

Research Article

Saddle-Point Properties and Nash Equilibria for Channel Games

Rudolf Mathar¹ and Anke Schmeink²

¹Institute for Theoretical Information Technology, RWTH Aachen University, 52056 Aachen, Germany

²UMIC Research Center, RWTH Aachen University, 52056 Aachen, Germany

Correspondence should be addressed to Rudolf Mathar, mathar@ti.rwth-aachen.de

Received 15 September 2008; Accepted 4 March 2009

Recommended by Holger Boche

In this paper, transmission over a wireless channel is interpreted as a two-person zero-sum game, where the transmitter gambles against an unpredictable channel, controlled by nature. Mutual information is used as payoff function. Both discrete and continuous output channels are investigated. We use the fact that mutual information is a convex function of the channel matrix or noise distribution densities, respectively, and a concave function of the input distribution to deduce the existence of equilibrium points for certain channel strategies. The case that nature makes the channel useless with zero capacity is discussed in detail. For each, the discrete, continuous, and mixed discrete-continuous output channel, the capacity-achieving distribution is characterized by help of the Karush-Kuhn-Tucker conditions. The results cover a number of interesting examples like the binary asymmetric channel, the Z-channel, the binary asymmetric erasure channel, and the n -ary symmetric channel. In each case, explicit forms of the optimum input distribution and the worst channel behavior are achieved. In the mixed discrete-continuous case, all convex combinations of some noise-free and maximum-noise distributions are considered as channel strategies. Equilibrium strategies are determined by extending the concept of entropy and mutual information to general absolutely continuous measures.

Copyright © 2009 R. Mathar and A. Schmeink. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Transmission over a band-limited wireless channel is often considered as a game where players compete for a scarce medium, the channel capacity. Nash bargaining solutions are determined for interference games with Gaussian additive noise. In the works [1, 2], different fairness and allocation criteria arise from this paradigm leading to useful access control policies for wireless networks.

The engineering problem of transmitting messages over a channel with varying states may also be gainfully considered from a game-theoretic point of view, particularly if the channel state is unpredictable. Here, two players are entering the scene, the transmitter and the channel state selector. The transmitter gambles against the channel state, chosen by a malicious nature, for example. Mutual information $I(X; Y)$ is considered as payoff function, the transmitter aims at maximizing, nature at minimizing $I(X; Y)$. A simple motivating example is the additive scalar channel with input X and additive Gaussian noise Z subject to average power

constraints $E(X^2) \leq P$ and $E(Z^2) \leq \sigma^2$. By standard arguments from information theory, it follows that

$$\begin{aligned} & \max_{X: E(X^2) \leq P} \min_{Z: E(Z^2) \leq \sigma^2} I(X; X + Z) \\ &= \min_{Z: E(Z^2) \leq \sigma^2} \max_{X: E(X^2) \leq P} I(X; X + Z) \\ &= \frac{1}{2} \log \left(1 + \frac{P}{\sigma^2} \right) \end{aligned} \quad (1)$$

is the capacity of the channel. Hence an equilibrium point exists and capacity is the value of the two-person zero-sum game. The corresponding equilibrium strategies are to increase power and noise, respectively, to their maximum values.

A similar game is considered in [3], where the *coder* controls the input and the *jammer* the noise, both from allowable sets. Saddle points, hence equilibria, and ϵ -optimal strategies are determined for binary input and output quantization under power constraints for both the coder and the jammer. An extension of the mutual information game (1) to vector channels with convex covariance constraints

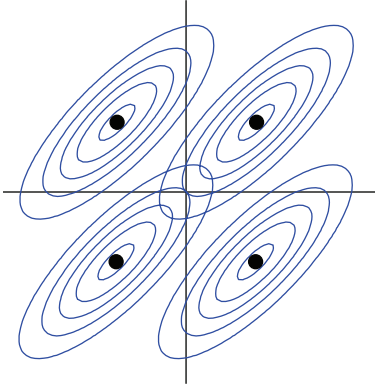


FIGURE 1: 4-QAM as an example of a continuous channel model. Signaling points (black circles) and contour lines of a two-dimensional Gaussian noise distribution with unit variances and correlation $\rho = 0.8$ are shown.

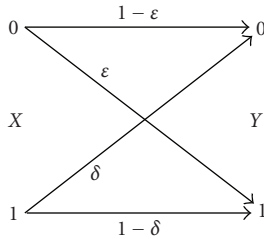


FIGURE 2: The binary asymmetric channel.

is considered in [4]. Jorswieck and Boche [5] investigate a similar minimax setup for a single link in a MIMO system with different types of interference. Further extensions to vector channels and different kinds of games are considered (e.g., [6, 7]).

In this paper, we choose the approach that nature gambles against the transmitter, which aims at conveying information across the channel in an optimal way. “Nature” and “channel” are used synonymously to characterize the antagonist of the transmitter. We consider two models of the channel which yield comparable results. First, transmission is considered purely on a symbol basis. Symbols from a finite set are transmitted and decoded with certain error probabilities. The model is completely discrete, and strategies of nature are described by certain channel matrices chosen from the set of stochastic matrices. The binary asymmetric erasure channel as shown in Figure 4 may serve as a typical example.

On the other hand, continuous channel models are considered. The strategies of the channel are then given by a set of densities, each describing the conditional distribution of received values given a transmitted symbol. The finite input additive white Gaussian noise channel is a standard example hereof, and also 4-QAM with correlated noise (e.g., as shown in Figure 1) is covered by this model.

For both models, equilibrium points are sought, where the strategy of the transmitter consists of selecting the optimum input distribution against the worst-case behavior

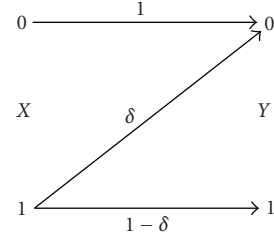


FIGURE 3: The Z-channel.

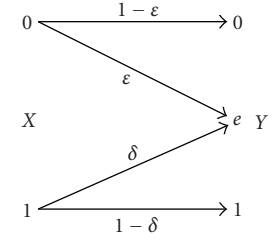


FIGURE 4: The binary asymmetric erasure channel.

of the channel, vice versa, and both have the same game value.

The contributions of this paper are as follows. In Section 2, we demonstrate that mutual information is a convex function of the channel matrix, or the noise densities, respectively. For discrete channels, transmission is considered as a game in Section 3. Some typical binary and n -ary channels are covered by this theory, as shown in Section 5. It is demonstrated that equilibrium points exist and the according optimum strategies for both players are determined. The entropy of mixture distributions is considered in Section 6, which finally, in Section 7, leads to equilibrium points for mixed discrete-continuous channel strategies.

2. Channel Models and Mathematical Foundations

Denote the set of stochastic vectors of dimension m by

$$\mathcal{D}^m = \left\{ \mathbf{p} = (p_1, \dots, p_m) \mid p_i \geq 0, \sum_{i=1}^m p_i = 1 \right\}. \quad (2)$$

Each $\mathbf{p} \in \mathcal{D}^m$ represents a discrete distribution with m support points. The entropy H of \mathbf{p} is defined as

$$H(\mathbf{p}) = -\sum_{i=1}^m p_i \log p_i. \quad (3)$$

If \mathbf{p} characterizes the distribution of some discrete random variable X , we synonymously write $H(X) = H(\mathbf{p})$. It is well known that the entropy H is a concave function of \mathbf{p} , and furthermore, even Schur-concave over the set of distributions \mathcal{D}^m , since it is symmetric (see [8]).

Let random variable \mathbf{X} denote the discrete channel input with symbol set $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ and distribution \mathbf{p} . Accordingly, random variable \mathbf{Y} denotes the output of the channel.

2.1. Discrete Output Channels. We first deal with discrete channels. If the output set consists of n symbols $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$, then the behavior of the channel is completely characterized by the $(m \times n)$ channel matrix:

$$\mathbf{W} = (w_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}, \quad (4)$$

consisting of conditional probabilities $w_{ij} = P(\mathbf{Y} = \mathbf{y}_j | \mathbf{X} = \mathbf{x}_i)$. Matrix \mathbf{W} is an element of the set of stochastic $(m \times n)$ matrices, denoted by $\mathcal{S}^{m \times n}$. Its rows are stochastic vectors, denoted by $\mathbf{w}_1, \dots, \mathbf{w}_m \in \mathcal{D}^n$. The distribution of \mathbf{Y} is then given by the stochastic vector $\mathbf{q} = \mathbf{p}\mathbf{W}$.

Mutual information for this channel model reads as

$$\begin{aligned} I(\mathbf{X}; \mathbf{Y}) &= H(\mathbf{Y}) - H(\mathbf{Y} | \mathbf{X}) \\ &= H(\mathbf{p}\mathbf{W}) - \sum_{i=1}^m p_i H(\mathbf{w}_i) \\ &= \sum_{i=1}^m p_i D(\mathbf{w}_i \| \mathbf{p}\mathbf{W}), \end{aligned} \quad (5)$$

where $D(\cdot \| \cdot)$ denotes the Kulback-Leibler divergence,

$$D(\mathbf{p} \| \mathbf{q}) = \sum_{i=1}^m p_i \log \frac{p_i}{q_i} \quad (6)$$

with $\mathbf{p}, \mathbf{q} \in \mathcal{D}^m$.

Obviously, mutual information depends on the input distribution \mathbf{p} , controlled by the transmitter, and channel matrix \mathbf{W} , controlled by nature. To emphasize this dependence, we also write $I(\mathbf{X}; \mathbf{Y}) = I(\mathbf{p}; \mathbf{W})$. The following result is quoted from [9, Lemma 3.5].

Proposition 1. *Mutual information $I(\mathbf{p}; \mathbf{W})$ is a concave function of $\mathbf{p} \in \mathcal{D}^m$ and a convex function of $\mathbf{W} \in \mathcal{S}^{m \times n}$.*

The proof relies on the representation in the third line of (5), convexity of the Kulback-Leibler divergence $D(\mathbf{p} \| \mathbf{q})$ as a function of the pair (\mathbf{p}, \mathbf{q}) , and concavity of the entropy H .

The problem of maximizing $I(\mathbf{p}; \mathbf{W})$ over \mathbf{p} or minimizing $I(\mathbf{p}; \mathbf{W})$ over \mathbf{W} subject to convex constraints hence fall into the class of convex optimization problems.

2.2. General Output Channels. Entropy definition (3) generalizes to densities f of absolutely continuous distributions with respect to a σ -finite measure μ as

$$H(f) = \int f(y) \log f(y) d\mu(y) \quad (7)$$

(see [10]). Practically relevant cases are the discrete case (3), where μ is taken as the counting measure, densities f , with respect to the Lebesgue measure λ^n on the σ -field of Borel sets over \mathbb{R}^n , and mixtures hereof. These cases correspond to discrete, continuous, and mixed discrete-continuous random variables.

The approach in Section 2.1 carries over to densities of absolutely continuous distributions with respect to μ , as used in (7). The channel output \mathbf{Y} is randomly distorted by noise,

for symbol i governed by μ -density f_i . Hence, the distribution of \mathbf{Y} given input $\mathbf{X} = \mathbf{x}_i$ has μ -density

$$f(\mathbf{y} | \mathbf{x}_i) = f_i(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^n. \quad (8)$$

The AWGN channel $\mathbf{Y} = \mathbf{X} + \mathbf{N}$ is a special case hereof with $f_i(\mathbf{y}) = \varphi(\mathbf{y} - \mathbf{x}_i)$. Here, φ denotes the Lebesgue density of a Gaussian distribution $N_n(\mathbf{0}, \mathbf{\Sigma})$.

Mutual information between channel input and output as a function of $\mathbf{p} = (p_1, \dots, p_m)$ and (f_1, \dots, f_m) may be written as

$$\begin{aligned} I(\mathbf{X}; \mathbf{Y}) &= I(\mathbf{p}; (f_1, \dots, f_m)) \\ &= H(\mathbf{Y}) - H(\mathbf{Y} | \mathbf{X}) \\ &= H\left(\sum_{i=1}^m p_i f_i\right) - \sum_{i=1}^m p_i H(f_i) \\ &= \sum_{i=1}^m p_i D\left(f_i \| \sum_{j=1}^m p_j f_j\right), \end{aligned} \quad (9)$$

where $D(f \| g) = \int f \log(f/g) d\mu$ denotes the Kullback-Leibler divergence between μ -densities f and g .

Let \mathcal{F} denote the set of all μ -densities. From the convexity of $t \log t$, $t \geq 0$, it is easily concluded that

$$H\left(\sum_{i=1}^m p_i f_i\right) \text{ is a concave function of } \mathbf{p} \in \mathcal{D}^m. \quad (10)$$

By applying the log-sum inequality (cf. [9]), we also obtain

$$\begin{aligned} \alpha f_1 \log \frac{f_1}{g_1} + (1 - \alpha) f_2 \log \frac{f_2}{g_2} \\ \geq (\alpha f_1 + (1 - \alpha) f_2) \log \frac{\alpha f_1 + (1 - \alpha) f_2}{\alpha g_1 + (1 - \alpha) g_2}, \end{aligned} \quad (11)$$

pointwise for any pairs of densities $(f_1, g_1), (f_2, g_2) \in \mathcal{F}^2$. Integrating both sides of the aforementioned inequality shows that

$$D(f \| g) \text{ is a convex function of the pair } (f, g) \in \mathcal{F}^2. \quad (12)$$

Applying (10) and (12) to the third and fourth lines of representation (9), respectively, gives the following proposition.

Proposition 2. *Mutual information $I(\mathbf{p}; (f_1, \dots, f_m))$ is a concave function of $\mathbf{p} \in \mathcal{D}^m$ and a convex function of $(f_1, \dots, f_m) \in \mathcal{F}^m$.*

Proposition 2 generalizes its discrete counterpart, Proposition 1. The latter is obtained from the former by identifying the rows of \mathbf{W} as densities with respect to the counting measure with support given by the output symbol set.

In summary, determining the capacity of the channel for fixed channel noise densities f_1, \dots, f_m leads to a concave optimization problem, namely,

$$C = \max_{\mathbf{p} \in \mathcal{D}^m} I(\mathbf{p}; (f_1, \dots, f_m)). \quad (13)$$

Further, minimizing $I(\mathbf{p}; (f_1, \dots, f_m))$ over a convex set of densities f_1, \dots, f_m for some fixed input distribution $\mathbf{p} \in \mathcal{D}^m$ yields a convex optimization problem.

3. Discrete Output Channel Games

In what follows, we regard transmission over a channel as a two-person zero-sum game. A malicious nature is gambling against the transmitter. If nature is controlling the channel, the transmitter wants to protect itself against a worst-case behavior of nature in the sense of maximizing the capacity of the channel by an appropriate choice of the input distribution. The question arises whether this type of channel game has an equilibrium. If the transmitter moves first and maximizes capacity under the present channel conditions, is the same game value achieved if nature deteriorates the channel against the chosen strategy of the transmitter? Hence, $I(\mathbf{X}; \mathbf{Y})$ plays the role of the payoff function.

We will show that for different classes of channels equilibria exist. The basis is formed by the following minimax or saddle point theorem.

Proposition 3. *Let $\mathcal{T} \subseteq \mathcal{S}^{m \times r}$ be a closed convex subset of channel matrices. Then the according channel game has an equilibrium point with value*

$$\max_{\mathbf{p} \in \mathcal{D}^m} \min_{\mathbf{W} \in \mathcal{T}} I(\mathbf{p}; \mathbf{W}) = \min_{\mathbf{W} \in \mathcal{T}} \max_{\mathbf{p} \in \mathcal{D}^m} I(\mathbf{p}; \mathbf{W}). \quad (14)$$

The proof is an immediate consequence of von Neumann's minimax theorem (cf. [11, page 131]). Since \mathcal{D}^m and \mathcal{T} are closed and convex, the main premises are concavity in \mathbf{p} and convexity in \mathbf{W} , both properties assured by Proposition 1.

If $\mathcal{T} = \mathcal{S}^{m \times r}$, the value of the game is zero. Nature will make the channel useless by selecting

$$\mathbf{W} = \begin{pmatrix} \mathbf{w} \\ \vdots \\ \mathbf{w} \end{pmatrix}, \quad (15)$$

with constant rows \mathbf{w} yielding $I(\mathbf{p}; \mathbf{W}) = 0$ independent of the input distribution. Obviously, (15) holds if and only if input \mathbf{X} and output \mathbf{Y} are stochastically independent.

We first consider the case that nature plays a singleton strategy, hence $\mathcal{T} = \{\mathbf{W}\}$, a set consisting of only one strategy. However, (14) then reduces to determining $\max_{\mathbf{p} \in \mathcal{D}^m} I(\mathbf{p}; \mathbf{W})$, the capacity C of the channel for fixed channel matrix \mathbf{W} . In order to characterize nonzero capacity channels, we use the variational distance between the i th and j th row of \mathbf{W} , defined as

$$d(\mathbf{w}_i, \mathbf{w}_j) = \sum_{k=1}^r |w_{ik} - w_{jk}|. \quad (16)$$

The condition

$$\max_{1 \leq i, j \leq m} d(\mathbf{w}_i, \mathbf{w}_j) = \gamma(\mathbf{W}) > 0 \quad (17)$$

on the channel matrix \mathbf{W} ensures that the according channel has nonzero capacity, as demonstrated in the following proposition.

Proposition 4. *If \mathbf{W} satisfies (17) for some $\gamma(\mathbf{W}) > 0$, then*

$$C = \max_{\mathbf{p} \in \mathcal{D}^m} I(\mathbf{p}; \mathbf{W}) \geq \frac{\gamma^2(\mathbf{W})}{8 \ln 2} > 0, \quad (18)$$

where information is measured in nats.

Proof. Let the maximum in (17) be attained at indices i_0 and j_0 . Further, set $\mathbf{p} = (1/2)(\mathbf{e}_{i_0} + \mathbf{e}_{j_0})$ where \mathbf{e}_i denotes the i th unit row vector in \mathbb{R}^m . The third line in (5) then gives

$$I(\mathbf{p}; \mathbf{W}) = \frac{1}{2} D\left(\mathbf{w}_{i_0} \parallel \frac{\mathbf{w}_{i_0} + \mathbf{w}_{j_0}}{2}\right) + \frac{1}{2} D\left(\mathbf{w}_{j_0} \parallel \frac{\mathbf{w}_{i_0} + \mathbf{w}_{j_0}}{2}\right). \quad (19)$$

Since

$$D(\mathbf{w}_i \parallel \mathbf{w}_j) \geq \frac{1}{2 \ln 2} d^2(\mathbf{w}_i, \mathbf{w}_j) \quad (20)$$

(see [9, page 58]), and

$$d\left(\mathbf{w}_i, \frac{\mathbf{w}_i + \mathbf{w}_j}{2}\right) = \frac{1}{2} d(\mathbf{w}_i, \mathbf{w}_j), \quad (21)$$

it follows that

$$I(\mathbf{p}; \mathbf{W}) \geq \frac{1}{8 \ln 2} d^2(\mathbf{w}_{i_0}, \mathbf{w}_{j_0}) = \frac{\gamma^2}{8 \ln 2} > 0. \quad (22)$$

□

In summary, some channel with transition probabilities \mathbf{W} has nonzero capacity if and only if $\gamma(\mathbf{W}) > 0$. The same condition turns out important when determining the capacity of arbitrary discrete channels.

Proposition 5. *Let channel matrix \mathbf{W} satisfy condition (17). Then $C = \max_{\mathbf{p} \in \mathcal{D}^m} I(\mathbf{p}; \mathbf{W})$ is attained at $\mathbf{p}^* = (p_1^*, \dots, p_m^*)$ if and only if*

$$D(\mathbf{w}_i \parallel \mathbf{p}^* \mathbf{W}) = \zeta \quad (23)$$

for some $\zeta > 0$ and all i with $p_i^* > 0$. Moreover, $C = I(\mathbf{p}^*; \mathbf{W}) = \zeta$ holds.

Proof. Mutual information $I(\mathbf{p}; \mathbf{W})$ is a concave function of \mathbf{p} . Hence the KKT conditions (cf., e.g., [12]) are necessary and sufficient for optimality of some input distribution \mathbf{p} . Using (5), some elementary algebra shows that

$$\frac{\partial}{\partial p_i} I(\mathbf{p}; \mathbf{W}) = D(\mathbf{w}_i \parallel \mathbf{p} \mathbf{W}) - 1. \quad (24)$$

The full set of KKT conditions now reads as

$$\begin{aligned} \mathbf{p} &\in \mathcal{D}^m, \\ \lambda_i &\geq 0, \quad i = 1, \dots, m, \\ \lambda_i p_i &= 0, \quad i = 1, \dots, m, \end{aligned} \quad (25)$$

$$D(\mathbf{w}_i \parallel \mathbf{p} \mathbf{W}) + \lambda_i + \nu = 0, \quad i = 1, \dots, m,$$

which shows the assertion. □

Proposition 5 has an interesting interpretation. For an input distribution $\mathbf{p}^* = (p_1^*, \dots, p_m^*)$ to be capacity-achieving, the Kulback-Leibler distance between the rows of \mathbf{W} and the weighted average with weights p_i^* has to be the same for all i with positive p_i^* . Hence, capacity-achieving distribution \mathbf{p}^* places the mixture distribution $\mathbf{p}^* \mathbf{W}$ somehow in the middle of all rows \mathbf{w}_i^* .

4. Elementary Channel Models

Discrete binary input channels are considered in this section. From the according channel games capacity-achieving distributions against worst-case channels are obtained.

4.1. The Binary Asymmetric Channel. As an example, we consider the binary asymmetric channel with channel matrix:

$$\mathbf{W} = \mathbf{W}(\varepsilon, \delta) = \begin{pmatrix} 1 - \varepsilon & \varepsilon \\ \delta & 1 - \delta \end{pmatrix} = \begin{pmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{pmatrix}, \quad (26)$$

with $0 < \varepsilon, \delta < 1$ such that condition (17) is satisfied (see Figure 2). By (23), the capacity-achieving input distribution $\mathbf{p} = (p_0, p_1)$ satisfies

$$D(\mathbf{w}_1 \| \mathbf{p}\mathbf{W}) = D(\mathbf{w}_2 \| \mathbf{p}\mathbf{W}). \quad (27)$$

This is an equation in the variables p_0, p_1 which jointly with the condition $p_0 + p_1 = 1$ has the solution

$$p_0^* = \frac{1}{1+b}, \quad p_1^* = \frac{b}{1+b}, \quad (28)$$

with

$$b = \frac{a\varepsilon - (1 - \varepsilon)}{\delta - a(1 - \delta)}, \quad a = \exp\left(\frac{h(\delta) - h(\varepsilon)}{1 - \varepsilon - \delta}\right), \quad (29)$$

and $h(\varepsilon) = H(\varepsilon, 1 - \varepsilon)$, the entropy of $(\varepsilon, 1 - \varepsilon)$. This result has been derived by cumbersome methods in the early paper [13].

Now assume that the strategy set of nature is given by

$$\mathcal{T}_{\hat{\varepsilon}, \hat{\delta}} = \{\mathbf{W}(\varepsilon, \delta) \mid 0 \leq \varepsilon \leq \hat{\varepsilon}, 0 \leq \delta \leq \hat{\delta}\}, \quad (30)$$

where $0 \leq \hat{\varepsilon}, \hat{\delta} < 1/2$ are given. Hence, error probabilities are bounded from the worst case by $\hat{\varepsilon}$ and $\hat{\delta}$.

Since $I(\mathbf{p}; \mathbf{W})$ is a convex function of \mathbf{W} , $I(\mathbf{p}; \mathbf{W}(\varepsilon, \delta))$ is a convex function of the argument $(\varepsilon, \delta) \in [0, 1]^2$. The minimum value 0 is obviously attained whenever $\varepsilon + \delta = 1$. This shows that $I(\mathbf{p}; \mathbf{W}(\varepsilon, \delta))$ is decreasing in $\varepsilon \in [0, \hat{\varepsilon}]$ for fixed δ , and vice versa, is a decreasing function of $\delta \in [0, \hat{\delta}]$ with ε fixed. Accordingly, it holds that

$$\min_{\mathbf{W} \in \mathcal{T}_{\hat{\varepsilon}, \hat{\delta}}} I(\mathbf{p}; \mathbf{W}) = I(\mathbf{p}; \mathbf{W}(\hat{\varepsilon}, \hat{\delta})) \quad (31)$$

for any $\mathbf{p} \in \mathcal{D}^2$. Further,

$$\max_{\mathbf{p} \in \mathcal{D}^2} \min_{\mathbf{W} \in \mathcal{T}_{\hat{\varepsilon}, \hat{\delta}}} I(\mathbf{p}; \mathbf{W}) = \max_{\mathbf{p} \in \mathcal{D}^2} I(\mathbf{p}; \mathbf{W}(\hat{\varepsilon}, \hat{\delta})) \quad (32)$$

is attained at $\mathbf{p}^* = (p_0^*, p_1^*)$ from (28) with the replacements $\varepsilon = \hat{\varepsilon}$ and $\delta = \hat{\delta}$.

Since $\mathcal{T}_{\hat{\varepsilon}, \hat{\delta}}$ is a convex set, we obtain from Proposition 3 that a saddle point exists and the value of the game is given by

$$\begin{aligned} \max_{\mathbf{p} \in \mathcal{D}^2} \min_{\mathbf{W} \in \mathcal{T}_{\hat{\varepsilon}, \hat{\delta}}} I(\mathbf{p}; \mathbf{W}) &= \min_{\mathbf{W} \in \mathcal{T}_{\hat{\varepsilon}, \hat{\delta}}} \max_{\mathbf{p} \in \mathcal{D}^2} I(\mathbf{p}; \mathbf{W}) \\ &= I(\mathbf{p}^*; \mathbf{W}(\hat{\varepsilon}, \hat{\delta})). \end{aligned} \quad (33)$$

The so-called *Z-channel* with error probability $\varepsilon = 0$ and $\delta \in [0, 1]$ (see Figure 3) is a special case hereof. We have

$$\begin{aligned} \max_{\mathbf{p} \in \mathcal{D}^2} \min_{\delta \leq \hat{\delta}} I(\mathbf{p}; \mathbf{W}(0, \delta)) &= \max_{\mathbf{p} \in \mathcal{D}^2} I(\mathbf{p}; \mathbf{W}(0, \hat{\delta})) \\ &= I(\mathbf{p}^*; \mathbf{W}(0, \hat{\delta})). \end{aligned} \quad (34)$$

After some algebra, from (28)

$$p_0^* = 1 - p_1^*, \quad p_1^* = \frac{1/(1 - \hat{\delta})}{1 - 2^{h(\hat{\delta})/(1 - \hat{\delta})}} \quad (35)$$

is obtained with capacity

$$I(\mathbf{p}^*; \mathbf{W}(0, \hat{\delta})) = \log_2(1 + 2^{-h(\hat{\delta})/(1 - \hat{\delta})}), \quad (36)$$

where information is measured in bits (cf. [14, Example 9.11]).

4.2. The Binary Asymmetric Erasure Channel. The binary asymmetric erasure channel (BEC) with bit error probabilities $\varepsilon, \delta \in [0, 1]$, and channel matrix

$$\mathbf{W} = \mathbf{W}(\varepsilon, \delta) = \begin{pmatrix} 1 - \varepsilon & \varepsilon & 0 \\ 0 & \delta & 1 - \delta \end{pmatrix} \quad (37)$$

is depicted in Figure 4.

According to Proposition 4, this channel has zero capacity if and only if $\varepsilon = \delta = 1$. Excluding this case, by Proposition 5, the capacity-achieving distribution $\mathbf{p}^* = (p_0^*, p_1^*)$, $p_0^* + p_1^* = 1$ is given by the solution of

$$\begin{aligned} (1 - \varepsilon) \log \frac{1 - \varepsilon}{p_0(1 - \varepsilon)} + \varepsilon \log \frac{\varepsilon}{p_0\varepsilon + p_1\delta} \\ = \delta \log \frac{\delta}{p_0\varepsilon + p_1\delta} + (1 - \delta) \log \frac{1 - \delta}{p_0(1 - \delta)}. \end{aligned} \quad (38)$$

Substituting $x = p_0/p_1$, (38) reads equivalently as

$$\varepsilon \log \varepsilon - \delta \log \delta = (1 - \delta) \log(\delta + \varepsilon x) - (1 - \varepsilon) \log\left(\varepsilon + \frac{\delta}{x}\right). \quad (39)$$

By differentiating with respect to x , it is easy to see that the right-hand side is monotonically increasing such that exactly one solution $\mathbf{p}^* = (p_1^*, p_2^*)$ exists, which can be numerically computed.

If $\varepsilon = \delta$, the solution is given by $p_0^* = p_1^* = 1/2$, as easily verified from (38).

Resembling the arguments used for the binary asymmetric channel and adopting the notation, we see that

$$\min_{\mathbf{W} \in \mathcal{T}_{\hat{\varepsilon}, \hat{\delta}}} I(\mathbf{p}; \mathbf{W}) = I(\mathbf{p}; \mathbf{W}(\hat{\varepsilon}, \hat{\delta})) \quad (40)$$

for any $\mathbf{p} \in \mathcal{D}^2$. Further,

$$\max_{\mathbf{p} \in \mathcal{D}^2} \min_{\mathbf{W} \in \mathcal{T}_{\hat{\varepsilon}, \hat{\delta}}} I(\mathbf{p}; \mathbf{W}) = \max_{\mathbf{p} \in \mathcal{D}^2} I(\mathbf{p}; \mathbf{W}(\hat{\varepsilon}, \hat{\delta})) \quad (41)$$

is attained at $\mathbf{p}^* = (p_0^*, p_1^*)$, the solution of (38) with ε substituted by $\hat{\varepsilon}$ and δ by $\hat{\delta}$. Finally, the game value amounts to

$$\begin{aligned} \max_{\mathbf{p} \in \mathcal{D}^2} \min_{\mathbf{W} \in \mathcal{T}_{\hat{\varepsilon}, \hat{\delta}}} I(\mathbf{p}; \mathbf{W}) &= \min_{\mathbf{W} \in \mathcal{T}_{\hat{\varepsilon}, \hat{\delta}}} \max_{\mathbf{p} \in \mathcal{D}^2} I(\mathbf{p}; \mathbf{W}) \\ &= I(\mathbf{p}^*; \mathbf{W}(\hat{\varepsilon}, \hat{\delta})). \end{aligned} \quad (42)$$

If $\delta = \varepsilon \leq \hat{\varepsilon}$, the result is

$$I(\mathbf{p}^*; \mathbf{W}(\hat{\varepsilon}, \hat{\delta})) = 1 - \hat{\varepsilon}, \quad (43)$$

and the equilibrium strategies are $p_0^* = p_1^* = 1/2$ for the transmitter and $\varepsilon = \delta = \hat{\varepsilon}$ for nature (cf. [15, Example 8.5]).

5. The n -Ary Symmetric Channel

Consider the n -ary symmetric channel with symbol set $\{0, 1, \dots, n-1\}$ and channel matrix

$$\mathbf{W}(\boldsymbol{\varepsilon}) = \begin{pmatrix} \varepsilon_0 & \varepsilon_1 & \cdots & \varepsilon_{n-1} \\ \varepsilon_{n-1} & \varepsilon_0 & \cdots & \varepsilon_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_0 \end{pmatrix} \quad (44)$$

by cyclically shifting some error vector $\boldsymbol{\varepsilon} = (\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{n-1}) \in \mathcal{D}^n$. Let $\mathcal{E} \subseteq \mathcal{D}^n$ denote the set of strategies that nature can choose the channel state from by selecting some $\boldsymbol{\varepsilon} \in \mathcal{E}$.

If $\mathcal{E} = \mathcal{D}^n$, the value of the game is zero. As mentioned earlier, nature will cripple the channel by selecting

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_u = \left(\frac{1}{n}, \dots, \frac{1}{n} \right), \quad (45)$$

yielding $I(\mathbf{X}; \mathbf{Y}) = 0$ independent of the input distribution. Note that $\boldsymbol{\varepsilon}_u$ is the unique minimum element with respect to majorization, that is, $\boldsymbol{\varepsilon}_u < \boldsymbol{\varepsilon}$ for all $\boldsymbol{\varepsilon} \in \mathcal{D}^n$. We briefly recall the corresponding definitions (see [8]). Let $p_{[i]}$ and $q_{[i]}$ denote the components of \mathbf{p} and \mathbf{q} in decreasing order, respectively. Distribution $\mathbf{p} \in \mathcal{S}$ is said to be majorized by $\mathbf{q} \in \mathcal{S}$, in symbols $\mathbf{p} < \mathbf{q}$, if $\sum_{i=1}^k p_{[i]} \leq \sum_{i=1}^k q_{[i]}$ for all $k = 1, \dots, m$.

Hence, to avoid trivial cases, the set of strategies for nature has to be separated from this worst case.

5.1. Separation by Schur Ordering. We first investigate the set

$$\begin{aligned} \mathcal{E}_{>\hat{\varepsilon}} &= \{ \boldsymbol{\varepsilon} = (\varepsilon_0, \dots, \varepsilon_{n-1}) \in \mathcal{D}^n \mid \\ &\hat{\varepsilon} < \boldsymbol{\varepsilon}, \varepsilon_{\pi(0)} \leq \dots \leq \varepsilon_{\pi(n-1)} \} \end{aligned} \quad (46)$$

for some fixed $\hat{\varepsilon} \neq \boldsymbol{\varepsilon}_u$ and permutation π . This means that the error probabilities are at least spread out, or separated from uniformity as $\hat{\varepsilon}$, with error probabilities increasing in the fixed order determined by π .

Since $\mathcal{E}_{>\hat{\varepsilon}}$ is convex and closed, the set of corresponding matrices

$$\mathcal{T}_{>\hat{\varepsilon}} = \{ \mathbf{W}(\boldsymbol{\varepsilon}) \mid \boldsymbol{\varepsilon} \in \mathcal{E}_{>\hat{\varepsilon}} \} \quad (47)$$

is convex and closed as well.

Proposition 3 ensures the existence of an equilibrium point:

$$\max_{\mathbf{p} \in \mathcal{D}^n} \min_{\mathbf{W} \in \mathcal{T}_{>\hat{\varepsilon}}} I(\mathbf{p}; \mathbf{W}) = \min_{\mathbf{W} \in \mathcal{T}_{>\hat{\varepsilon}}} \max_{\mathbf{p} \in \mathcal{D}^n} I(\mathbf{p}; \mathbf{W}). \quad (48)$$

To determine the value ν of the game, we first consider $\max_{\mathbf{p} \in \mathcal{D}^n} I(\mathbf{p}; \mathbf{W}(\boldsymbol{\varepsilon}))$ for some fixed $\boldsymbol{\varepsilon} \in \mathcal{E}_{>\hat{\varepsilon}}$. From (5), it follows that the maximum is attained at input distribution $\mathbf{p} = (1/n, \dots, 1/n)$ with value

$$\max_{\mathbf{p} \in \mathcal{D}^n} I(\mathbf{p}; \mathbf{W}(\boldsymbol{\varepsilon})) = \log n - H(\boldsymbol{\varepsilon}). \quad (49)$$

As the entropy is Schur concave, $\min_{\boldsymbol{\varepsilon} \in \mathcal{E}_{>\hat{\varepsilon}}} (\log n - H(\boldsymbol{\varepsilon}))$ is attained at $\hat{\varepsilon}$ such that the value of the game is obtained as

$$\min_{\mathbf{W} \in \mathcal{T}_{>\hat{\varepsilon}}} \max_{\mathbf{p} \in \mathcal{D}^n} I(\mathbf{p}; \mathbf{W}) = \log n - H(\hat{\varepsilon}) \quad (50)$$

with according equilibrium strategies $\mathbf{p} = (1/n, \dots, 1/n)$ and the components of $\boldsymbol{\varepsilon}$ equal to those of $\hat{\varepsilon}$ rearranged according to π .

5.2. Directional Separation. In what follows, we consider channel states separated from the worst-case $\boldsymbol{\varepsilon}_u$ into the direction of some prespecified $\hat{\varepsilon} \in \mathcal{D}^n$, $\hat{\varepsilon} \neq \boldsymbol{\varepsilon}_u$. This set of strategies is formally described as

$$\mathcal{E}_{\hat{\alpha}, \hat{\varepsilon}} = \{ \boldsymbol{\varepsilon} = (1 - \alpha)\boldsymbol{\varepsilon}_u + \alpha\hat{\varepsilon} \mid \hat{\alpha} \leq \alpha \leq 1 \} \quad (51)$$

for some given $\hat{\alpha} > 0$. It is obviously convex and closed. The set of corresponding channel matrices

$$\mathcal{T}_{\hat{\alpha}, \hat{\varepsilon}} = \{ \mathbf{W}(\boldsymbol{\varepsilon}) \mid \boldsymbol{\varepsilon} \in \mathcal{E}_{\hat{\alpha}, \hat{\varepsilon}} \} \quad (52)$$

is also closed and convex such that an equilibrium exists by Proposition 3. It remains to determine the game value.

Since $I(\mathbf{p}; \mathbf{W})$ is a convex function of \mathbf{W} , hence decreasing in $\alpha \in [\hat{\alpha}, 1)$:

$$\min_{\mathbf{W} \in \mathcal{T}_{\hat{\alpha}, \hat{\varepsilon}}} I(\mathbf{p}; \mathbf{W}) \quad (53)$$

is attained at $\mathbf{W}(\boldsymbol{\varepsilon}_{\hat{\alpha}})$ with $\boldsymbol{\varepsilon}_{\hat{\alpha}} = (1 - \hat{\alpha})\boldsymbol{\varepsilon}_u + \hat{\alpha}\hat{\varepsilon}$. From representation (5), it can be easily seen that

$$\max_{\mathbf{p} \in \mathcal{D}^n} \min_{\mathbf{W} \in \mathcal{T}_{\hat{\alpha}, \hat{\varepsilon}}} I(\mathbf{p}; \mathbf{W}) \quad (54)$$

is attained at $\mathbf{p} = (1/n, \dots, 1/n)$.

Vice versa, from (5), it follows that for any $\mathbf{W} = \mathbf{W}(\boldsymbol{\varepsilon})$,

$$\max_{\mathbf{p} \in \mathcal{D}^n} I(\mathbf{p}; \mathbf{W}(\boldsymbol{\varepsilon})) = \log n - H(\boldsymbol{\varepsilon}) \quad (55)$$

is attained at $\mathbf{p} = (1/n, \dots, 1/n)$ for any $\boldsymbol{\varepsilon} \in \mathcal{E}_{\hat{\alpha}, \hat{\varepsilon}}$. By monotonicity in $\alpha \in [\hat{\alpha}, 1)$, it holds that

$$\min_{\mathbf{W} \in \mathcal{T}_{\hat{\alpha}, \hat{\varepsilon}}} \max_{\mathbf{p} \in \mathcal{D}^n} I(\mathbf{p}; \mathbf{W}) = \log n - H(\boldsymbol{\varepsilon}_{\hat{\alpha}}), \quad (56)$$

which determines the game value. The equilibrium strategies are the uniform distribution for the transmitter and the extreme error vector $\boldsymbol{\varepsilon}_{\hat{\alpha}}$ for nature.

The n -ary symmetric channel with error probabilities

$$\left(1 - \delta, \frac{\delta}{n-1}, \dots, \frac{\delta}{n-1}\right) \quad (57)$$

is a special case of the aforementioned by identifying $\hat{\mathbf{e}} = (1, 0, \dots, 0)$ and $\alpha = 1 - (n/(n-1))\delta$.

The binary symmetric channel (BSC) with error probability $0 < \delta < 1/2$ is obtained by setting $n = 2$, $\hat{\mathbf{e}} = (1, 0)$ and $\alpha = 1 - 2\delta$.

6. Entropy of Mixture Distributions

Let \mathbf{U} be an absolutely continuous random variable with density $g(\mathbf{y})$ with respect to the Lebesgue measure λ^n , and let random variable \mathbf{V} have a discrete distribution with discrete density $h(\mathbf{y}) = p_i$, if $\mathbf{y} = \mathbf{x}_i$, $i = 1, \dots, m$, and $h(\mathbf{y}) = 0$ otherwise, $p_i \geq 0$, $\sum_{i=1}^m p_i = 1$. Furthermore, assume that B is Bernoulli distributed with parameter α , $0 \leq \alpha \leq 1$, hence $P(B = 1) = \alpha$, $P(B = 0) = 1 - \alpha$. Further, let \mathbf{U} , \mathbf{V} , B be stochastically independent, then

$$\mathbf{W} = B\mathbf{U} + (1 - B)\mathbf{V} \quad (58)$$

has density

$$f(\mathbf{y}) = \alpha g(\mathbf{y}) + (1 - \alpha)h(\mathbf{y}) \quad (59)$$

with respect to the measure $\mu = \lambda^n + \chi$, where χ denotes the counting measure with support $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$. According to [10], the entropy of \mathbf{W} is defined as

$$H(\mathbf{W}) = - \int f(\mathbf{y}) \log f(\mathbf{y}) d\mu(\mathbf{y}). \quad (60)$$

It easily follows (see [16]) that

$$\begin{aligned} H(\mathbf{W}) &= -\alpha \int g(\mathbf{y}) \log g(\mathbf{y}) d\mathbf{y} - \alpha \log \alpha \\ &\quad - (1 - \alpha) \sum_{i=1}^m p_i \log p_i - (1 - \alpha) \log(1 - \alpha) \\ &= H(B) + \alpha H(\mathbf{U}) + (1 - \alpha)H(\mathbf{V}). \end{aligned} \quad (61)$$

The following proposition will be useful when investigating equilibria of channel games with continuous noise densities.

Proposition 6. Let $\mathbf{p} = (p_1, \dots, p_m)$ denote some stochastic vector, and g_1, \dots, g_m be densities with respect to some measure μ . It holds that

$$H\left(\sum_{i=1}^m p_i g_i\right) - \sum_{i=1}^m p_i H(g_i) \leq H(\mathbf{p}). \quad (62)$$

The proof is provided by the following chain of equalities and inequalities. The argument \mathbf{y} of g_i is omitted for reasons of brevity:

$$\begin{aligned} & - \int \left(\sum_i p_i g_i\right) \log \left(\sum_j p_j g_j\right) d\mu + \sum_i p_i \int g_i \log g_i d\mu \\ &= - \sum_i p_i \int g_i \left(\log \left(\sum_j p_j g_j\right) - \log g_i\right) d\mu \\ &= \sum_i p_i \int g_i \log \frac{g_i}{\sum_j p_j g_j} d\mu \\ &\leq \sum_i p_i \int g_i \log \frac{g_i}{p_i g_i} d\mu \\ &= - \sum_i p_i \log p_i = H(\mathbf{p}). \end{aligned} \quad (63)$$

7. A Mixed Discrete-Continuous Channel Game

Let g_1, \dots, g_m be given λ^n -densities. Distribution $\mathbf{p}^* = (p_1^*, \dots, p_m^*)$ achieves capacity, that is, maximizes mutual information if and only if $I(\mathbf{X}; \mathbf{Y})$ is maximized by \mathbf{p}^* in the set of all stochastic vectors. By representation (9), we need to solve

$$\begin{aligned} & \text{maximize } \left\{ - \int \left(\sum_{i=1}^m p_i g_i(\mathbf{y})\right) \log \left(\sum_{i=1}^m p_i g_i(\mathbf{y})\right) d\mathbf{y} \right. \\ & \quad \left. + \sum_{i=1}^m p_i \int g_i(\mathbf{y}) \log g_i(\mathbf{y}) d\mathbf{y} \right\} \end{aligned} \quad (64)$$

subject to $p_i \geq 0$, $i = 1, \dots, m$,

$$\sum_{i=1}^m p_i = 1.$$

The aforementioned is a convex problem since by Proposition 2, the objective function is concave and the constraint set is convex. The Lagrangian is given by

$$\begin{aligned} L(\mathbf{p}, \boldsymbol{\mu}, \nu) &= - \int \left(\sum_{i=1}^m p_i g_i(\mathbf{y})\right) \log \left(\sum_{i=1}^m p_i g_i(\mathbf{y})\right) d\mathbf{y} \\ &\quad - \sum_{i=1}^m p_i \int g_i(\mathbf{y}) \log g_i(\mathbf{y}) d\mathbf{y} \\ &\quad + \sum_{i=1}^m \mu_i p_i + \nu \left(\sum_{i=1}^m p_i - 1\right), \end{aligned} \quad (65)$$

with the notation $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)$. The optimality conditions are (cf. [12, Chapter 5.5.3])

$$\begin{aligned} \frac{\partial L(\mathbf{p}, \boldsymbol{\mu}, \nu)}{\partial p_i} &= 0, \\ p_i \mu_i &\geq 0, \\ \mu_i p_i &= 0, \end{aligned} \quad (66)$$

for all $i = 1, \dots, m$. Partial derivatives of the Lagrangian with respect to p_i are easily obtained as

$$\begin{aligned} \frac{\partial L(\mathbf{p}, \boldsymbol{\mu}, \nu)}{\partial p_i} &= -(\log e) - \int g_i(\mathbf{y}) \log \left(\sum_{j=1}^m p_j g_j(\mathbf{y}) \right) d\mathbf{y} \\ &+ \int g_i(\mathbf{y}) \log g_i(\mathbf{y}) d\mathbf{y} + \mu_i + \nu, \end{aligned} \quad (67)$$

for $i = 1, \dots, m$. Hence (66) leads to the conditions $p_i = 0$ or

$$\int g_i(\mathbf{y}) \left(\log g_i(\mathbf{y}) - \log \left(\sum_{j=1}^m p_j g_j(\mathbf{y}) \right) \right) d\mathbf{y} = \log e - \nu, \quad (68)$$

for all $i = 1, \dots, m$. In summary, we have demonstrated the following result.

Proposition 7. *Let g_1, \dots, g_m be Lebesgue λ^n -densities. Input distribution \mathbf{p}^* is capacity-achieving if and only if*

$$D\left(g_i \parallel \sum_{j=1}^m p_j^* g_j\right) = \zeta, \quad (69)$$

for some $\zeta > 0$, for all i such that $p_i^* > 0$. Furthermore, if $H(g_i)$ is independent of i , then \mathbf{p}^* is capacity-achieving if and only if

$$\int g_i(\mathbf{y}) \log \left(\sum_{j=1}^m p_j^* g_j(\mathbf{y}) \right) d\mathbf{y} = \xi \quad (70)$$

for some $\xi \in \mathbb{R}$, for all i such that $p_i^* > 0$.

Now, assume that the strategy set of the channel consists of the densities

$$\begin{aligned} \mathcal{F} &= \{ (f_1^{(\alpha)}(\mathbf{y}), \dots, f_m^{(\alpha)}(\mathbf{y})) \mid \\ &f_i^{(\alpha)}(\mathbf{y}) = \alpha g_i(\mathbf{y}) + (1 - \alpha) h_i(\mathbf{y}), \quad 0 \leq \alpha \leq 1 \}, \end{aligned} \quad (71)$$

where g_i are densities with respect to λ^n and h_i represents the singleton distribution with support point \mathbf{x}_i . $f_i^{(\alpha)}$ are hence densities with respect to the measure $\lambda^n + \chi$.

\mathcal{F} represents a closed convex line segment in the space of all densities, reaching from error distribution (g_1, \dots, g_m) at $\alpha = 1$ to the error-free singleton distribution (h_1, \dots, h_m) at $\alpha = 0$. The strategy set is analogous to the m -ary discrete output case with directional separation in Section 5.2. In Figure 5, the mixture of a standard Gaussian and the singleton distribution in 0 is depicted for $\alpha \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$. Densities are with respect to the measure $\lambda^1 + \chi(0)$.

Intuitively, it seems to be clear that the channel would choose the extreme value $\alpha = 1$ as the worst case to jam the transmitter. A precise proof of this fact, however, is amazingly complicated.

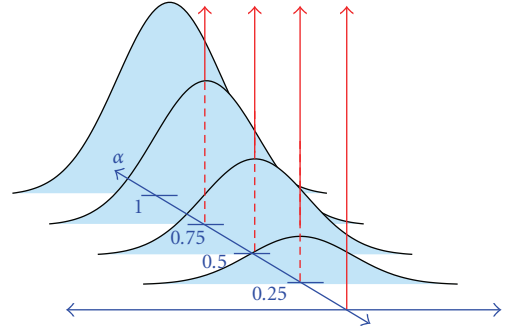


FIGURE 5: Mixture density of a standard Gaussian and a singleton distribution in 0. Densities are with respect to the measure $\lambda^1 + \chi$.

By (61), mutual information is given as

$$\begin{aligned} I(\mathbf{X}; \mathbf{Y}) &= I(\mathbf{p}; (f_1^{(\alpha)}, \dots, f_m^{(\alpha)})) \\ &= H\left(\sum_{i=1}^m p_i f_i^{(\alpha)}\right) - \sum_{i=1}^m p_i H(f_i^{(\alpha)}) \\ &= H\left(\alpha \sum_{i=1}^m p_i g_i + (1 - \alpha) \sum_{i=1}^m p_i h_i\right) \\ &\quad - \sum_{i=1}^m p_i H(\alpha g_i + (1 - \alpha) h_i) \\ &= H(\alpha, 1 - \alpha) + \alpha H\left(\sum_{i=1}^m p_i g_i\right) + (1 - \alpha) H(\mathbf{p}) \\ &\quad - \sum_{i=1}^m p_i (H(\alpha, 1 - \alpha) + \alpha H(g_i)) \\ &= \alpha \left\{ H\left(\sum_{i=1}^m p_i g_i\right) - \sum_{i=1}^m p_i H(g_i) - H(\mathbf{p}) \right\} + H(\mathbf{p}). \end{aligned} \quad (72)$$

Since by Proposition 6, the term in curly brackets in the last line of (72) is nonpositive, for any \mathbf{p} , the minimum of $I(\mathbf{p}; (f_1^{(\alpha)}, \dots, f_m^{(\alpha)}))$ over $\alpha \in [0, 1]$ is attained at $\alpha = 1$ with value

$$H\left(\sum_{i=1}^m p_i g_i\right) - \sum_{i=1}^m p_i H(g_i) = \sum_{i=1}^m p_i D\left(g_i \parallel \sum_{j=1}^m p_j g_j\right). \quad (73)$$

From Proposition 7, it follows that the right-hand side is maximized at $\mathbf{p}^* \in \mathcal{D}^m$ whenever

$$D\left(g_i \parallel \sum_{j=1}^m p_j^* g_j\right) = \zeta \quad (74)$$

for all i with $p_i^* > 0$.

In summary, the channel game has an equilibrium point

$$\begin{aligned} & \max_{\mathbf{p} \in \mathcal{D}^m} \min_{(f_1^{(\alpha)}, \dots, f_m^{(\alpha)}) \in \mathcal{F}} I(\mathbf{p}; (f_1^{(\alpha)}, \dots, f_m^{(\alpha)})) \\ & = \min_{(f_1^{(\alpha)}, \dots, f_m^{(\alpha)}) \in \mathcal{F}} \max_{\mathbf{p} \in \mathcal{D}^m} I(\mathbf{p}; (f_1^{(\alpha)}, \dots, f_m^{(\alpha)})). \end{aligned} \quad (75)$$

The equilibrium strategy for the channel is given by $\alpha = 1$. The optimum strategy \mathbf{p}^* for the transmitter is characterized by (74). For certain error distributions g_j this condition can be explicitly evaluated (see [17]).

8. Conclusions

We have investigated Nash equilibria for a two-person zero-sum game where the channel gambles against the transmitter. The transmitter strategy set consists of all input distributions over a finite symbol set, while the channel strategy sets are formed by certain convex subsets of channel matrices or noise distributions, respectively. Mutual information is used as payoff function.

Basically, it is assumed that a malicious nature is controlling the channel such that equilibria are achieved when the transmitter plays the capacity-achieving distribution against worst-case attributes of the channel. In practice, however, a wireless channel is only partially controlled by nature, for example, by shadowing and attenuation effects, further, diffraction and reflection. A major contribution to the channel properties, however, is made by interference from other users. It will be a subject of future research to investigate how these effects may be combined in a single strategy set of the channel. The question arises if equilibria for the game “one transmitter against a group of others plus random effects from nature” still exist.

Acknowledgments

Part of the material in this paper was presented at IEEE ISIT 2008, Toronto. This work was partially supported by the UMIC Research Center at RWTH Aachen University.

References

- [1] A. Leshem and E. Zehavi, “Bargaining over the interference channel,” in *Proceedings of IEEE International Symposium on Information Theory (ISIT '06)*, pp. 2225–2229, Seattle, Wash, USA, July 2006.
- [2] S. Mathur, L. Sankaranarayanan, and N. B. Mandayam, “Coalitional games in Gaussian interference channels,” in *Proceedings of IEEE International Symposium on Information Theory (ISIT '06)*, pp. 2210–2214, Seattle, Wash, USA, July 2006.
- [3] J. M. Borden, D. M. Mason, and R. J. McEliece, “Some information theoretic saddlepoints,” *SIAM Journal on Control and Optimization*, vol. 23, no. 1, pp. 129–143, 1985.
- [4] S. N. Diggavi and T. M. Cover, “The worst additive noise under a covariance constraint,” *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 3072–3081, 2001.
- [5] E. A. Jorswieck and H. Boche, “Performance analysis of capacity of MIMO systems under multiuser interference based on worst-case noise behavior,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2004, no. 2, pp. 273–285, 2004.
- [6] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, “Uniform power allocation in MIMO channels: a game-theoretic approach,” *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1707–1727, 2003.
- [7] A. Feiten and R. Mathar, “Minimax problems and directional derivatives for MIMO channels,” in *Proceedings of the 63rd IEEE Vehicular Technology Conference (VTC '06)*, vol. 5, pp. 2231–2235, Melbourne, Australia, May 2006.
- [8] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York, NY, USA, 1979.
- [9] I. Csiszar and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, London, UK, 1981.
- [10] M. Pinsker, *Information and Information Stability of Random Variables and Processes*, Holden-Day, San Francisco, Calif, USA, 1964.
- [11] A. W. Roberts and D. E. Varberg, *Convex Functions*, Academic Press, New York, NY, USA, 1973.
- [12] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.
- [13] R. A. Silverman, “On binary channels and their cascades,” *IRE Transactions on Information Theory*, vol. 1, no. 3, pp. 19–27, 1955.
- [14] D. J. C. MacKay, *Information Theory, Inference, and Learning Theory*, Cambridge University Press, Cambridge, UK, 2003.
- [15] R. W. Yeung, *A First Course in Information Theory*, Kluwer Academic Publishers/Plenum Press, New York, NY, USA, 1st edition, 2002.
- [16] D. N. Politis, “Maximum entropy modelling of mixture distributions,” *Kybernetes*, vol. 23, no. 1, pp. 49–54, 1994.
- [17] A. Feiten and R. Mathar, “Capacity-achieving discrete signaling over additive noise channels,” in *Proceedings of IEEE International Conference on Communications (ICC '07)*, pp. 5401–5405, Glasgow, UK, June 2007.