On the Uniqueness of the Posterior Matching Feedback Communication Scheme

Amichai Painsky and Saharon Rosset
School of Mathematical Sciences, Tel Aviv University
Tel Aviv, Israel
amichaip@eng.tau.ac.il, saharon@post.tau.ac.il

Meir Feder
School of Electrical Engineering, Tel Aviv University
Tel Aviv, Israel
meir@eng.tau.ac.il

Abstract— The Posterior Matching Scheme provides an optimal feedback communication method. It introduces a fundamental principle whereby the information still missing at the receiver is extracted from the a-posteriori density function, and matched to any desirable input distribution. Posterior Matching achieves the channel capacity for any memoryless channel in the presence of noiseless feedback, and can be viewed as a generalization of the well-known Schalkwijk-Kailath scheme for the AWGN with average power constraint and the Horstein scheme for the BSC channel. The importance of the Posterior Matching principle goes beyond optimal feedback communication as it deals with a broader family of density matching problems. In this paper we discuss the uniqueness of the Posterior Matching principle and show that this scheme is unique for the set of monotonically increasing functions. Moreover, we show that any non-increasing function that satisfies the conditions of the Posterior Matching scheme is necessarily a "permutation" of the suggested scheme.

Keywords – Posterior Matching, feedback communication, memoryless encoding.

I. INTRODUCTION

The challenge of optimal communication in the presence of noiseless feedback has been studied quite broadly throughout the years. Early on Horstein [1] presented a feedback communication approach for a Binary Symmetric Channel (BSC). In his scheme, a message point inside the unit interval is used to represent the data bits, and is conveyed to the receiver by always indicating whether it lies to the left or to the right of the receiver's posterior median, which is also known to the transmitter via feedback. That way, the transmitter always answers the most informative question that can be posed by the receiver based on the information the latter has. This technique proved to attain the capacity of the BSC channel, and can be easily adapted to any Discrete Memoryless Channel (DMC) with feedback. The Schalkwijk-Kailath method [2-5] suggests an elegant capacity achieving feedback scheme for the Additive White Gaussian Noise (AWGN) channel with an average power constraint. At each time point the suggested method finds the Minimum Mean Square Error (MMSE) estimate of the message point at the receiver, and transmits the MMSE error on the next channel use, amplified to match the permissible input power constraint. This Scheme too, was shown to attain the channel capacity. In their Posterior Matching scheme [6-7] Shayevitz and Feder argue that there exists an underlying principle connecting these two methods. They claim that both these methods strive to transmit what the receiver is missing. This principle is the basic premise of the Posterior Matching scheme and is formulized in quite a simple manner. Let $X \in [0,1)$ be a message point whose binary expression represents an infinite bitstream to be reliably conveyed to the receiver. Denote the transmitted signal at time n as V_n, sent through a noisy memoryless channel $F_{Y|V_n}$. A random variable Y is therefore received by the receiver and is also evident to the transmitter through the noiseless feedback channel. The Posterior Matching principle suggests that the re-transmitted message V_{n+1} should hold three conditions:

There exists a random variable U such that,

(I) U is statistically independent of $Y^n = \{Y_i\}_{i=1}^n$.

(II) X can be uniquely recovered from U given Y^n.

(III) V_{n+1} is a deterministic function of U satisfying $V_{n+1} \sim F_V(v)$.

The first condition guarantees that V_{n+1} represents "new information" not yet observed by the receiver while the second condition makes sure the information is "relevant" in terms of describing the message. The third condition yields the re-transmitted message V_{n+1} which is distributed according to a known commutative distribution function F_V, matched to the noisy channel (hence the name Posterior Matching). In their paper, Shayevitz and Feder show there exist a method satisfying the above:

$$V_{n+1} = F_V^{-1}\left(F_{X|Y^n}(x|y^n)\right) \tag{1}$$

where $F_{X|Y^n}(x|y^n)$ is the posterior of $X|Y^n$ and F_V^{-1} is the inverse of F_V. It is easy to show that applying the function $F_{X|Y^n}(x|y^n)$ on the message X given the value of Y^n yields a uniformly distributed random variable. Moreover, this random variable is in fact independent of Y^n as for any realization of Y^n, it is distributed in a similar manner. In addition, since a uniform distribution has a monotonically increasing CDF, the message X can be uniquely recovered from V_{n+1} given the realization of Y^n. Finally, applying the inverse of F_V on this uniformly distributed random variable

guarantees V_{n+1} is distributed according to required distribution. Since F_V^{-1} is also a deterministic function of $F_{X|Y^n}(x|y^n)$, the Posterior Matching principle is satisfied Shayevitz and Feder also proved this scheme achieves the channel capacity and provided closed form expressions for the error probably of it. They showed both Horstein and Schalkwijk-Kailath methods can be derived as special cases of the Posterior Matching scheme. In this paper we discuss the uniqueness of the Posterior Matching scheme. I.e. we argue whether or not there are other functions that satisfy the conditions stated above. The importance of this question goes beyond the context of optimal feedback communication as it deals with quite a fundamental setup of probability functions matching. In other words, given a set of distribution functions, is there a unique method of matching them to the same distribution function? If not, is it possible to describe the family of functions that satisfy it? In this paper we provide a comprehensive discussion over these questions and characterize a set of solutions for the Posterior Matching conditions. We show that the scheme proposed by Shayevitz and Feder is unique under a natural monotonicity condition on the relation between *X* (given Y^n) and *U*, and explicitly derive its connection to all possible non-monotone transformations.

II. Problem Formulation

The essence of the Posterior Matching scheme is presented by the following problem formulation. Assume a message point *X* is to be transmitted through a noisy memoryless channel. A random variable *Y* is received by the receiver (representing Y^n as it was previously denoted to reduced notation) and is also evident to the transmitter through the noiseless feedback channel. From these two random variables we would like to construct a re-transmitted random variable, $V = g(X,Y)$ with a known $F_V(v)$ such that there exists a random variable *U* which satisfy:

(I) *U* is statistically independent in *Y*.
(II) *X* can be uniquely recovered from *U* given *Y*.
(III) *V* is a deterministic function of *U* satisfying $V \sim F_V(v)$.

Our goal is therefore finding such $V = g(X,Y)$ and discussing its uniqueness.

III. The Uniform Distribution Case

In this section we consider a special case where *V*=*U* and is uniformally distributed

$$F_V(v) = F_U(u) = u\,. \tag{2}$$

For *U* to be statistically independent of Y it must satisfy

$$F_{U|Y}(u|Y=y) = F_U(u). \tag{3}$$

Deriving the left hand side we have

$$\begin{aligned} F_{U|Y}(u|Y=y) &= Prob(U \le u|Y=y) \\ &= Prob(g(X,Y) \le u|Y=y). \end{aligned} \tag{4}$$

A. Uniqueness of Monotonically Increasing Transformations

The second constraint suggests *X* can be uniquely recovered from *U* and *Y*, which implies $X = g_Y^{-1}(U)$. Assume $g(X,Y)$ is monotonically increasing with respect to *X*. In this case we have

$$\begin{aligned} F_{U|Y}(u|Y=y) &= Prob(g(X,Y) \le u|Y=y) \\ &= Prob(X \le g_Y^{-1}(u)|Y=y) \\ &= F_{X|Y}(g_Y^{-1}(u)|Y=y) \end{aligned} \tag{5}$$

where the second equality follows from the monotonically increasing behavior of $g(X,Y)$ with respect to *X*.
Therefore, we are looking for a monotonically increasing transformation $x = g_Y^{-1}(u)$ such that

$$F_{X|Y}(g_Y^{-1}(u)|Y=y) = F_U(u) = u. \tag{6}$$

Lemma 1: Assume *X* is a random variable with a CDF $F_X(x)$. Assume there exists a transformation on its domain, $x = h(u)$ such that

$$F_X(x)|_{x=h(u)} = F_U(u). \tag{7}$$

In addition assume $F_X(x)$ is monotonically increasing where the transformation is defined, then

(i) $x = h(u)$ is unique.
(ii) $h(u)$ is monotonically non decreasing (increasing, if $F_U(u)$ is strictly increasing).

Proof:

(i) The transformation $x = h(u)$ satisfies

$$F_X(x)|_{x=h(u)} = F_X\big(h(u)\big) = prob\big(X \le h(u)\big) = F_U(u). \tag{8}$$

Suppose there is another transformation $x = g(u)$ that satisfies the conditions stated above then

$$F_X(x)|_{x=g(u)} = F_X\big(g(u)\big) = prob(X \le g(u)) = F_U(u). \tag{9}$$

Therefore,

$$prob\big(X \le g(u)\big) = prob(X \le h(u))\ \forall u. \tag{10}$$

Suppose $h(u) \neq g(u)$. This means there exist at least a single $u = \tilde{u}$ where $g(\tilde{u}) = h(\tilde{u}) + \delta$ and $\delta \neq 0$. It follows that

$$prob(X \le h(\tilde{u}) + \delta\ ) = prob(X \le \tilde{h}(u)) \qquad (11)$$

or in other words

$$F_X\big(h(\tilde{u})\big) = F_X(h(\tilde{u}) + \delta) \qquad (12)$$

which contradicts the monotonically increasing behavior of $F_X(x)$ where the transformation is defined.

(ii) We have $F_X\big(h(u)\big) = F_U(u)\ \forall u.$

Therefore

$$F_X\big(h(u+\delta)\big) = F_U(u+\delta). \qquad (13)$$

$F_U(u)$ is a CDF which satisfies $F_U(u+\delta) \ge F_U(u)$, therefore

$$F_X\big(h(u+\delta)\big) \ge F_X\big(h(u)\big) \qquad (14)$$

(strictly larger if $F_U(u)$ is monotonically increasing).
Since $F_X(x)$ is monotonically increasing we have

$$h(u+\delta) \ge\ h(u) \qquad (15)$$

(strictly larger if $F_U(u)$ is monotonically increasing) ■

Lemma 2: Assume *X* is a *discrete* random variable with a CDF $F_X(x)$. Suppose there exists a transformation on its domain $x = h(u)$ such that

$$F_X(x)|_{x=h(u)} = F_U(u) \qquad (16)$$

then we have:

(i) $x = h(u)$ is unique up to transformations in zero probability areas of the *X*.

(ii) $h(u)$ is monotonically non decreasing (increasing, if $F_U(u)$ is strictly increasing).

Proof:

(i) As in Lemma 1, let's assume there is another transformation $x = g(u)$ that satisfies the requested conditions. Therefore we have

$$prob\big(X \le g(u)\big) = prob(X \le h(u))\ \forall u. \qquad (17)$$

Assuming $h(u) \ne g(u)$ we have at least a single $u = \tilde{u}$ where $g(\tilde{u}) = h(\tilde{u}) + \delta$ and $\delta \ne 0$.
If both $h(\tilde{u})$ and $g(\tilde{u})$ are valid values in *X*'s alphabet (positive probability) then we have

$$prob(X \le x_1) = prob(X \le x_2). \qquad (18)$$

This contradicts $prob(X = x_1) > 0$ and $prob(X = x_2) > 0$ unless $x_1 = x_2$. Moreover, if $g(\tilde{u}) \in [x_1, x_2]$ and $h(\tilde{u}) \notin [x_1, x_2]$ then again it contradicts $prob(X = x_1) > 0$ and $prob(X = x_2) > 0$ unless $x_1 = x_2$.
The only situation in which we are not facing a contradiction is where $g(\tilde{u}), h(\tilde{u}) \in [x_1, x_2]$.
In other words, $x = g(u)$ is unique up to transformations in zero probability areas of the random variable *X* (areas which satisfy $prob\big(X = g(\tilde{u})\big) = 0$)

(ii) The monotonicity proof follows the same derivation as in Lemma 1 ■

To conclude, assuming there exists a transformation $x = g_Y^{-1}(u)$ such that

$$F_{X|Y}(g_Y^{-1}(u)|Y = y) = F_U(u) = u \qquad (19)$$

then it is unique and monotonically increasing. In this case we have

$$\begin{aligned} F_U(u) = F_{X|Y}(g_Y^{-1}(u)|Y = y) = \qquad (20)\\ Prob(X \le g_Y^{-1}(u)|Y = y) = \\ Prob(g(X,Y) \le u|Y = y) = F_{U|Y}(u|Y = y) \end{aligned}$$

which means *U* is statistically independent of *Y*.
Equivalently, if we find a monotonically increasing transformation *U=g(X,Y)* that satisfies conditions (I), (II) and (III) then it is unique.

B. Non Monotonically Increasing Transformations

In the previous section we discussed the case in which we limit ourselves to functions *g(X,Y)* which are monotone in their first argument. For this set of functions equation (5) is a sufficient condition for satisfying (I) and (II).
We may find, however, non monotonically increasing transformations *U=h(X,Y)* which satisfy conditions (I), (II) and (III) but do not satisfy equation (5). For example: *h(X,Y)=1-g(X,Y)*. Notice these transformations are necessarily measurable, as they map one distribution to another, and reversible with respect to X given Y (condition II). In this case, the following properties hold:

Lemma 3: Assume h(X,Y) satisfies the three conditions mentioned above but does not satisfy equation (5). Then:

(i) h(X, Y) is not monotonically increasing in X.

(ii) h(X, Y) is necessarily "reordering" of *g(X,Y)*.

Proof:

(i) Assume there exists a transformation *U=h(X,Y)* which satisfy the three conditions (I), (II) and (III). Moreover assume *h(X,Y)≠g(X,Y).*
We know that

$$F_{U|Y}(u|Y = y) = Prob(h(X,Y) \le u|Y = y) = F_U(u) \qquad (21)$$

but on the other hand, *h(X,Y)≠g(X,Y)* which implies

$$F_{X|Y}(h_Y^{-1}(u)|Y=y) \neq F_U(u) \quad (22)$$

since *g(X,Y)* is unique. Therefore,

$$Prob(h(X,Y) \leq u|Y=y) \neq Prob(X \leq h_Y^{-1}(u)|Y=y) \quad (23)$$

which means $h(X,Y)$ cannot be monotonically increasing.

(ii) We can always create a (reversible) transformation on *h(X,Y)* that will make it monotonically increasing with respect to *X*, since *X* is uniquely recoverable from *h(X,Y)* and *Y*. Consider this transformation as $S(h(X,Y))$. Therefore, we found $U = S(h(X,Y))$ such that u is monotonically increasing, independent of *Y* and *X* is uniquely recoverable from *U* and *Y*. This contradicts the uniqueness of *g(X,Y)* unless $S(h(X,Y))$=*g(X,Y)*, which means

$$\mathrm{h(X,Y)} = \mathrm{S^{-1}(g(X,Y))} \quad (24)$$

■

C. Existance of a Monotonically Increasing Transformation

Following the properties we showed in the previous sections, it is enough to find *U=g(X,Y)* which is invertible and monotonically increasing with respect to *X* given *Y=y*, and satisfies

$$F_{U|Y}(u|Y=y) = F_{X|Y}(g_Y^{-1}(u)|Y=y) = F_U(u) = u. \quad (25)$$

If such *U=g(X,Y)* exists then

(i) If $F_{X|Y}(x|Y=y)$ is monotonically increasing, then *U=g(X,Y)* is unique according to Lemma1.

(ii) If *X|Y* takes on discrete values, then again *U=g(X,Y)* is unique, up to different transformations in zero probability areas of the *X|Y*.

(iii) Any other *h(X,Y)* that may satisfy conditions (I),(II) and (III) is necessarily a function of *g(X,Y)* (and not monotonically increasing).

Lemma 4: Let $X \sim P_X(x)$ and $\Theta \sim \mathcal{U}nif[0,1]$ be statistically independent. Then

(i) $F_X^{-1}(\Theta) \sim P_X(x)$

(ii) $F_X(x) - \Theta \cdot P_X(x) \sim \mathcal{U}nif[0,1]$. Specifically, if *X* is proper ($F_X(x)$ is strictly continuous) then $F_X(x) \sim \mathcal{U}nif[0,1]$

Proof : can be located in [7].

Define $U = F_{X|Y}(x|y) - \Theta \cdot P_{X|Y}(x|y)$, where $\Theta \sim \mathcal{U}nif[0,1]$ is statistically independent of *X* and *Y*. Then we get

$$F_{U|Y}(u|y) = \quad (26)$$
$$Prob\big(F_{X|Y}(x|y) - \Theta \cdot P_{X|Y}(x|y) \leq u \mid Y=y\big) \underbrace{=}_{1}$$
$$Prob\left(F_{X|Y}(x|y) - \Theta \cdot P_{X|Y}(x|y) \leq h^{-1}(u)\right) \underbrace{=}_{2}$$
$$u \underbrace{=}_{3} F_U(u).$$

Where:

(1) all the terms in $F_{X|Y}(x|y) - \Theta \cdot P_{X|Y}(x|y) \leq h^{-1}(u)$ are already conditioned in y, or statistically independent of *y*.

(2) $F_{X|Y}(x|y) - \Theta \cdot P_{X|Y}(x|y) \sim \mathcal{U}nif[0,1]$, according to lemma 1.

(3) The condition we want to satisfy

Also, it is easy to see that $U = F_{X|Y}(x|y) - \Theta \cdot P_{X|Y}(x|y)$ is reversible with respect to *X* given *Y=y*.

Therefore, we found a monotonically increasing transformation *U=g(X,Y)* that satisfies

$$F_{U|Y}(u|Y=y) = F_{X|Y}(g_Y^{-1}(u)|Y=y) = F_U(u) = u. \quad (27)$$

II. The Non-Uniform Case

Going back to our original task, we are interested in finding such *V=g(X,Y)* that there exists a random variable U which satisfy:

(I) *U* is statistically independent in *Y*.

(II) *X* can be uniquely recovered from *U* given *Y*.

(III) *V* s a deterministic function of *U* satisfying $V \sim F_V(v)$.

Throughout the previous sections we discussed the uniqueness of the case in which *V=U* is uniformly distributed. Assume we are now interested in a non-uniformly distributed *V*. We know that if we set $V = F_V^{-1}(U)$ all three conditions are satisfied according to [6]. Therefore, we still need to discuss the uniqueness of this (deterministic) mapping from *U* to *V*. In other words, given a uniformly distributed *U* and a desired random variable $V \sim F_V(v)$, is the mapping $V = F_V^{-1}(U)$ unique? This question was already answered by Lemma 1; if we limit ourselves to monotonically increasing transformation, the solution we found is unique. However, assume we do not limit ourselves to monotonically increasing transformations, and assume we have a transformation $V = G(U)$ that satisfies $V \sim F_V(v)$. Since *U* is uniformly distributed we can always shift between local transformations on sets of the same lengths while maintaining the transformation measurable. Then we can always find $S(G(U))$ which makes it monotonically increasing with respect to *U*. This contradicts the uniqueness of the monotonically increasing set unless $S(G(U))$ equals the single unique transformation we found.

Putting this all together we have a two stage process in which we first generate a random variable U and then shape it to a desired distribution V through a deterministic mapping. We show that in both stages, if we limit ourselves to monotonically increasing transformations the solution presented in [6] is unique. However, if we allow a broader family of functions we necessarily end up with either the same solution, or a reordering of it which is not monotonically increasing.

SUMMARY

In this paper we discussed the uniqueness of the Posterior Matching scheme over a noiseless feedback channel. We presented the three conditions that define the Posterior Matching principle, and discussed the family of functions that satisfy them. We showed that if we limit ourselves to a family of monotonically increasing functions with respect to the message X, the solution presented in the original Posterior Matching scheme is unique. We also showed that even though there are other non-increasing functions that satisfy the three conditions, these are necessarily "permutations" of the original Posterior Matching solution. The uniqueness of the Posterior Matching scheme is essential for the discussion and comparison of this scheme with other optimal feedback methods that may be derivatives of this fundamental idea. It is also important in the context of a broader family of problems dealing with probability distributions matching, as in memoryless representation for Markov processes [8] for example.

ACKNOWLEDGMENT

This work was funded in part by Israeli Science Foundation grant 634-09 and by a grant to Amichai Painsky from the Israeli Center for Absorption in Science.

REFERENCES

[1] M. Horstein, "Sequential transmission using noiseless feedback," IEEE Trans. Info. Theory, pp. 136–143, 1963.

[2] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback part I: No bandwidth constraint," *IEEE Trans.Info. Theory*, vol. IT-12, pp. 172 – 182, 1966.

[3] J. P. M. Schalkwijk, "A class of simple and optimal strategies for block coding on the binary symmetric channel with noiseless feedback," IEEE Trans. Info. Theory, vol. 17, no. 3, pp. 283–287, 1971.

[4] J. P. M. Schalkwijk and K. A. Post, "On the error probability for a class of binary recursive feedback strategies," IEEE Trans. Info. Theory, vol. IT-19, pp. 498–511, 1973.

[5] A.D. Wyner, "On the schalkwijk-kailath coding scheme with a peak energy constraint," IEEE Trans. on Info. Theory, vol. IT-14, no. 1, pp. 129–134, 1968.

[6] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," IEEE Trans. Info. Theory. vol 57, no. 3, pp. 1186-1222, 2011.

[7] O. Shayevitz and M. Feder, "The posterior matching feedback scheme Capacity achieving and error analysis," in Proc. of the International Symposium on Information Theory, 2008.

[8] A. Painsky, S. Rosset and M.Feder, "Memoryless representation for markov processes", submitted for the Internation Symposium on Information Theory , 2013.