Game Equilibria for Discrete Channels

Rudolf Mathar
Inst. for Theoretical Information Technology
RWTH Aachen University
D-52056 Aachen, Germany
mathar@ti.rwth-aachen.de

Anke Schmeink
Inst. for Theoretical Information Technology
RWTH Aachen University
D-52056 Aachen, Germany
anke.schmeink@rwth-aachen.de

***Abstract*— In this paper, the saddle point behavior of mutual information is investigated for discrete channel models. We use the fact that mutual information is a convex function of the channel matrix, and a concave function of the input distribution. Interpreting transmission as a game, nature against the transmitter with payoff given by mutual information, equilibria are shown to exist for certain strategy sets of nature. The case that nature makes the channel useless with zero capacity is discussed in detail. If nature uses a singleton nonzero capacity strategy, a characterization of the capacity-achieving input distribution is derived. Relevant channel classes covered by this approach include the binary asymmetric and erasure channel with bounded error probabilities. Furthermore, for the symmetric n-symbol channel two classes of separation constraints are introduced and the according game equilibria are determined.**

I. Introduction

In recent literature, resource allocation in wireless channels is often considered as a game where players compete for a scarce medium, the capacity provided by some channel. Usually, Nash bargaining solutions are sought for interference games with Gaussian additive noise, cf. [1], [2]. Different fairness and allocation criteria arise from this paradigm leading to useful control policies for wireless networks.

Other approaches include the channel itself as a player. The transmitter then gambles against a malicious nature. 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 power constraints $\mathrm{E}(X^2) \le P$ and $\mathrm{E}(Z^2) \le \sigma^2$. By standard arguments from information theory it follows that

$$\begin{aligned} &\max_{X:\mathrm{E}(X^2)\le P} \; \min_{Z:\mathrm{E}(Z^2)\le\sigma^2} I(X;X+Z) \\ =&\min_{Z:\mathrm{E}(Z^2)\le\sigma^2} \; \max_{X:\mathrm{E}(X^2)\le P} I(X;X+Z) \\ =&\frac{1}{2}\log\big(1+\frac{P}{\sigma^2}\big) \end{aligned} \tag{1}$$

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 its 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 is considered in [4]. The authors [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 types of games are considered, e.g., in [6], [7].

In this paper, we consider the channel as a player, the malicious nature. Nature gambles against the transmitter which conveys information across the channel. Here, the channel is characterized by error probabilities of symbol transmission with symbols from a finite set. Hence, we confine ourself to discrete channel models and investigate the question whether there are equilibrium points.

The contributions of this paper are as follows. We use the fact that mutual information is a convex function of the channel matrix and a concave function of the input distribution. If nature is free in choosing any strategy it will make the channel useless by selecting a channel matrix with equal rows. This is exactly the zero-capacity case, as is demonstrated by help of the variational distance. A characterization of the capacity-achieving distribution is derived if nature plays a singleton strategy. For general strategy sets with bounded error probabilities relevant examples which have equilibria include the binary asymmetric and erasure channel. Further, the n-symbol symmetric channel is investigated in detail. Using the majorization concept, strategy sets are defined which describe possible channel states in a realistic way excluding the useless channel.

II. Preliminaries

Denote the set of stochastic vectors by

$$\mathcal{D}^m = \big\{\boldsymbol{p} = (p_1,\dots,p_m) \mid p_i \ge 0, \sum_{i=1}^m p_i = 1\big\}.$$

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

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

If $\boldsymbol{p}$ characterizes the distribution of some discrete random variable X we synonymously write $H(X) = H(\boldsymbol{p})$.

978-1-4244-2571-6/08/$25.00 ©2008 IEEE

It is well known that the entropy H is a concave function of $\boldsymbol{p}$, furthermore, since it is symmetric, even Schur-concave over the set of distributions $\mathcal{D}^m$.

We briefly recall the corresponding definitions, see [8]. Let $p_{[i]}$ and $q_{[i]}$ denote the components of $\boldsymbol{p}$ and $\boldsymbol{q}$ in decreasing order, respectively. Distribution $\boldsymbol{p} \in \mathcal{S}$ is said to be majorized by $\boldsymbol{q} \in \mathcal{S}$, in symbols $\boldsymbol{p} \prec \boldsymbol{q}$, if $\sum_{i=1}^{k} p_{[i]} \le \sum_{i=1}^{k} q_{[i]}$ for all $k = 1, \ldots, m$. A function $f : \mathcal{D}^m \to \mathbb{R}$ is called Schur-concave, if $f(\boldsymbol{p}) \ge f(\boldsymbol{q})$ whenever $\boldsymbol{p} \prec \boldsymbol{q}$.

In the following we deal with discrete channels. The input is described by a discrete random variable X with values from a finite alphabet of m symbols and input distribution $\boldsymbol{p}$, the output denoted by random variable Y ranging over an alphabet of size n. The channel matrix comprising the conditional probabilities $w_{ij} = P(Y = j \mid X = i)$ is denoted by

$$\boldsymbol{W} = \big(w_{ij}\big)_{1 \le i \le m, 1 \le j \le n}.$$

Matrix $\boldsymbol{W}$ is an element of the set of stochastic matrices, denoted by $\mathcal{S}^{m \times n}$, its rows are stochastic vectors, denoted by $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m \in \mathcal{D}^m$. The distribution of Y is then given by the stochastic vector $\boldsymbol{q} = \boldsymbol{p}\boldsymbol{W}$.

Mutual information for this channel model reads as

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y \mid X) \\ &= H(\boldsymbol{p}\boldsymbol{W}) - \sum_{i=1}^{m} p_i H(\boldsymbol{w}_i) \\ &= \sum_{i=1}^{m} p_i D(\boldsymbol{w}_i \,\|\, \boldsymbol{p}\boldsymbol{W}) \end{aligned} \tag{2}$$

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

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

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

Obviously, mutual information depends on $\boldsymbol{p}$, controlled by the transmitter, and $\boldsymbol{W}$, controlled by nature. To emphasize this dependence, we also write $I(X;Y) = I(\boldsymbol{p}; \boldsymbol{W})$. We quote from [9, Lemma 3.5].

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

The proof relies on the representation in the third line of (2), convexity of the Kulbach-Leibler divergence $D(\boldsymbol{p} \,\|\, \boldsymbol{q})$ as a function of the pair $(\boldsymbol{p}, \boldsymbol{q})$, and concavity of the entropy H.

III. Channel Games

In the following 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, $I(X;Y)$ 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 2: Let $\mathcal{T} \subseteq \mathcal{S}^{m \times n}$ be a closed convex subset of channel matrices. Then the according channel game has an equilibrium point with value

$$\max_{\boldsymbol{p} \in \mathcal{D}^m} \min_{\boldsymbol{W} \in \mathcal{T}} I(\boldsymbol{p}; \boldsymbol{W}) = \min_{\boldsymbol{W} \in \mathcal{T}} \max_{\boldsymbol{p} \in \mathcal{D}^m} I(\boldsymbol{p}; \boldsymbol{W}). \tag{3}$$

The proof is an immediate consequence of von Neumann's minimax theorem (cf. [10, p. 131]. Since $\mathcal{D}^m$ and $\mathcal{T}$ are compact and convex, the main premises are concavity in $\boldsymbol{p}$ and convexity in $\boldsymbol{W}$, both properties assured by Proposition 1.

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

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{w} \\ \vdots \\ \boldsymbol{w} \end{pmatrix} \tag{4}$$

with constant rows $\boldsymbol{w}$ yielding $I(\boldsymbol{p}; \boldsymbol{W}) = 0$ independent of the input distribution. Obviously, (4) holds if and only if input X and output Y are stochastically independent.

We first consider the case that nature plays a singleton strategy, hence $\mathcal{T} = \{\boldsymbol{W}\}$, a set consisting of exactly one element. Equation (3) then reduces to determining the capacity C of the channel. In order to characterize non-zero capacity channels we use the variational distance between the i-th and j-th row of $\boldsymbol{W}$, defined as

$$d(\boldsymbol{w}_i, \boldsymbol{w}_j) = \sum_{k=1}^{n} |w_{ik} - w_{jk}|.$$

The condition

$$\max_{1 \le i,j \le m} d(\boldsymbol{w}_i, \boldsymbol{w}_j) = \gamma(\boldsymbol{W}) > 0. \tag{5}$$

on the channel matrix $\boldsymbol{W}$ ensures that the according channel has non-zero capacity, as is demonstrated in the following.

Proposition 3: If $\boldsymbol{W}$ satisfies (5) for some $\gamma(\boldsymbol{W}) > 0$, then

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

where information is measured in nats.

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

$$I(\boldsymbol{p}; \boldsymbol{W}) = \frac{1}{2} D\big(\boldsymbol{w}_{i_0} \| \frac{\boldsymbol{w}_{i_0} + \boldsymbol{w}_{j_0}}{2}\big) + \frac{1}{2} D\big(\boldsymbol{w}_{j_0} \| \frac{\boldsymbol{w}_{i_0} + \boldsymbol{w}_{j_0}}{2}\big).$$

Since

$$D(\boldsymbol{w}_i \| \boldsymbol{w}_j) \ge \frac{1}{2 \ln 2} d^2(\boldsymbol{w}_i, \boldsymbol{w}_j),$$

see [9, p. 58], and

$$d\big(\boldsymbol{w}_i, \frac{\boldsymbol{w}_i + \boldsymbol{w}_j}{2}\big) = \frac{1}{2} d(\boldsymbol{w}_i, \boldsymbol{w}_j)$$

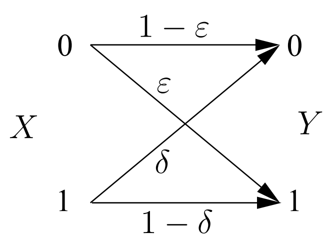

Fig. 1. The binary asymmetric channel.

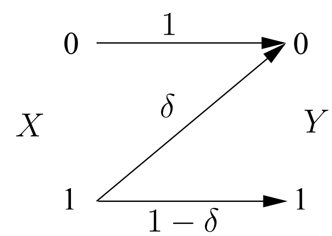

Fig. 2. The Z-channel.

it follows that

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

■

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

Proposition 4: Let channel matrix $\boldsymbol{W}$ satisfy condition (5). Then $C = \max_{\boldsymbol{p} \in \mathcal{D}^m} I(\boldsymbol{p}; \boldsymbol{W})$ is attained at $\boldsymbol{p}^* = (p_1^*, \ldots, p_m^*)$ if and only if

$$D(\boldsymbol{w}_i \| \boldsymbol{p}^* \boldsymbol{W}) = \zeta \tag{6}$$

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

Proof: Mutual information $I(\boldsymbol{p}; \boldsymbol{W})$ is a concave function of $\boldsymbol{p}$. Hence the KKT conditions (cf., e.g., [11]) are necessary and sufficient for optimality of some input distribution $\boldsymbol{p}$. Using representation (2) some elementary algebra shows that

$$\frac{\partial}{\partial p_i} I(\boldsymbol{p}; \boldsymbol{W}) = D(\boldsymbol{w}_i \| \boldsymbol{p} \boldsymbol{W}) - 1.$$

The full set of KKT conditions now reads as

$$\begin{aligned} \boldsymbol{p} &\in \mathcal{D}^m \\ \lambda_i &\geq 0, \ i = 1, \ldots, m \\ \lambda_i p_i &= 0, \ i = 1, \ldots, m \\ D(\boldsymbol{w}_i \| \boldsymbol{p} \boldsymbol{W}) + \lambda_i + \nu &= 0, \ i = 1, \ldots, m \end{aligned}$$

which shows the assertion. ■

Proposition 4 has an interesting interpretation. For an input distribution $\boldsymbol{p}^* = (p_1^*, \ldots, p_m^*)$ to be capacity-achieving, the Kulback-Leibler distance between the rows of $\boldsymbol{W}$ and the weighted average with weights p_i^* has to be the same for all i with positive p_i^*. Hence, capacity-achieving distribution $\boldsymbol{p}^*$ places the mixture distribution $\boldsymbol{p}^* \boldsymbol{W}$ somehow in the middle of all rows $\boldsymbol{w}_i^*$.

A. The Binary Asymmetric Channel

As an example we consider the binary asymmetric channel with channel matrix

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

with $0 < \varepsilon, \delta < 1$ such that condition (5) is satisfied (see Fig. 1). By (6) the capacity-achieving input distribution $\boldsymbol{p} = (p_0, p_1)$ satisfies

$$D(\boldsymbol{w}_1 \| \boldsymbol{p} \boldsymbol{W}) = D(\boldsymbol{w}_2 \| \boldsymbol{p} \boldsymbol{W}).$$

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

$$p_0^* = \frac{1}{1+b}, \quad p_1^* = \frac{b}{1+b}, \tag{7}$$

with

$$b = \frac{a\varepsilon - (1-\varepsilon)}{\delta - a(1-\delta)} \quad \text{and} \quad a = \exp\Big(\frac{h(\delta) - h(\varepsilon)}{1-\varepsilon-\delta}\Big),$$

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 [12].

Now assume that the strategy set of nature is given by

$$\mathcal{T}_{\hat{\varepsilon}, \hat{\delta}} = \big\{\boldsymbol{W}(\varepsilon, \delta) \mid 0 \leq \varepsilon \leq \hat{\varepsilon}, 0 \leq \delta \leq \hat{\delta}\big\}$$

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

Since $I(\boldsymbol{p}; \boldsymbol{W})$ is a convex function of $\boldsymbol{W}$, $I(\boldsymbol{p}; \boldsymbol{W}(\varepsilon, \delta))$ 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 $I(\boldsymbol{p}; \boldsymbol{W}(\varepsilon, \delta))$ 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

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

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

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

is attained at $\boldsymbol{p}^* = (p_0^*, p_1^*)$ from (7) 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 2 that a saddle point exists and the value of the game is given by

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

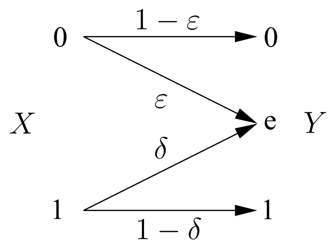

Fig. 3. The binary asymmetric erasure channel.

The so called *Z-channel* with error probability $\varepsilon = 0$ and $\delta \in [0, 1]$ (see Figure 2) is a special case hereof. We have

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

After some algebra, from (7)

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

is obtained with capacity

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

where information is measured in bits, cp. [13, Example 9.11].

B. The Binary Asymmetric Erasure Channel

The *binary asymmetric erasure channel (BEC)* with bit error probabilities $\varepsilon, \delta \in [0, 1]$, and channel matrix

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

is depicted in Figure 3.

According to Proposition 3 this channel has zero capacity if and only if $\varepsilon = \delta = 1$. Excluding this case, by Proposition 4 the capacity-achieving distribution $\boldsymbol{p}^* = (p_0^*, p_1^*)$, $p_0^* + p_1^* = 1$ is given by the solution of

$$\begin{aligned}(1-\varepsilon)\log\frac{1-\varepsilon}{p_0(1-\varepsilon)} &+ \varepsilon\log\frac{\varepsilon}{p_0\varepsilon + p_1\delta} \\ = \delta\log\frac{\delta}{p_0\varepsilon + p_1\delta} &+ (1-\delta)\log\frac{1-\delta}{p_0(1-\delta)},\end{aligned} \tag{8}$$

Substituting $x = \frac{p_0}{p_1}$ equation (8) reads equivalently as

$$\varepsilon\log\varepsilon - \delta\log\delta = (1-\delta)\log(\delta + \varepsilon x) - (1-\varepsilon)\log(\varepsilon + \delta/x)$$

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

If $\varepsilon = \delta$, the solution is given by $p_0^* = p_1^* = \frac{1}{2}$, as is easily verified from equation (8).

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

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

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

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

is attained at $\boldsymbol{p}^* = (p_0^*, p_1^*)$, the solution of (8) with ε substituted by $\hat{\varepsilon}$ and δ by $\hat{\delta}$. Finally, the game value amounts to

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

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

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

and the equilibrium strategies are $p_0^* = p_1^* = \frac{1}{2}$ for the transmitter and $\varepsilon = \delta = \hat{\varepsilon}$ for nature, cf. [14, Example 8.5].

IV. THE n-ARY SYMMETRIC CHANNEL

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

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

by cyclically shifting some error vector $\boldsymbol{\varepsilon} = (\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{n-1}) \in \mathcal{D}^n$. Let $\mathcal{E} \subseteq \mathcal{D}^n$ denote the set of strategies 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 above, nature will cripple the channel by selecting

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_u = (\frac{1}{n}, \ldots, \frac{1}{n}),$$

yielding $I(X; Y) = 0$ independent of the input distribution. Note that $\boldsymbol{\varepsilon}_u$ is the unique minimum element with respect to majorization, i.e., $\boldsymbol{\varepsilon}_u \prec \boldsymbol{\varepsilon}$ for all $\boldsymbol{\varepsilon} \in \mathcal{D}^n$.

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

A. Separation by Schur Ordering

We first investigate the set

$$\begin{aligned}\mathcal{E}_{\succ\hat{\boldsymbol{\varepsilon}}} = \{\boldsymbol{\varepsilon} = (\varepsilon_0, \ldots, \varepsilon_{n-1}) \in \mathcal{D} \mid \\ \hat{\boldsymbol{\varepsilon}} \prec \boldsymbol{\varepsilon}, \varepsilon_{\pi(0)} \leq \cdots \leq \varepsilon_{\pi(n-1)}\}\end{aligned}$$

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

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

$$\mathcal{T}_{\succ\hat{\boldsymbol{\varepsilon}}} = \{\boldsymbol{W}(\boldsymbol{\varepsilon}) \mid \boldsymbol{\varepsilon} \in \mathcal{E}_{\succ\hat{\boldsymbol{\varepsilon}}}\}$$

is convex and closed as well.

Proposition 2 ensures the existence of an equilibrium point.

$$\max_{\boldsymbol{p}\in\mathcal{D}} \min_{\boldsymbol{W}\in\mathcal{T}_{\succ\hat{\boldsymbol{\varepsilon}}}} I(\boldsymbol{p}; \boldsymbol{W}) = \min_{\boldsymbol{W}\in\mathcal{T}_{\succ\hat{\boldsymbol{\varepsilon}}}} \max_{\boldsymbol{p}\in\mathcal{D}} I(\boldsymbol{p}; \boldsymbol{W}).$$

To determine the value v of the game, we first consider $\max_{\boldsymbol{p}\in\mathcal{D}} I(\boldsymbol{p};\boldsymbol{W}(\boldsymbol{\varepsilon}))$ for some fixed $\boldsymbol{\varepsilon} \in \mathcal{E}_{\succ\hat{\boldsymbol{\varepsilon}}}$. From (2) it follows that the maximum is attained at input distribution $\boldsymbol{p} = (\frac{1}{n},\ldots,\frac{1}{n})$ with value

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

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

$$\min_{\boldsymbol{W}\in\mathcal{T}_{\succ\hat{\boldsymbol{\varepsilon}}}} \max_{\boldsymbol{p}\in\mathcal{D}} I(\boldsymbol{p};\boldsymbol{W}) = \log n - H(\hat{\varepsilon})$$

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

B. Directional Separation

Next we consider channel states separated from the worst case $\boldsymbol{\varepsilon}_u$ into the direction of some prespecified $\hat{\boldsymbol{\varepsilon}} \in \mathcal{D}$, $\hat{\boldsymbol{\varepsilon}} \neq \boldsymbol{\varepsilon}_u$. This set of strategies is formally described as

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

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

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

is also closed and convex such that an equilibrium exists by Proposition 2. It remains to determine the game value.

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

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

is attained at $\boldsymbol{W}(\varepsilon_{\hat{\alpha}})$ with $\boldsymbol{\varepsilon}_{\hat{\alpha}} = (1-\hat{\alpha})\boldsymbol{\varepsilon}_0 + \hat{\alpha}\hat{\boldsymbol{\varepsilon}}$. From representation (2) it can be easily seen that

$$\max_{\boldsymbol{p}\in\mathcal{D}} \min_{\boldsymbol{W}\in\mathcal{T}_{\hat{\alpha},\hat{\boldsymbol{\varepsilon}}}} I(\boldsymbol{p};\boldsymbol{W})$$

is attained at $\boldsymbol{p} = (\frac{1}{n},\ldots,\frac{1}{n})$.

Vice versa, from (2) it follows that for any $\boldsymbol{W} = \boldsymbol{W}(\boldsymbol{\varepsilon})$

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

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

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

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

$$(1-\delta, \frac{\delta}{n-1}, \ldots, \frac{\delta}{n-1})$$

is a special case of the above by identifying $\hat{\boldsymbol{\varepsilon}} = (1, 0, \ldots, 0)$ and $\alpha = 1 - \frac{n}{n-1}\delta$.

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

V. Conclusions

In this paper, we have built upon the fact that mutual information is a concave function of the input distribution and a convex function of the channel matrix. Hence, the game with players transmitter against channel has an equilibrium point, provided the channel strategy is selected from a compact convex set. By help of the variational distance we have discussed the case that nature makes the channel useless with zero capacity. If the channel has nonzero capacity, a characterization of the capacity-achieving input distribution against a fixed strategy of nature has been derived. As we have shown, equilibria exist for a series of relevant examples including the asymmetric binary and erasure channel. Moreover, we have investigated symmetric channels over n symbols for two types of separation constraints preventing nature from blocking out transmission.

References

[1] A. Leshem and E. Zehavi, "Bargaining over the interference channel," in *Proceedings ISIT'06*, Seattle, Washington, July 2006, pp. 2225–2229.

[2] S. Mathur, L. Sankaranarayanan, and N. Mandayam, "Coalitional games in Gaussian interfrence channels," in *Proceedings ISIT'06*, Seattle, Washington, July 2006, pp. 2210–2214.

[3] J. Borden, D. Mason, and R. McEliece, "Some information theoretic saddlepoints," *SIAM J. 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, November 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*, 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 VTC 2006 Spring*, Melbourne, May 2006.

[8] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press, 1979.

[9] I. Csiszar and J. Körner, *Information Theory, Coding theorems for discrete memoryless sources*. London: Academic Press, 1981.

[10] A. W. Roberts and D. E. Varberg, *Convex Functions*. New York: Academic Press, 1973.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge University Press, 2004.

[12] R. A. Silverman, "On binary channels and their cascades," *IRE Transactions - Information Theory*, vol. IT-1, pp. 19–27, September 1955.

[13] D. J. C. MacKay, *Information Theory, Inference, and Learning Theory*. Cambridge: Cambridge University Press, 2003.

[14] R. W. Yeung, *A First Course in Information Theory*, 1st ed. Kluwer Academic/Plenum Publishers, 2002.