On Proving the Water Pouring Theorem for Information Rate Optimization

Mark F. Flanagan
Department of Electronic and Electrical Engineering,
University College Dublin,
Belfield, Dublin 4, Ireland,
e-mail: mark.flanagan@ee.ucd.ie

***Abstract*— This paper presents a novel proof of the famous water pouring theorem for choosing transmit power distribution for the optimization of information rate on a coloured additive Gaussian noise channel. The discrete case (optimization of information rate for multicarrier modulation) and the continuous case (determination of the capacity of a coloured noise channel) are treated separately. In contrast to previous proofs, the proof for the discrete case does not require calculus and for the continuous case requires only elementary calculus. Finally, in the discrete case, the algorithm for calculating the optimal transmit power distribution follows as a natural corollary.**

I. Introduction

Rigorous proofs of the water pouring theorem, such as the classic proof in [1], generally require use of the Karhunen-Loève expansion and Toeplitz distribution theorems. A simpler approach to proving the theorem, often used in textbooks, is to make the intuitively reasonable assumption that when the channel is divided in frequency, each subchannel is independent of the set of all others, even in the limit as the number of subchannels tends to infinity. Simpler proofs are then possible which usually rely on the method of Lagrange multipliers with the application of the Kuhn-Tucker conditions [2], [3], [4]. The proofs presented here shall make use of this assumption, but shall proceed using more elementary techniques. Also, the proofs are presented for the baseband channel only; extension to the bandpass channel follows from the same analysis with only slight modification.

II. Expression for the Information Rate of a Multitone System

According to Shannon's information capacity theorem [5] the capacity in bits/s of an AWGN channel of bandwidth B is given by

$$C = B\log_2\left(1 + SNR\right) \tag{1}$$

In practice, a physically realisable system (employing e.g. PAM or QAM) transmits data at a rate $R < C$, incorporating a modulation gap $\Gamma > 1$:

$$R = B\log_2\left(1 + \frac{SNR}{\Gamma}\right) \tag{2}$$

Consider a channel bandlimited to B Hz with frequency response function $H\left(f\right)$, two-sided transmit power spectral density $P\left(f\right)$ and two-sided additive noise power spectral density $S_N\left(f\right)$, $|f| \leq B$. We will assume that all three of these functions are continuous on $[0, B]$, and that $|H\left(f\right)|$ and $S_N\left(f\right)$ are positive on $[0, B]$. We divide the bandwidth $[0, B]$ into N subchannels $\left[f_k - \frac{\Delta f}{2}, f_k + \frac{\Delta f}{2}\right]$ of bandwidth $\Delta f = \frac{B}{N}$, where the center frequencies are $f_k = \frac{\Delta f}{2} + k\Delta f$, $k = 0, 1, \ldots N-1$. We then make the approximation that each subchannel is a narrowband AWGN channel. This approximation is good for large N. The information rate for subchannel k is then

$$R_k = \Delta f \log_2\left(1 + \frac{\left|H\left(f_k\right)\right|^2 P_k}{\sigma_k^2 \Gamma_k}\right) \tag{3}$$

where $P_k = 2P\left(f_k\right)\Delta f$ is the transmit signal power allocated to subchannel k, $\sigma_k^2 = 2S_N\left(f_k\right)\Delta f$ is the subchannel noise variance and $\Gamma_k \geq 1$ is the subchannel modulation gap. In addition we shall assume a power constraint

$$\sum_{k=0}^{N-1} P_k = P \tag{4}$$

The gain to noise ratio for subchannel k is defined as

$$\text{GNR}_k = \frac{1}{\alpha_k} = \frac{\left|H\left(f_k\right)\right|^2}{\sigma_k^2 \Gamma_k} \tag{5}$$

so we can write the information rate for subchannel k as

$$R_k = \Delta f \log_2\left(1 + \frac{P_k}{\alpha_k}\right) \tag{6}$$

We shall assume that the N subchannels are independent of one another. The overall information rate for the DMT system is then

$$R = \sum_{k=0}^{N-1} R_k = \Delta f \log_2\left\{\prod_{k=0}^{N-1}\left(\frac{1}{\alpha_k}\right) \cdot \prod_{k=0}^{N-1} x_k\right\} \tag{7}$$

where we define $x_k = P_k + \alpha_k$ for each k.

III. Optimal power loading of subchannels for DMT

Theorem 1 (Water Pouring Theorem - Discrete Case): Given the power constraint (4), necessary and sufficient conditions for the maximization of information rate are

$$P_k > 0 \Rightarrow x_k \leq x_m \quad \forall m \in \{0, 1, \dots N-1\} \tag{8}$$

(As an illustration of this theorem, the optimal transmit power distribution is shown in Fig. 1 for the case of $N = 5$ subchannels.)

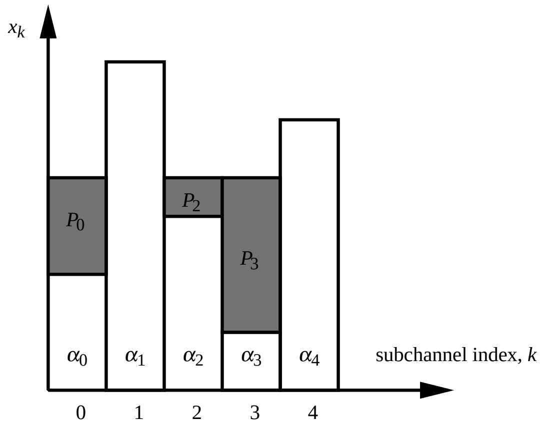

Fig. 1. Illustration of the water pouring theorem for DMT. The diagram shows the optimal transmit distribution for the case of $N = 5$ subchannels. Here only P_0, P_2 and P_3 are positive.

Proof: The problem is easily seen to reduce to that of maximizing the function

$$Q((x_0, x_1, \dots x_{N-1})) = \prod_{k=0}^{N-1} x_k$$

over the set

$$S = \left\{ (x_0, x_1, \dots x_{N-1}) \,| P_k = x_k - \alpha_k \geq 0, \sum_{k=0}^{N-1} P_k = P \right\}$$

where $\{\alpha_k \mid k = 0, 1, \dots N-1\}$ is a set of fixed positive real numbers.

Necessity of the Conditions:

The domain S is a finite-dimensional, closed and bounded set, and Q is continuous on S; therefore a maximum exists. To establish necessity we shall use proof by contradiction. Assume the configuration $(x_0, x_1, \dots x_{N-1}) \in S$ is optimal. Suppose $x_k > x_m$ for some k with $x_k > \alpha_k$ and for some m. Then, for any δ with $0 < \delta < \text{Min}\{x_k - \alpha_k, x_k - x_m\}$, we can perform the mapping

$$\begin{aligned} x_k &\rightarrow x'_k = x_k - \delta \\ x_m &\rightarrow x'_m = x_m + \delta \\ x_i &\rightarrow x'_i = x_i \qquad\qquad i \neq k, m \end{aligned}$$

We can immediately see that $x'_i \geq \alpha_i$ for $i \in \{0, 1, \dots N-1\} \setminus \{k\}$, and $x'_k \geq \alpha_k$ because of our choice of δ. The configuration $\left(x'_0, x'_1, \dots x'_{N-1}\right)$ also satisfies $\sum_{i=0}^{N-1} \{x'_i - \alpha_i\} = P$. Now

$$\begin{aligned} &Q(\left(x'_0, x'_1, \dots x'_{N-1}\right)) - Q((x_0, x_1, \dots x_{N-1})) \\ &= \left(\prod_{i \neq k,m} x_i \right) \left[(x_k - \delta)(x_m + \delta) - x_k x_m \right] \\ &= \left(\prod_{i \neq k,m} x_i \right) \left[\delta (x_k - x_m - \delta) \right] \\ &> 0 \end{aligned}$$

This contradicts the assumption that the configuration is optimal, and thus proves necessity of the conditions.

Sufficiency of the conditions:

Let $P_k > 0$ and $P_m > 0$ for some k and m. Then by necessary conditions $x_k \leq x_m$ and $x_m \leq x_k$, and thus $x_k = x_m$. Therefore there exists a constant L such that

$$x_k = L \quad \forall\, k \text{ with } P_k > 0 \tag{9}$$

Let $P_k > 0$ and let $\alpha_m < L$. Suppose $P_m = 0$. Then $x_m = \alpha_m < x_k$. But the necessary conditions imply $x_k \leq x_m$. This contradiction proves that

$$P_m > 0 \quad \forall\, m \text{ with } \alpha_m < L \tag{10}$$

Let $P_k > 0$ and let $\alpha_m \geq L$. Suppose $P_m > 0$. Then $x_m > \alpha_m \geq L$. But we should have, from (9), that $x_m = L$. This contradiction proves that

$$P_m = 0 \quad \forall\, m \text{ with } \alpha_m \geq L \tag{11}$$

Now, without loss of generality, we can assume that $\{\alpha_k\}$ forms a monotonically nondecreasing sequence. From (9), (10) and (11) we deduce that there exists $r \in \{0, 1, \dots N-1\}$ such that $P_k > 0$ for all $k \leq r$ and $P_k = 0$ for all $k > r$. Therefore, defining $\alpha_N = \infty$, for any $j \in \{0, 1, \dots r\}$ we have

$$\alpha_r < x_j \leq \alpha_{r+1} \tag{12}$$

Summing these inequalities for all $j \in \{0, 1, \dots r\}$, we get

$$(r+1)\,\alpha_r < \sum_{j=0}^{r} (P_j + \alpha_j) \leq (r+1)\,\alpha_{r+1} \tag{13}$$

Letting

$$T_r = \begin{cases} 0 & r = 0 \\ r\alpha_r - \sum_{j=0}^{r-1} \alpha_j & r \in \{1, \dots N-1\} \\ \infty & r = N \end{cases} \tag{14}$$

this can be written as

$$T_r < P \leq T_{r+1} \tag{15}$$

Since $\{T_j\}$, $j \in \{0, 1, \dots N\}$ is a monotonically nondecreasing sequence, the value of $r \in \{0, 1, \dots N-1\}$ which satisfies the previous equation is unique. Given this value of r, the solution for the $\{P_j\}$ is, by (9)

$$P_j = \begin{cases} L - \alpha_j & j \leq r \\ 0 & j > r \end{cases} \tag{16}$$

where (combining (4) and (16))

$$L = \frac{P + \sum_{j=0}^{r} \alpha_j}{(r+1)} \tag{17}$$

Thus the $\{P_j\}$ which satisfies the conditions is unique, and this proves sufficiency of the conditions. ∎

The algorithm for calculating the optimal power distribution follows as a corollary. Given $\{\alpha_k\}$, $k \in \{0, 1, \ldots N-1\}$ and P, the procedure is as follows: using (14), compute $\{T_j\}$, $j \in \{0, 1, \ldots N\}$ and find that unique $r \in \{0, 1, \ldots N-1\}$ which satisfies (15). Then compute L via (17), thus yielding the optimal power distribution (16).

It is not a difficult matter to extend this method of proof to obtaining analagous necessary and sufficient conditions for the continuous case; however this method of proof fails to establish the existence of a maximum, a problem which the current authors have as yet been unable to resolve.

IV. Expression for the Capacity of the Coloured Noise Channel subject to Fixed Transmit Power Spectral Density

The capacity of the coloured noise channel for a fixed transmit power spectral density $P(f)$ is obtained by taking the limit of the information rate (with $\Gamma_k = 1$ for all k) as the number of subchannels, N, tends to infinity. Define

$$GNR(f) = \frac{1}{\alpha(f)} = \frac{|H(f)|^2}{S_N(f)} \tag{18}$$

Then, assuming that the subchannels are independent in the limit as $N \to \infty$,

$$\begin{aligned} C &= \lim_{N\to\infty} \sum_{k=0}^{N-1} R_k \bigg|_{\Gamma_k = 1\, \forall\, k} \\ &= \lim_{N\to\infty} \sum_{k=0}^{N-1} \Delta f \log_2\left(1 + \frac{|H(f_k)|^2 P(f_k)}{S_N(f_k)}\right) \\ &= \int_0^B \log_2\left(1 + \frac{|H(f)|^2 P(f)}{S_N(f)}\right) df \\ &= \int_0^B \log_2\left(P(f) + \alpha(f)\right) df - \int_0^B \log_2 \alpha(f)\, df \end{aligned}$$

To determine the capacity of the colored noise channel, we must maximize this expression over the variable transmit power distribution $P(f)$, subject to the limiting form of the power constraint (4), i.e.

$$\lim_{N\to\infty} \sum_{k=0}^{N-1} 2P(f_k)\,\Delta f = 2\int_0^B P(f)\,df = P \tag{19}$$

V. Capacity of Coloured Noise Channel

Theorem 2 (Water Pouring Theorem - Continuous Case): Given the power constraint (19), the unique continuous transmit power distribution $P(f)$ for the maximization of the capacity of the coloured noise channel is:

$$P_0(f) = \begin{cases} L - \alpha(f) & f \in S_L \\ 0 & f \notin S_L \end{cases} \tag{20}$$

where, for any l,

$$S_l = \{f \in [0, B]\, |\alpha(f) < l\} \tag{21}$$

and L is such that

$$2\int_0^B P_0(f)\,df = P \tag{22}$$

(As an illustration of this theorem, the optimal transmit power distribution is shown in Fig. 2 for an example channel GNR.)

Proof: This problem is easily seen to reduce to that of maximizing the functional

$$Q(P(f)) = \int_0^B \ln\left(P(f) + \alpha(f)\right) df$$

over the set

$$\begin{aligned} S = \Big\{ & P(f) : [0, B] \to \mathbb{R} |\, P(f) \geq 0\, \forall f \in [0, B], \\ & \int_0^B P(f) df = P/2, P(f) \text{ continuous} \Big\} \end{aligned}$$

where $\alpha(f)$ is a fixed positive continuous function on $[0, B]$. To prove that the function $P_0(f)$ defined above is the unique $P(f) \in S$ which maximizes Q, we define for any candidate function $P(f) \neq P_0(f)$,

$$P_t(f) = P_0(f) + t\Delta P(f) \quad t \in [0, 1] \tag{23}$$

where

$$\Delta P(f) = P(f) - P_0(f) \tag{24}$$

Note that

$$\int_0^B \Delta P(f) df = \int_0^B P(f) df - \int_0^B P_0(f) df = 0 \tag{25}$$

since $P(f), P_0(f)$ are both in S. Let

$$\begin{aligned} G(t) &= Q(P_t(f)) \\ &= \int_0^B \ln\left(P_t(f) + \alpha(f)\right) df \end{aligned}$$

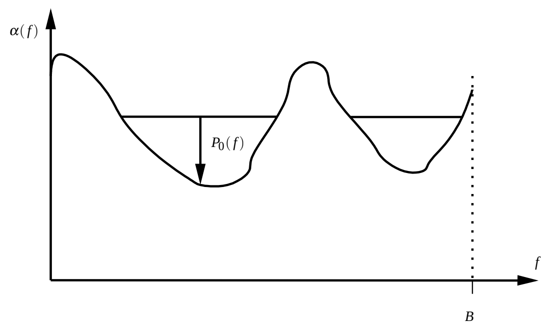

Fig. 2. Illustration of the water pouring theorem for capacity of the coloured Gaussian noise channel. The diagram shows the optimal transmit distribution $P_0(f)$. Here $P_0(f)$ is positive only over two subintervals of $[0, B]$.

This integral exists since $P_t(f) \geq 0$ for all $f \in [0, B]$. Differentiating,

$$G'(t) = \int_0^B \left(\frac{\Delta P(f)}{P_t(f) + \alpha(f)} \right) df \tag{26}$$

$$\begin{aligned} G'(0) &= \int_0^B \left(\frac{\Delta P(f)}{P_0(f) + \alpha(f)} \right) df \\ &= \int_{f \notin S_L} \left(\frac{\Delta P(f)}{\alpha(f)} \right) df + \int_{f \in S_L} \left(\frac{\Delta P(f)}{L} \right) df \\ &= \int_{f \notin S_L} \left(\frac{\Delta P(f)}{\alpha(f)} \right) df - \int_{f \notin S_L} \left(\frac{\Delta P(f)}{L} \right) df \\ &= \int_{f \notin S_L} \Delta P(f) \left(\frac{1}{\alpha(f)} - \frac{1}{L} \right) df \end{aligned}$$

Thus $G'(0) \leq 0$, since $\Delta P(f) = P(f)$ for all $f \notin S_L$. Also

$$G''(t) = -\int_0^B \left(\frac{\Delta P(f)}{P_t(f) + \alpha(f)} \right)^2 df < 0 \tag{27}$$

So $G'(0) \leq 0$ and $G''(t) < 0$ for all $t \in [0, 1]$. Therefore $G(0) > G(1)$ and so $P_0(f)$ is the global maximum. Now if we define, for any $l > \min\{\alpha(f)\}$,

$$T(l) = \int_{S_l} (l - \alpha(f))\, df \tag{28}$$

then $k > l$ implies

$$\begin{aligned} T(k) &= \int_{S_k} (k - \alpha(f))\, df \\ &\geq \int_{S_l} (k - \alpha(f))\, df \\ &> \int_{S_l} (l - \alpha(f))\, df \end{aligned}$$

and thus $T(k) > T(l)$. Therefore $T(x)$ is strictly monotonically increasing on $x > \min\{\alpha(f)\}$. Applying the power constraint (19) to $P_0(f)$ yields $T(L) = P/2$. This equation, coupled with the monotonicity of $T(x)$ on $x > \min\{\alpha(f)\}$, imply that the solution for L is unique, and thus so is the solution for $P_0(f)$. ■

We note finally that this method of proof can quite easily be extended to cover the discrete case.

References

[1] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, New York, 1968.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[3] J. G. Proakis, *Digital Communications*. USA: McGraw-Hill, fourth ed., 2001.

[4] S. Haykin, *Communication Systems*. New York: Wiley, fourth ed., 2001.

[5] C. E. Shannon, “A mathematical theory of communication,” *Bell System Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.