Linear Coding for Fading Channels

Jin-Jun Xiao *

May 7, 2005

Abstract

We consider the linear coding of a memoryless Gaussian source transmitted through a discrete memoryless fading channel with additive white Gaussian noise (AWGN). The goal is to minimize the mean squared error (MSE) of the source reconstruction at the destination while there is average power constraint imposed on the channel input symbols. We show that among all single-letter codes, linear coding achieves the smallest MSE, and is thus optimal. But when block length increases, the linear coding can not approach the Shannon's bound, and it turns out to share the same performance with a single-letter coding. In spite of its suboptimality, the performance loss of linear coding compared to the optimal coding can be quantitively bounded in terms of the variance of the fading gain and received signal to noise ratio (SNR). We also show that for linear coding, when there is no transmitter channel state information (CSI), uniform power allocation is optimal, and in the presence of transmitter CSI, the optimal power allocation can be analytical solved in terms of the channel fading gains.

Keywords

Linear coding, joint source-channel coding, fading channels.

1 Introduction

Shannon has shown in [1, Theorem 21] that in a point-to-point link, when a discrete memoryless source is transmitted through a discrete memoryless channel, the optimal tradeoff between (channel input) cost and (source reconstruction distortion) can be achieved by separate source and channel coding. Despite its conceptual beauty, in practice, to approach the optimal pair of cost and distortion, the separate source and channel coding leads to high complexity and long delay when block length increases.

Although joint source-channel coding does not have pleasing separation property, but it sometimes can lead to very simple optimal coding strategy. A well-known example is when a memoryless

*Course project of EE 8510: Multiuser information theory, spring 2005, University of Minnesota; Instructor: Prof. Nihar Jindal.

Gaussian source transmitted through an AWGN channel, an amplify and forward transmission strategy achieves the optimal power-distortion tradeoff [2,3]. The perfect match between the source and channel leads to a very simple but optimal coding strategy which is both theoretically and practically appealing. Unfortunately, when source and channel do not come up with such a natural match, the simple but optimal coding is not easy to find. In this work, we study the case when the source is still Gaussian but there is fading in the channel. We analyze the performance of linear coding in such a source-channel communication system, and focus on the following questions: How optimal is the linear coding? Can linear coding achieve the Shannon's bound when block length increases? If not, how can the performance loss be bounded?

Specifically, we consider a memoryless Gaussian source $\{S(t) : t \in \mathbb{Z}^+\}$ which has instantaneously distribution $N(0, \sigma_S^2)$, and is transmitted through a discrete memoryless fading channel

$$Y(i) = h(i)X(i) + W(i), \quad i = 1, 2, \dots,$$

where

i. $W(i)$ are AWGN with unitary variance,

ii. $h(i)$ are i.i.d. fading with known distribution h.

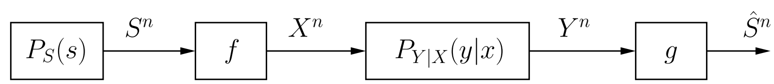

Figure 1: A source-channel coding system where f and g are encoder and decoder respectively.

We suppose there is channel state information (CSI) at receiver only. A general source-channel coding scheme of block length n is depicted in Fig. 1. When we limit to the class of linear coding, the encoder is then a $n \times n$ matrix which maps the source symbols $S^{(n)}$ to channel input symbols $X^{(n)}$. The decoder is then the mean square error estimator (MMSE) estimating $S^{(n)}$ based on $Y^{(n)}$.

In next section, we analyze the performance of linear coding in two cases: i. $n = 1$; ii. $n \geq 2$, where n is the coding block length.

2 Linear Coding

In this section we consider the linear coding when there is receiver CSI only. We start with the case when block length $n = 1$.

$$S(i) \longrightarrow \boxed{f(i)} \longrightarrow X(i) \longrightarrow \boxed{\text{channel}} \longrightarrow Y(i) \longrightarrow \boxed{g(i)} \longrightarrow \hat{S}(i)$$

Figure 2: A linear coding system with block length $n = 1$.

2.1 Linear Coding of Block Length $n = 1$

When the block length $n = 1$, we need to design encoders $f(i)$ for all $i \in \mathbb{Z}$ (see Fig. 2). In such a case, $f(i)$ simply scales $S(i)$ to satisfy the power constraint. Define $P(i) \stackrel{\text{def}}{=} E(|X(i)|^2)$. Then the power constraint implies

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} P(i) \leq P. \tag{1}$$

The coding scheme is given as

$$X(i) = \sqrt{\frac{P(i)}{\sigma_s^2}} S(i); \quad \hat{S}(i) = \frac{P(i)h(i)^2}{P(i)h(i)^2 + 1} y(i). \tag{2}$$

It is easy to calculate that the achieved mean squared distortion

$$D_l(P(i)) \stackrel{\text{def}}{=} E(|S(i) - \hat{S}(i)|^2) = \sigma_S^2 E_{h(i)} \left\{ \frac{1}{1 + h(i)^2 P(i)} \right\} = \sigma_S^2 E_h \left\{ \frac{1}{1 + h^2 P(i)} \right\}. \tag{3}$$

Theorem 1. *Among all single-letter codes of the source-channel coding of a memoryless Gaussian source that is transmitted through a fading channel with AWGN, the linear coding given in* (2) *is optimal.*

Before proving the above theorem, we need the following lemma.

Lemma 2. *Let S be a Gaussian random variable with variance σ_S^2, and $\hat{S}$ be any random variable jointly distributed with S. Then*

$$\frac{E(|S - \hat{S}|^2)}{\sigma_S^2} \geq \exp\left(-2I(S; \hat{S})\right).$$

Proof. We have the following chain of inequalities:

$$\begin{aligned}
I(S; \hat{S}) &= h(S) - h(S|\hat{S}) \\
&= h(S) - h(S - \hat{S}|\hat{S}) \\
&\geq h(S) - h(S - \hat{S}) \\
&\geq h(S) - \frac{1}{2}\log\left(2\pi e E(|S - \hat{S}|^2)\right) \\
&= \frac{1}{2}\log(2\pi e \sigma_S^2) - \frac{1}{2}\log\left(2\pi e E(|S - \hat{S}|^2)\right) \\
&= \frac{1}{2}\log \frac{\sigma_S^2}{E(|S - \hat{S}|^2)}.
\end{aligned}$$

Thus the proof is complete. □

Proof of Theorem 1: For any single-letter codes $\{f(i), g(i)\}$ with $X(i) = f(S(i))$ and $\hat{S}(i) = g(Y(i))$, we have the following Markov chain for any i:

$$S(i) \longrightarrow X(i) \longrightarrow Y(i) \longrightarrow \hat{S}(i).$$

By data processing inequality, we have

$$I(S(i);\hat{S}(i)) \le I(X(i);Y(i)) \le \frac{1}{2}\log(1+h^2(i)P(i)). \tag{4}$$

Combing (4) and Lemma 2, we obtain that when there is power constraint $P(i)$, and the fading coefficient is $h(i)$, the achieved distortion at time i:

$$E\left(|S(i)-\hat{S}(i)|^2\Big|\, h(i)\right) \ge \sigma_S^2 \exp\left(-\frac{1}{2}I(S(i);\hat{S}(i))\right) \ge \frac{\sigma_S^2}{1+h^2(i)P(i)}.$$

Therefore,

$$E(|S(i)-\hat{S}(i)|^2) = E_{h(i)}\left\{E\left(|S(i)-\hat{S}(i)|^2\Big|\, h(i)\right)\right\} \ge \sigma_S^2 E_{h(i)}\left\{\frac{1}{1+h(i)^2P(i)}\right\}.$$

It is easy to see that the equality is obtained by linear coding (c.f. (3)). Therefore among all single-letter codes, linear coding is optimal. □

We have shown the optimality of linear coding among all single-letter codes when the power allocation $\{P(i): i\in\mathbb{Z}\}$ is given. We further show that uniform power allocation is optimal. Suppose P is the average power distortion, for the linear coding, the mean square distortion averaged over time is

$$\begin{aligned} E(|S-\hat{S}|^2) &\stackrel{\text{def}}{=} \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E(|S(i)-\hat{S}(i)|^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E_{h(i)}\left\{\frac{\sigma_S^2}{1+h^2(i)P(i)}\right\} \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E_h\left\{\frac{\sigma_S^2}{1+h^2P(i)}\right\} \ge \sigma_S^2 E_h\left\{\frac{1}{1+h^2P}\right\}, \end{aligned}$$

where the last step is due to the convexity of the function $E_h\left\{\frac{\sigma_S^2}{1+h^2P(i)}\right\}$ in terms of $P(i)$. Therefore linear coding with uniform power allocation is optimal among all single-letter codes. Recalling the definition of $D_l(\cdot)$ in (3), we get that the linear coding with uniform power allocation achieves MSE

$$D_l(P) = \sigma_S^2 E_h\left\{\frac{1}{1+h^2P}\right\}. \tag{5}$$

When there is transmitter CSI, the optimal power loading is not uniform anymore. It can be analytically solved in terms of the channel state h, and is given in the Appendix.

2.2 Linear Coding of Finite Block Length

In this section we consider the linear coding with block length $n \ge 2$. The encoder is given by a $n\times n$ matrix F, and the decoder is an MMSE decoder (see Fig. 3). Let $\Omega_S \stackrel{\text{def}}{=} E(S^{(n)}S^{(n)^T})$, and $H \stackrel{\text{def}}{=} \text{diag}(h^{(n)})$, then we have

$$X^{(n)} = FS^{(n)}, \quad Y^{(n)} = HX^{(n)} + W^{(n)} = HFS^{(n)} + W^{(n)},$$

and

$$\hat{S}^{(n)} = \left(HF\Omega_S F^T H^T + I\right)^{-1} HF\Omega_S F^T H^T Y^{(n)}.$$

The achieved MSE, in terms of the encoding matrix is F and channel matrix is H, can be expressed as:

$$\begin{aligned} D(F,H) &\overset{\text{def}}{=} \frac{1}{n}\mathrm{tr}\left\{E\left((S^{(n)}-\hat{S}^{(n)})(S^{(n)}-\hat{S}^{(n)})^T\right)\right\} \\ &= \frac{1}{n}\mathrm{tr}\left\{\left(HF\Omega_S F^T H^T + I\right)^{-1}\Omega_S\right\}. \end{aligned} \tag{6}$$

The power constraint implies $\mathrm{tr}(F\Omega_S F^T) \leq nP$.

$$S^{(n)} \longrightarrow \boxed{F} \longrightarrow X^{(n)} \longrightarrow \boxed{\text{channel}} \longrightarrow Y^{(n)} \longrightarrow \boxed{\text{MMSE}} \longrightarrow \hat{S}^{(n)}$$

Figure 3: A linear coding system with block length $n \geq 2$.

Thus, introducing $Q = FF^T$, and noticing $\Omega_S = \sigma_S^2 I$, we can solve the following problem to obtain the optimal Q^*, and get the optimal encoding matrix F^*.

$$\begin{aligned} \min \quad & E_H\left\{\mathrm{tr}\left(HQH^T + \sigma_S^{-2}I\right)^{-1}\right\} \\ \text{s.t.} \quad & \mathrm{tr}(Q) \leq \frac{nP}{\sigma_S^2}. \end{aligned} \tag{7}$$

To solve (7), we quote the following two lemmas in matrix algebra without proof.

Lemma 3. *For any square matrix $R \succ 0$, it holds that $\mathrm{tr}(R^{-1}) \geq \sum_{i=1}^{n} R_{ii}^{-1}$, and equality holds iff R is diagonal.*

Lemma 4. *For any square matrices A and B, it holds that $\mathrm{tr}(I+AB)^{-1} = \mathrm{tr}(I+BA)^{-1}$.*

Theorem 5. *Considering the source-channel coding of a memoryless Gaussian source transmitted through an AWGN fading channel, we obtain that any linear coding with finite block length can be performed in single-letter form without performance loss.*

Proof. To solve (7), we first apply Lemma 3 and obtain $\mathrm{tr}\left(HQH^T + \sigma_S^{-2}I\right)^{-1} = \mathrm{tr}\left(QH^TH + \sigma_S^{-2}I\right)^{-1}$ for any H. Then by Lemma 4, we obtain

$$\mathrm{tr}\left(QH^TH + \sigma_S^{-2}I\right)^{-1} \geq \sum_{i=1}^{n} \frac{1}{Q_{ii}h(i)^2 + \sigma_S^{-2}}, \tag{8}$$

where equality holds iff Q is diagonal. Therefore, the optimal solution gives diagonal $Q^* = FF^T$. Thus, any $F^* = \sqrt{Q^*}U$ where U is unitary is an optimal solution. Specifically, if we take $U = I$, we can obtain a diagonal F^*. So any linear coding can be achieved in a single-letter form without performance loss. □

2.3 Comparison of the Performance of Linear Coding with the Shannon's Bound

In this section we examine the performance of linear coding, and compare it with the Shannon's bound (which is theoretically the best achievable). According to the separation theorem, the Shannon's bound can be obtained by combining the rate-distortion function and channel capacity. The rate-distortion function of a memoryless Gaussian source with variance σ_S^2 is

$$R(D) = \frac{1}{2}\log^{+}\frac{\sigma_S^2}{D}. \tag{9}$$

Combining it with the channel capacity (when there is only Receiver CSI, and power constraint is P)

$$C(P) = E_h\left\{\frac{1}{2}\log(1+h^2P)\right\},$$

we obtain the best achievable distortion in terms of P is

$$D^*(P) = \sigma_S^2 \exp\left(E_h\left\{\log\frac{1}{1+h^2P}\right\}\right). \tag{10}$$

Recalling (5), it is easy to see that $D_l(P) \geq D^*(P)$ from concavity of the log-function. The equality holds iff

$$E_h\left\{\log\frac{1}{1+h^2P}\right\} = \log\left(E_h\left\{\frac{1}{1+h^2P}\right\}\right) \Longleftrightarrow \frac{1}{1+h^2P} = const.$$

Theorem 6. *Linear coding (with block length $n=1$) is optimal in Shannon's sense iff $|h|$ is deterministic.*

Therefore, we see that linear coding is optimal only if h is real, $h \equiv \pm c$, or if h is complex, h is distributed on a circle. For all other cases, we know that linear coding is suboptimal, and can not achieve Shannon's bound. In what follows, we bound the performance gap in terms of P and the statistic of h.

Theorem 7. *The MSE performance of linear coding can be bounded away from the theoretically best achievable MSE (i.e., the Shannon's bound) as follows. Define*

$$\gamma(P,h) \stackrel{\text{def}}{=} \frac{D_l(P) - D^*(P)}{D^*(P)},$$

which is the relative gap of the MSE achieved by linear coding from the Shannon's bound. Then we have

$$0 \leq \gamma(P,h) \leq P\sqrt{Var(|h|^2)}.$$

Proof. Introducing $h_0^2 = E(|h^2|)$, then from (10) we obtain

$$D^*(P) = \sigma_S^2 \exp\left(E_h\left\{\log\frac{1}{1+h^2P}\right\}\right) \geq \sigma_S^2\frac{1}{1+h_0^2P} \stackrel{\text{def}}{=} D_0(P).$$

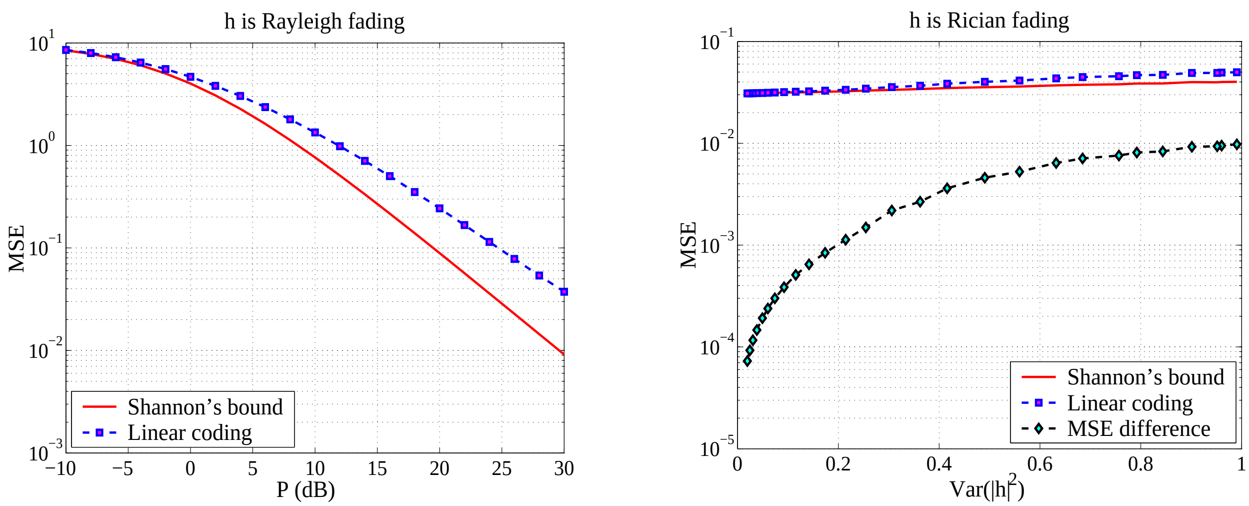

Figure 4: MSE performance as P or $\mathrm{Var}(|h|^2)$ increases. In the left plot, we assume h is Rayleigh fading with $\mathrm{Var}(|h|^2) = 1$, but the transmit power P increases. In the right plot, we assume P=15dB, and h is Rician fading with increasing variance $\mathrm{Var}(|h|^2)$ (by taking different Rician factors).

Notice that the right term in the above formula is the best achievable performance when channel is AWGN with the same average path gain, thus it is a natural performance bound for the fading case. Recalling (5), we obtain

$$\begin{aligned} 0 \le \gamma(h,P) &= \frac{D_l(P) - D^*(P)}{D^*(P)} \le \frac{D_l(P) - D_0(P)}{D_0(P)} \\ &= E_h\left\{\frac{(h_0^2 - h^2)P}{1+h^2P}\right\} \le E_h\left\{|h^2 - h_0^2|P\right\} \le P\sqrt{\mathrm{Var}(|h|^2)}. \end{aligned}$$

The proof is complete. □

Therefore, the performance of linear coding is close to Shannon's bound if $\mathrm{Var}(|h|^2)$ is small, or if P is small, where the latter happens, for example, in the applications of sensor networks. The plot of the MSE performance of linear coding compared to the Shannon's bound in the case of Rayleigh fading with increasing power (when fading is fixed), or Rician fading with increasing variance (when power P is fixed) are plotted in Fig. 4. We can see that when either the transmit power or the variance of fading gain is relatively small, the MSE performance gap is negligible. In Fig. 5, we numerically examine the behavior of the gap coefficient $\gamma(P,h)$ for different fading distributions. The universal upper bound of $\gamma(P,h)$ given in Theorem 6 is quite loose for Rician distribution, but is better when the fading is on-off where h takes values from a two-point set $\{0,1\}$.

3 Conclusion and Future Work

We have studied the performance linear coding of a memoryless Gaussian source transmitted through a discrete memoryless fading channel. We show that linear coding is optimal among all single-letter

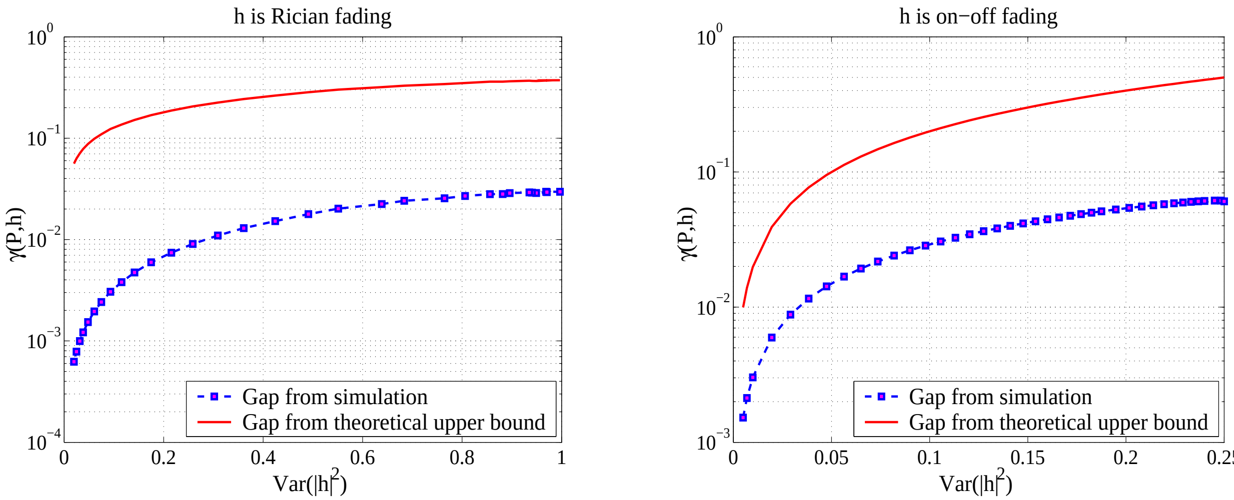

Figure 5: Performance gap of linear coding compared to the Shannon's bound. The red curves are the universal bound of the gap $\gamma(P,h)$ given in Theorem 6 for all possible distributions. In addition, in the left plot, we assume $P = 0$dB, and h is Rician fading with increasing $\mathrm{Var}(|h|^2)$ (by taking different Rician factors); in the right plot, we assume $P = 0$dB, and h is on-off fading with increasing $\mathrm{Var}(|h|^2)$.

codes, but the performance of linear coding can not be improved by increasing the block length. Thus, in general, the linear coding can not achieve the Shannon's bound unless the magnitude of the fading coefficient is a constant. We bound the performance gap of linear coding from the optimal coding in terms of the variance of the fading gain and transmit SNR. Simulation shows that the gap is negligible if either transmit power is or the variance of fading gain is relatively small.

As future directions, we are still investigating simple joint source-channel coding that can obtain or approach the Shannon's bound for the fading channels. We have shown in this paper that linear coding is not a valid candidate, thus, other simple coding schemes need to be proposed. Also we have shown that linear coding is optimal among all single-letter codes when a Gaussian source is matched with a fading channel with AWGN. In general, it is worthwhile to find the best coding when there is constraint imposed on block length. In such a scenario, separation theorem does not hold anymore. We are interested in characterizing necessary/sufficient conditions for such best coding schemes with limited block length for general source-channel pairs.

Appendix: Optimal Power Loading in Presence of Transmitter CSI

If there is CSI at the transmitter, the optimal power loading (along the time i) is not uniform anymore. Instead, the optimal power loading is according to the CSI. Suppose when the channel state is h, the corresponding power loading is $P(h)$. Then the achieved average distortion is

$$D = E_h \left\{ \frac{\sigma_s^2}{1 + h^2 P_2(h)} \right\}.$$

The optimal power loading $P^*(h)$ can be solved from the following problem.

$$\begin{aligned} \min \quad & E_h\left\{\frac{\sigma_s^2}{1+h^2P_2(h)}\right\} \\ \text{s.t.} \quad & E_h\{P(h)\} = P, \quad P(h) \geq 0. \end{aligned}$$

If h has finite states, and $P(h = h_i) = f_i,\ i = 1, 2, \ldots, L$, then we have

$$\begin{aligned} \min \quad & \sum_{i=1}^{L} \frac{\sigma_s^2}{1+h_i^2P_i} f_i \\ \text{s.t.} \quad & \sum_{i=1}^{L} P_i f_i = P, \quad P_i \geq 0. \end{aligned}$$

The Lagrangian is

$$G(P, \lambda, \mu) = \sum_{i=1}^{L} \frac{\sigma_s^2}{1+h_i^2P_i} f_i + \mu\left(\sum_{i=1}^{L} P_i f_i - P\right) - \sum_{i=1}^{L} \lambda_i P_i.$$

It is easy to see that at the optimal point,

$$\frac{\partial G}{\partial P_i} = -\frac{\sigma_s^2 h_i^2}{(1+h_i^2P_i)^2} f_i + \mu f_i - \lambda_i = 0.$$

For those $P_i \neq 0$, we have $\lambda_i = 0$, and

$$\frac{\sigma_s^2 h_i^2}{(1+h_i^2P_i)^2} = \mu.$$

Therefore,

$$P_i^{opt} = \frac{1}{h_i}\left(u_0 - \frac{1}{h_i}\right)^+,$$

where u_0 is a common threshold for all states. If h has a non-discrete pdf, then by discretizing the pdf, we can obtain

$$P^{opt}(h) = \frac{1}{h}\left(u_0 - \frac{1}{h}\right)^+, \tag{11}$$

and u_0 is solved from

$$\int_0^\infty P^{opt}(h) f(h) dh = \int_{\frac{1}{u_0}}^\infty \frac{1}{h}\left(u_0 - \frac{1}{h}\right) f(h) dh = P.$$

Since the obtained power loading is not water-filling (which achieves the capacity when there is transmitter CSI), therefore the linear coding with power loading given in (11) is not optimal either. It can be further verified as follows. The performance of separate source and channel coding is obtained by combining the channel capacity

$$C(P) = E_h\left\{\frac{1}{2}\log(1 + h^2 P_w(h))\right\}.$$

and rate-distortion function in (9), which gives

$$D_w^*(P) = \sigma_S^2 \exp\left(E_h \left\{ \log \frac{1}{1 + h^2 P_w(h)} \right\} \right),$$

where $P_w(h)$ is water-filling power loading that maximizes the channel capacity under average power constraint P. The linear coding with any power loading strategy $P_u(h)$ has performance

$$D_u(P) = \sigma_S^2 E_h \left\{ \frac{1}{1 + h^2 P_u(h)} \right\}.$$

It is easy to see that

$$E_h \left\{ \frac{1}{1 + h^2 P_u(h)} \right\} \geq \exp\left(E_h \left\{ \log \frac{1}{1 + h^2 P_u(h)} \right\} \right) \geq \exp\left(E_h \left\{ \log \frac{1}{1 + h^2 P_w(h)} \right\} \right),$$

where the last inequality holds since $P_w(h)$ is the capacity achieving power loading. Therefore $D_w^*(P) \leq D_u(P)$, where equality holds if and only if $|h| \equiv 1$, or $P = 0$.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379-423, 623-656, 1948.

[2] T. J. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Transactions on Information Theory*, vol. 11, pp. 558-567, Oct. 1965.

[3] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source channel communication revisited", *IEEE Transactions on Information Theory*, Vol. 49, pp. 1147-1158, May 2003.