbf{X};\mathbf{Y}|Z=0) + P(Z=1)I(\mathbf{X};\mathbf{Y}|Z=1) \\ &= p(U|1)\log s + p(A|1)I^*(\mathbf{X};\mathbf{Y}). \end{aligned}$$

The lemma follows since the capacity achieving input distribution is uniform for both channels. □

Hence we can decompose the n-ary PPM channel into an unambiguous channel, which contributes $p(U|1)\log s$ to the capacity, and a reduced channel, with transition probabilities $p^*(y|x)$. In the remainder, we assume the channel is reduced, which allows a simple corollary of Theorem 2

Corollary 1 *For the reduced binary PPM channel,* $I(n) = D(p(y|1)||p(y|0)) - D(p(\mathbf{y})||p(\mathbf{y}|\mathbf{0}))$.

Corollary 1 allows a straightforward proof of the asymptotic behavior of the memoryless PPM channel.

Theorem 3 $\lim_{n\to\infty} I(n) = D(p(y|1)||p(y|0))$.

Proof: Let $\mathbf{1}$ denote the unit vector with a 1 in the first position.

$$\begin{aligned} 0 &\leq D(p(\mathbf{y})||p(\mathbf{y}|\mathbf{0})) \\ &= E_{\mathbf{Y}|\mathbf{X}=\mathbf{1}}\left[\log \frac{1}{n}\sum_{i=1}^{n}\frac{p(Y_i|1)}{p(Y_i|0)}\right] \\ &\leq \log \frac{1}{n}\sum_{i=1}^{n} E_{Y_i|\mathbf{X}=\mathbf{1}}\left[\frac{p(Y_i|1)}{p(Y_i|0)}\right] \qquad \text{(Jensen's inequality)} \\ &\leq \log\left(\frac{1}{n}E_{Y_1|X_1=1}\left[\frac{p(Y_1|1)}{p(Y_1|0)}\right] + \frac{n-1}{n}\right) \end{aligned}$$

where the last inequality follows since

$$E_{Y_i|X_i=0}\left[\frac{p(Y_i|1)}{p(Y_i|0)}\right] = \int_{y:p(y|0)>0} p(y|1)\ dy \leq 1$$

for $i \neq 1$. For a reduced channel there exists a constant K such that $E_{Y_1|X_1=1}\left[\frac{p(Y_i|1)}{p(Y_i|0)}\right] < K$, hence

$$0 \leq \lim_{n\to\infty} D(q^n||q_0^n) \leq \lim_{n\to\infty} \log\left(\frac{K}{n} + \frac{n-1}{n}\right) = 0$$

□

3 Capacity Inequalities

Theorem 4 *If $n \leq m$ then $I(kn) - I(n) \geq I(km) - I(m)$.*

This theorem says that if we multiply the number of slots by k, the increase in PPM capacity will be larger if the original number of slots was smaller. The conclusion can be equivalently stated as $I(kn) + I(m) \geq I(km) + I(n)$.

Proof: Let Z_k, Z_m, and Z_n be random variables uniformly distributed on $\{1, \ldots, k\}$, $\{1, \ldots, m\}$, and $\{1, \ldots, n\}$, respectively. Let $\mathbf{Y}_n$ be the (random) output vector when Z_n drives an n-PPM channel (that is, the input to the channel is an n-vector with a 1 in position Z_n and zeros elsewhere). Similarly, let $\mathbf{Y}_{kn}$ and $\mathbf{Y}_{km}$ be the output vectors when the ordered pairs (Z_k, Z_n) and (Z_k, Z_m) drive kn-PPM and km-PPM channels, respectively. In these two cases, it is useful conceptually to regard the slots as being arranged in a rectangular grid, with, for example, (Z_k, Z_n) specifying the column and row of the 1.

Observe that $H(Z_n|Z_k, \mathbf{Y}_{kn}) = H(Z_n|\mathbf{Y}_n)$, because in the left-hand side Z_k specifies the column of n slots in which the 1 is "hiding"; thus the other slots can be ignored. We therefore have

$$\begin{aligned} H(Z_k, Z_n|\mathbf{Y}_{kn}) &= H(Z_n|Z_k, \mathbf{Y}_{kn}) + H(Z_k|\mathbf{Y}_{kn}) \\ &= H(Z_n|\mathbf{Y}_n) + H(Z_k|\mathbf{Y}_{kn}). \end{aligned}$$

We then have

$$\begin{aligned} I(kn) - I(n) &= \log_2 kn - H(Z_k, Z_n|\mathbf{Y}_{kn}) - \log_2 n + H(Z_n|\mathbf{Y}_n) \\ &= \log_2 k - H(Z_k|\mathbf{Y}_{kn}). \end{aligned}$$

Similarly, $I(km) - I(m) = \log_2 k - H(Z_k|\mathbf{Y}_{km})$.

All that remains is to show that $H(Z_k|\mathbf{Y}_{km}) \geq H(Z_k|\mathbf{Y}_{kn})$. Intuitively, this is clear because there is more ambiguity about which column the 1 is in when there are more rows. A more formal line of reasoning would involve introducing side information W in the km-PPM case that specifies a random set of n rows, one of which contains the 1. Then $H(Z_k|\mathbf{Y}_{km}) \geq H(Z_k|\mathbf{Y}_{km}, W) = H(Z_k|\mathbf{Y}_{kn})$. □

Theorem 4 has the immediate consequence that if $m_1 m_2 = n_1 n_2$ and $m_1 + m_2 \leq n_1 + n_2$, then $I(m_1) + I(m_2) \geq I(n_1) + I(n_2)$. A special case of the theorem is that $I(mn) \leq I(m) + I(n)$.

We are also now able to say something interesting about 2^k-PPM:

Corollary 2 *For $k = 1, 2, \ldots$, the quantity $I(2^k)/k$ is decreasing in k.*

Proof: Theorem 4 implies that $I(2) - I(1), I(4) - I(2), I(8) - I(4), \ldots$ is a decreasing sequence. Therefore the average of the first k terms of the sequence is decreasing in k. But since $I(1) = 0$, the average of the first k terms is simply $I(2^k)/k$. □

Corollary 3 *For $k \in \mathbb{N}$, the bits-per-slot capacity of 2^k-PPM on a discrete-time memoryless channel is monotonically decreasing in k.*

Proof: The capacity of 2^k-PPM in bits per slot is $I(2^k)/2^k$. Thus this result follows from (and is much weaker than) Corollary 2. □

A close look at Theorem 4 suggests that the following is likely true: The quantity $(I(k+1) - I(k))/(\log(k+1) - \log k)$ is decreasing in k. Equivalently, the function $I(k)$ is convex-$\cap$ when plotted as a function of $\log k$. As of this writing, we have not proven this, so it is still a conjecture. Theorem 4 would essentially be a special case of this result. The result would also imply more general versions of Corollaries 2 and 3: the quantity $I(k)/\log k$ would be decreasing in k for $k \geq 2$, and $I(k)/k$ would be decreasing in k for $k \geq 3$.

References

[1] R. G. Lipes, "Pulse-position-modulation coding as near-optimum utilization of photon counting channel with bandwidth and power constraints," *DSN Progress Report*, vol. 42, pp. 108–113, Apr. 1980.

[2] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel–Part I," *IEEE Transactions on Information Theory*, vol. 34, pp. 1449–1461, Nov. 1988.

[3] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, 1968.

[4] J. A. Gallian, *Contempory Abstract Algebra*. D.C. Heath and Company, 3 ed., 1994.