HEWLETT® PACKARD

On Random Coding Error Exponents of Watermarking Codes

Neri Merhav*
Hewlett-Packard Laboratories Israel
HPL-98-169
October, 1998

E-mail: merhav@hp.technion.ac.il

steganography,
watermarking,
information hiding,
error exponent,
random coding

Watermarking codes are analyzed from an information-theoretic viewpoint as a game between an information hider and an active attacker. While the information hider embeds a secret message (watermark) in a covertext message (typically, text, image, sound, or video stream) within a certain distortion level, the attacker processes the resulting watermarked message, within limited additional distortion, in an attempt to invalidate the watermark. For a memoryless covertext source, we provide a single-letter characterization of the minimax-maximin game of the random coding error exponent associated with the average probability of erroneously decoding the watermark. This single-letter characterization is in effect because there is a "memoryless saddle point" in this game: The information hider utilizes a memoryless channel to generate random codewords for every covertext message, whereas the attacker implements a memoryless channel to disrupt the watermark information hidden in the covertext.

Internal Accession Date Only

*Electrical Engineering Department of the Technion–Israel Institute of Technology, Technion City, Haifa 32000 Israel
Part of this work was done while the author was a consultant at Hewlett-Packard Laboratories, Palo Alto, California

1 Introduction

The fast development of digital information technology, combined with the simplicity of duplication and distribution of digital data, has recently stimulated many research efforts towards the design and study of sophisticated copyright protection and information hiding methodologies (see, e.g., [1],[2],[4] and references therein).

Relatively little attention, however, has been devoted to the problem of information hiding from an information-theoretic perspective. An exception is a recent work by O'Sullivan, Moulin, and Ettinger [7], who characterized the highest achievable information rate of watermarking codes for the following system model: A secret message, encoded at rate R, is hidden in a memoryless covertext message within small degradation of quality, symbolized by distortion level D_1 w.r.t. some distortion measure. An active attacker may introduce additional distortion D_2 in attempt to disrupt the watermark. Finally, the resulting data set is analyzed using information shared with the information hider (i.e., the original covertext message) to extract the watermark. In [7], the highest information rate R has been found (as a function of D_1 and D_2) for which, even in the presence of optimum attack, there exist block encoders and decoders that guarantee arbitrarily small probability of erroneous watermark decoding.

In this paper, we go one step further and provide a single-letter characterization of the best achievable random coding error exponent under this model. Our results are based heavily on a fact, proved in Section 3, that for a memoryless covertext source and a memoryless random coding distribution, an optimum attack strategy subject to the above-mentioned distortion constraint, is to implement a *memoryless channel* on the watermarked data set. It is then clear that this memoryless channel is the one that minimizes the random coding error exponent subject to the single-letter

distortion constraint. We also show in Section 4 the dual property that if the attack strategy is given be a memoryless channel, then the optimal random coding distribution is memoryless as well. It turns out then that there is a minimax–maximin saddle point in the game between the information hider and the attacker, where in the presence of a memoryless covertext source, both parties implement memoryless channels.

2 Notation, Problem Formulation, and Preliminaries

Throughout this paper, we adopt the following notation conventions.

Scalar random variables will be denoted by captial letters (e.g., X) and specific values they may take will be denoted by the corresponding lower case letters (e.g., x). All scalar random variables in this paper are assumed to take on values in the same finite alphabet $\mathcal{A}$. Similarly, random vectors of length n (n – positive integer) will be denoted by boldface capital letters (e.g., $\boldsymbol{X} = (X_1, \ldots, X_n)$), and specific values that they take will be denoted by the respective boldface lower case letters (e.g., $\boldsymbol{x} = (x_1, \ldots, x_n)$), where it is understood that the alphabet of each such vector is always $\mathcal{A}^n$, the nth order Cartezian power of $\mathcal{A}$.

Probability mass functions (PMF's) of single letters will also be denoted by capital letters with probabilities of specific letters denoted by the respective lower case letters (e.g., $Q = \{q(x),\ x \in \mathcal{A}\}$). Similarly, vector sources, or joint PMF's of n-vectors, will be denoted by boldface capital letters with probabilities of specific vector values denoted by the corresponding boldface lower case letters (e.g., $\boldsymbol{Q} = \{\boldsymbol{q}(\boldsymbol{x}),\ \boldsymbol{x} \in \mathcal{A}^n\}$). A source $\boldsymbol{Q}$ over $\mathcal{A}^n$ is said to assume a *product form* if $\boldsymbol{Q}(\boldsymbol{x}) = \prod_{i=1}^n q(x_i)$ for all $\boldsymbol{x} \in \mathcal{A}^n$. For shorthand notation, this fact will be denoted by $\boldsymbol{Q} = Q^n$, where $Q = \{q(x), x \in \mathcal{A}\}$ will be referred to as the *single-letter component* of $\boldsymbol{Q}$.

Similar conventions apply to conditional PMF's of channels: Single-letter conditional PMF's will be denoted by capital letters that symbolize matrices of conditional probabilities (e.g., $W = \{w(y|x),\ x, y \in \mathcal{A}\}$) and vector channels will be denoted by the respective boldface letters (e.g., $\boldsymbol{W} = \{\boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}),\ \boldsymbol{x}, \boldsymbol{y} \in \mathcal{A}^n\}$. A channel $\boldsymbol{W}$ is said to admit a product form if $\boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^n w(y_i|x_i)$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{A}^n$. For shorthand notation, this fact will be denoted by $\boldsymbol{W} = W^n$, where $W = \{w(y|x),\ x, y \in \mathcal{A}\}$ will be referred to as the single-letter component of $\boldsymbol{W}$.

The probability of an event $\mathcal{E}$ will be denoted by $\Pr\{\mathcal{E}\}$ whenever the underlying probability measure is clear from the context, and the expectation operator will be denoted by $\mathbf{E}\{\cdot\}$. Summations and products of indexed terms over the entire index set will be sometimes subscripted only by the symbol of that index. For example, $\sum_x f(x)$ will mean that the summation is over the entire alphabet $\mathcal{A}$, $\prod_i q(x_i)$ means that the product is taken from $i = 1$ to n, etc.

Let $\boldsymbol{P} = P^n$ designate a stationary, memoryless vector source generating random vectors $\boldsymbol{U} = (U_1, \ldots, U_n)$ over $\mathcal{A}^n$. The data vector $\boldsymbol{U}$ will designate the *covertext* message within which the watermark will be hidden. Let $d_1 : \mathcal{A} \times \mathcal{A} \to \mathbb{R}_+$ denote a single-letter distortion measure.

A rate R *watermarking code* of size n, subject to distortion level D_1, is a triple $(\mathcal{M}_n, f_n, g_n)$ with the following ingredients. The set $\mathcal{M}_n$ is the watermarking message set whose size is $\lfloor 2^{nR} \rfloor$. The function $f_n : \mathcal{A}^n \times \mathcal{M}_n \to \mathcal{A}^n$ is the encoder that maps every combination of a covertext data set $\boldsymbol{u} \in \mathcal{A}^n$ and a *watermark message* $m \in \mathcal{M}_n$ into a *stegotext* message (or, watermarked message) $\boldsymbol{x} \in \mathcal{A}^n$ such that $\mathbf{E}d_1(u_i, X_i) \le D_1$ for all $\boldsymbol{u} \in \mathcal{A}^n$, $1 \le i \le n$, that is, the conditional expected distortion in each coordinate given $\boldsymbol{U} = \boldsymbol{u}$ never exceeds D_1 for all $\boldsymbol{u}$.[1] The function $g_n : \mathcal{A}^n \times \mathcal{A}^n \to \mathcal{M}_n$ is the decoder, which maps the original covertext message (shared with the

[1] Although these constraints seem quite restrictive, note that for stationary additive channels and difference distortion measures, $d_1(u, x) = d(x - u)$, they all boil down to a single constraint $\mathbf{E}d(V) \le D_1$, where $V = X - U$ designates the additive noise.

encoder) together with a stegotext message (possibly modified or corrupted by an attacker) back to a watermark message.

Let $d_2 : \mathcal{A} \times \mathcal{A} \to \mathbb{R}_+$ denote a single-letter distortion measure. An *attack* subject to distortion D_2 is a (possibly randomized) map $h_n : \mathcal{A}^n \to \mathcal{A}^n$ that satisfies $\mathbf{E}d_2(x_i, Y_i) \leq D_2$ for every $\boldsymbol{x} \in \mathcal{A}^n$, $1 \leq i \leq n$. We will also assume throughout the paper that $\max\{\max_{u,x} d_1(u,x), \max_{x,y} d_2(x,y)\} \stackrel{\Delta}{=} D_{\max} < \infty$.

Given distortion levels D_1 and D_2 w.r.t. distortion measures d_1 and d_2, respectively, the error probability is defined as

$$P_e(f_n, g_n, h_n) = \Pr\{g_n(\boldsymbol{U}, h_n(f_n(\boldsymbol{U}, M)) \neq M\}, \tag{1}$$

where M designates a random variable in $\mathcal{M}_n$ with a uniform PMF independently of $\boldsymbol{U}$.

An *achievable rate* R is one for which there exists a sequence of rate R watermarking codes $\{(\mathcal{M}_n, f_n, g_n)\}_{n\geq 1}$ subject to distortion level D_1 such that

$$\lim_{n\to\infty} \sup_{h_n} P_e(f_n, g_n, h_n) = 0, \tag{2}$$

where the supremum is over all attack strategies subject to distortion D_2. The capacity of the system, $C = C(D_1, D_2)$, is the supremum of all achievable rates.

Similarly as in [7],[2] it can be shown that for a memoryless covertext source $\boldsymbol{P} = P^n$, the capacity is given by

$$C(D_1, D_2) = \max \min I(X; Y|U) \tag{3}$$

where the maximin is taken over all Markov chains $U \ominus X \ominus Y$ such that: (i) the marginal PMF of U is given by P, (ii) $\mathbf{E}d_1(u, X) \leq D_1$ for all $u \in \mathcal{A}$, and (iii) $\mathbf{E}d_2(x, Y) \leq D_2$ for all $x \in \mathcal{A}$.

[2] The above-described model assumptions are somewhat different from those of [7].

In this paper, we refine the analysis of the probability of error at a given information rate $R < C(D_1, D_2)$. To this end, we assume a random coding regime and we focus on the average probability of error w.r.t. the ensemble of randomly chosen codes given $\boldsymbol{U}$. To meet the D_1 distortion constraint, we adopt the following random coding mechanism: For each $\boldsymbol{u} \in \mathcal{A}^n$, we generate independently, at random 2^{nR} codewords by using a certain channel $\boldsymbol{Q} = \{\boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u}),\ \boldsymbol{u}, \boldsymbol{x} \in \mathcal{A}^n\} \in \mathcal{Q}_n(D_1)$ fed by $\boldsymbol{u}$, where

$$\mathcal{Q}_n(D_1) \stackrel{\Delta}{=} \{\boldsymbol{Q} :\ \sum_{\boldsymbol{x}} \boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u}) d_1(u_i, x_i) \le D_1 \ \ \forall \boldsymbol{u} \in \mathcal{A}^n,\ 1 \le i \le n\}.$$

The channel $\boldsymbol{Q}$ will be referred to as the *watermarking channel* [7]. For $n = 1$, $\mathcal{Q}_n(D_1) = \mathcal{Q}_1(D_1)$, which is a set of single-letter channels Q, will be abbreviated by $\mathcal{Q}(D_1)$.

Since the attacker is allowed to use a randomized map, we will symbolize the attack strategy by a *channel* $\boldsymbol{W} = \{\boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}),\ \boldsymbol{x}, \boldsymbol{y} \in \mathcal{A}^n\}$, henceforth referred to as the *attack channel* [7]. The fact that the attack is subjected to distortion constraint D_2 means that $\boldsymbol{W}$ lies in the set

$$\mathcal{W}_n(D_2) = \{\boldsymbol{W} :\ \sum_{\boldsymbol{y}} \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}) d_2(x_i, y_i) \le D_2 \ \ \forall \boldsymbol{x} \in \mathcal{A}^n,\ 1 \le i \le n\}.$$

For $n = 1$, $\mathcal{W}_n(D_2) = \mathcal{W}_1(D_2)$, which is a set of single-letter channels W, will be abbreviated by $\mathcal{W}(D_2)$.

The average error probability $\bar{P}_e(\boldsymbol{Q}, \boldsymbol{W})$ is defined as the expectation of P_e w.r.t. the joint ensemble of random codes drawn by $\boldsymbol{Q}$ and the randomized attacks governed by $\boldsymbol{W}$.

We will be interested in the quantities

$$\min_{\boldsymbol{Q} \in \mathcal{Q}_n(D_1)} \max_{\boldsymbol{W} \in \mathcal{W}_n(D_2)} \bar{P}_e(\boldsymbol{Q}, \boldsymbol{W})$$

and

$$\max_{\boldsymbol{W} \in \mathcal{W}_n(D_2)} \min_{\boldsymbol{Q} \in \mathcal{Q}_n(D_1)} \bar{P}_e(\boldsymbol{Q}, \boldsymbol{W})$$

as well as on a characterization of pairs $(\boldsymbol{Q}, \boldsymbol{W})$ that achieve these minimax and maximin criteria. However, since an exact expression of $\bar{P}_e(\boldsymbol{Q}, \boldsymbol{W})$ is unavailable, we will focus on a Gallager–type upper bound [6] on $\bar{P}_e(\boldsymbol{Q}, \boldsymbol{W})$.

3 The Worst Attack Channel is Memoryless

In this section, we confine attention to the minimax criterion, corresponding to the assumption that the attacker has perfect knowledge of the watermarking channel, which is assumed here to admit a product form $\boldsymbol{Q} = Q^n$. Generally speaking, a more plausible assumption would be that the attacker knows not only $\boldsymbol{Q}$, but also the specific watermark code in use. Our above assumption on knowing Q but not the code itself is suitable, for example, if the random choice of the code is controlled by a secret key shared by the encoder and decoder, but not the attacker.

The first observation to make is that the problem at hand is no different from an ordinary channel coding problem with side information at both encoder and decoder: A message $m \in \mathcal{M}_n$ of length nR bits is mapped to a codeword $\boldsymbol{x} \in \mathcal{A}^n$ depending on the side information $\boldsymbol{u} \in \mathcal{A}^n$. (In other words, there is a different codebook for every $\boldsymbol{u}$.) Then, $\boldsymbol{x}$ is transmitted across a channel $\boldsymbol{W}$, whose output $\boldsymbol{y}$ is processed in the presence of $\boldsymbol{u}$ to decode the message.

Assuming the random coding regime described in Section 2 and maximum likelihood decoding, the following upper bound on the average error probability is easily derived as a straightforward extension of [6, Theorem 5.6.1, p. 135] to the case where side information is available at both encoder and decoder:

$$\bar{P}_e(\boldsymbol{Q}, \boldsymbol{W}) \leq B(\rho, \boldsymbol{Q}, \boldsymbol{W}) \stackrel{\Delta}{=} 2^{\rho n R} F(\rho, \boldsymbol{Q}, \boldsymbol{W}), \qquad 0 \leq \rho \leq 1 \tag{4}$$

where

$$F(\rho, \boldsymbol{Q}, \boldsymbol{W}) \stackrel{\Delta}{=} \sum_{\boldsymbol{u}} \boldsymbol{p}(\boldsymbol{u}) \sum_{\boldsymbol{y}} \left[\sum_{\boldsymbol{x}} \boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u}) \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x})^{1/(1+\rho)} \right]^{1+\rho} \tag{5}$$

with $\boldsymbol{P} = \{\boldsymbol{p}(\boldsymbol{u}),\ \boldsymbol{u} \in \mathcal{A}^n\}$ being a given memoryless source P^n, $\boldsymbol{Q} \in \mathcal{Q}_n(D_2)$, and $\boldsymbol{W} \in \mathcal{W}_n(D_2)$.

The objective function hereafter will be the upper bound on the average error probability $B(\rho, \boldsymbol{Q}, \boldsymbol{W})$. This function invloves the auxiliary parameter ρ, which can be chosen so as to obtain the tightest upper bound, namely, $\min_{0 \le \rho \le 1} B(\rho, \boldsymbol{Q}, \boldsymbol{W})$. Thus, for a memoryless watermarking channel $\boldsymbol{Q} = Q^n$, we will be interested in characterizing the minimax game

$$\min_{Q \in \mathcal{Q}(D_1)} \boldsymbol{W} \max_{\in \mathcal{W}_n(D_2)} \min_{0 \le \rho \le 1} B(\rho, Q^n, \boldsymbol{W}). \tag{6}$$

For later use, we also define $F(\rho, Q, W)$ and $B(\rho, Q, W)$ as the same functionals as above, applied to the case $n = 1$, that is:

$$B(\rho, Q, W) = 2^{\rho R} F(\rho, Q, W) = 2^{\rho R} \sum_{u} p(u) \sum_{y} \left[\sum_{x} q(x|u) w(y|x)^{1/(1+\rho)} \right]^{1+\rho}. \tag{7}$$

Note that for $\boldsymbol{P} = P^n$, $B(\rho, Q^n, W^n) = B(\rho, Q, W)^n$ and $F(\rho, Q^n, W^n) = F(\rho, Q, W)^n$. We also define the random coding error exponent as

$$E_r(\rho, Q, W) \stackrel{\Delta}{=} -\log B(\rho, Q, W) = -\log F(\rho, Q, W) - \rho R. \tag{8}$$

Our first result states that for a memoryless covertext source and a memoryless random coding distribution, the worst channel $\boldsymbol{W}$, in the sense of maximizing B subject a distortion constraint, is memoryless as well. Since B and F are the same within a factor that does depends on $\boldsymbol{Q}$ and $\boldsymbol{W}$, we can state this result in terms of F.

Theorem 1 *Given ρ, $\boldsymbol{P} = P^n$, Q, and distortion level D_2 w.r.t. distortion measure d_2, it is a stationary memoryless channel $\boldsymbol{W}_* = W_*^n$ that attains the maximum of $F(\rho, Q^n, \boldsymbol{W})$ w.r.t. $\boldsymbol{W}$*

subject to the constraints

$$\begin{aligned} \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}) &\geq 0 \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{A}^n & (9) \\ \sum_{\boldsymbol{y}} \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}) &= 1 \qquad \forall \boldsymbol{x} \in \mathcal{A}^n & (10) \\ \sum_{\boldsymbol{y}} \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}) d_2(x_i, y_i) &\leq D_2, \quad \forall \boldsymbol{x} \in \mathcal{A}^n, 1 \leq i \leq n. & (11) \end{aligned}$$

The single-letter component W_* of $\boldsymbol{W}_*$ is given by the maximizer of $F(\rho, Q, W)$ subject to the same constraints applied to the case $n = 1$, i.e.,

$$\max_{\boldsymbol{W} \in \mathcal{W}_n(D_2)} F(\rho, Q^n, \boldsymbol{W}) = \left[\max_{W \in \mathcal{W}(D_2)} F(\rho, Q, W) \right]^n. \tag{12}$$

The proof is given in Section 5. Theorem 1 leads to our first main result, which is the following.

Theorem 2 *For a memoryless source $\boldsymbol{P} = P^n$,*

$$\min_{Q \in \mathcal{Q}(D_2)} \max_{\boldsymbol{W} \in \mathcal{W}_n(D_2)} \min_{0 \leq \rho \leq 1} B(\rho, Q^n, \boldsymbol{W}) = 2^{-nE(R, D_1, D_2)}, \tag{13}$$

where

$$E(R, D_1, D_2) = \max_{0 \leq \rho \leq 1} \max_{Q \in \mathcal{Q}(D_1)} \min_{W \in \mathcal{W}(D_2)} E_r(\rho, Q, W). \tag{14}$$

Note that the order of the optimizations in eq. (14) is different from the one in eq. (13). The order in (14) has the convenience that for every fixed ρ, the functional $B(\rho, Q, W) = 2^{-E_r(\rho, Q, W)}$ as we shall see, is concave in W and convex in Q.

Proof of Theorem 2. For a given $Q \in \mathcal{Q}(D_2)$, consider the expression

$$\max_{\boldsymbol{W} \in \mathcal{W}_n(D_2)} \min_{0 \leq \rho \leq 1} B(\rho, Q^n, \boldsymbol{W}) = \exp_2 \left\{ \max_{\boldsymbol{W} \in \mathcal{W}_n(D_2)} \min_{0 \leq \rho \leq 1} \left[\log F(\rho, Q^n, \boldsymbol{W}) + \rho n R \right] \right\}. \tag{15}$$

Since $F(\rho, Q^n, \boldsymbol{W})$ is concave in $\boldsymbol{W}$ (see Section 5, Lemma 1) and the logarithmic function is monotonic and concave, then $\log F(\rho, Q^n, \boldsymbol{W})$ is concave as well. Also, a straightforward extension

of [5, p. 17, Proof of Theorem 2] to the case of side information, shows that $\log F(\rho, Q^n, \boldsymbol{W})$, and hence also $[\log F(\rho, Q^n, \boldsymbol{W}) + \rho nR]$ is convex in ρ. Also, $\mathcal{W}_n(D_2)$ and the interval $[0,1]$ are both compact convex sets. Therefore, by the minimax theorem [8, p. 232, Theorem (6.3.7)], the maximization and the minimization on the r.h.s. of eq. (15) are interchangable, and so, it can be written also as

$$\begin{aligned}\exp_2\left\{\min_{0\le\rho\le 1}\max_{\boldsymbol{W}\in\mathcal{W}_n(D_2)}[\log F(\rho,Q^n,\boldsymbol{W})+\rho nR]\right\} &= \min_{0\le\rho\le 1}\left[2^{\rho nR}\max_{\boldsymbol{W}\in\mathcal{W}_n(D_2)}F(\rho,Q^n,\boldsymbol{W})\right]\\ &= \left[\min_{0\le\rho\le 1}2^{\rho R}\max_{W\in\mathcal{W}(D_2)}F(\rho,Q,W)\right]^n \qquad (16)\end{aligned}$$

where the second equality follows from Theorem 1. The assertion of the theorem now follows by taking the minimum over $Q\in\mathcal{Q}(D_2)$. $\square$

4 The Best Watermarking Channel is Memoryless

In this section, we focus on the maximin criterion, which corresponds to a situation where the information hider knows in advance the attack channel and strives at maximizing the error exponent for this channel. In analogy to the previous section, here we assume that the attack channel is memoryless, and so, our objective function will be

$$\max_{W\in\mathcal{W}(D_2)}\min_{\boldsymbol{Q}\in\mathcal{Q}_n(D_1)}\min_{0\le\rho\le 1}B(\rho,\boldsymbol{Q},W^n).$$

We next show that the solution to this problem is dual to the solution of minimax problem of Section 3. As a first step towards this end, we need the following analogue to Theorem 1.

Theorem 3 *Given ρ, $\boldsymbol{P}=P^n$, W, and a distortion level D_1 w.r.t. distortion measure d_1, it is a stationary memoryless channel $\boldsymbol{Q}_*=Q_*^n$ that attains the minimum of $F(\rho,\boldsymbol{Q},W^n)$ w.r.t. $\boldsymbol{Q}$ subject*

to the constraints

$$\begin{aligned} \boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u}) &\geq 0 \qquad \forall \boldsymbol{u}, \boldsymbol{x} \in \mathcal{A}^n & (17) \\ \sum_{\boldsymbol{x}} \boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u}) &= 1 \qquad \forall \boldsymbol{u} \in \mathcal{A}^n & (18) \\ \sum_{\boldsymbol{x}} \boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u}) d_1(u_i, x_i) &\leq D_1, \qquad \forall \boldsymbol{u} \in \mathcal{A}^n,\ 1 \leq i \leq n & (19) \end{aligned}$$

The single-letter component Q_* of $\boldsymbol{Q}_*$ is given by the minimizer of $F(\rho, Q, W)$ subject to the same constraints applied to the case $n = 1$, i.e.,

$$\min_{\boldsymbol{Q} \in \mathcal{Q}_n(D_1)} F(\rho, \boldsymbol{Q}, W^n) = \left[\min_{Q \in \mathcal{Q}(D_1)} F(\rho, Q, W) \right]^n. \tag{20}$$

The proof of Theorem 3 is sketched in Section 6.

Comment: The fact that a memoryless input maximizes the random coding error exponent of a memoryless channel, in the absence of distortion constraints, is well-known as the *parallel channels theorem* due to Gallager [5, Theorem 5],[6, p. 149, Example 4]. Theorem 3 tells us that this continues to be the case in the presence of distortion constraints.

We are now ready to characterize the minimax–maximin game between the information hider and the attacker.

Theorem 4 *For a memoryless source $\boldsymbol{P} = P^n$,*

$$\begin{aligned} \max_{\boldsymbol{W} \in \mathcal{W}_n(D_2)} \min_{\boldsymbol{Q} \in \mathcal{Q}_n(D_1)} \min_{0 \leq \rho \leq 1} B(\rho, \boldsymbol{Q}, \boldsymbol{W}) &= \min_{\boldsymbol{Q} \in \mathcal{Q}_n(D_1)} \max_{\boldsymbol{W} \in \mathcal{W}_n(D_2)} \min_{0 \leq \rho \leq 1} B(\rho, \boldsymbol{Q}, \boldsymbol{W}) \\ &= 2^{-nE(R, D_1, D_2)}, & (21) \end{aligned}$$

where

$$\begin{aligned} E(R, D_1, D_2) &= \max_{0 \leq \rho \leq 1} \max_{Q \in \mathcal{Q}(D_1)} \min_{W \in \mathcal{W}(D_2)} E_r(\rho, Q, W) \\ &= \max_{0 \leq \rho \leq 1} \min_{W \in \mathcal{W}(D_2)} \max_{Q \in \mathcal{Q}(D_1)} E_r(\rho, Q, W). & (22) \end{aligned}$$

Moreover, both the minimax and the maximin of eq. (21) are attained by a saddle point $(\boldsymbol{Q}_*, \boldsymbol{W}_*) = (Q_*^n, W_*^n)$*, where* (Q_*, W_*) *is a saddle point of* $\min_{0\le\rho\le 1} B(\rho, Q, W)$*, or equivalently, a saddle point* (Q_ρ, W_ρ) *of* $B(\rho, Q, W)$ *for the value of* ρ *that minimizes* $B(\rho, Q_\rho, W_\rho)$*.*

Proof. For every fixed ρ, $F(\rho, \boldsymbol{Q}, \boldsymbol{W})$, and hence also $B(\rho, \boldsymbol{Q}, \boldsymbol{W})$, is convex in $\boldsymbol{Q}$ [5, Theorem 4] and concave in $\boldsymbol{W}$ (Section 5, Lemma 1). Also, $\mathcal{Q}_n(D_1)$ and $\mathcal{W}_n(D_2)$ are both compact convex sets. Therefore,

$$\max_{\boldsymbol{W}\in\mathcal{W}_n(D_2)} \min_{\boldsymbol{Q}\in\mathcal{Q}_n(D_1)} B(\rho, \boldsymbol{Q}, \boldsymbol{W}) = \min_{\boldsymbol{Q}\in\mathcal{Q}_n(D_1)} \max_{\boldsymbol{W}\in\mathcal{W}_n(D_2)} B(\rho, \boldsymbol{Q}, \boldsymbol{W}) \tag{23}$$

and hence there exists a saddle point [8, p. 229, Theorem 6.2.9(ii)] $(\boldsymbol{Q}_\rho, \boldsymbol{W}_\rho)$ depending on ρ, i.e.,

$$B(\rho, \boldsymbol{Q}_\rho, \boldsymbol{W}) \le B(\rho, \boldsymbol{Q}_\rho, \boldsymbol{W}_\rho) \le B(\rho, \boldsymbol{Q}, \boldsymbol{W}_\rho) \tag{24}$$

for every $(\boldsymbol{Q}, \boldsymbol{W}) \in \mathcal{Q}_n(D_1) \times \mathcal{W}_n(D_2)$. Taking the minimum over ρ, we get

$$\min_{0\le\rho\le 1} B(\rho, \boldsymbol{Q}_\rho, \boldsymbol{W}) \le \min_{0\le\rho\le 1} B(\rho, \boldsymbol{Q}_\rho, \boldsymbol{W}_\rho) \le \min_{0\le\rho\le 1} B(\rho, \boldsymbol{Q}, \boldsymbol{W}_\rho). \tag{25}$$

Since this is true for every $(\boldsymbol{Q}, \boldsymbol{W}) \in \mathcal{Q}_n(D_1) \times \mathcal{W}_n(D_2)$, one may maximize the l.h.s. over $\boldsymbol{W} \in \mathcal{W}_n(D_2)$ and minimize the r.h.s. over $\boldsymbol{Q} \in \mathcal{Q}_n(D_1)$, which leads to

$$\max_{\boldsymbol{W}\in\mathcal{W}_n(D_2)} \min_{0\le\rho\le 1} B(\rho, \boldsymbol{Q}_\rho, \boldsymbol{W}) \le \min_{0\le\rho\le 1} B(\rho, \boldsymbol{Q}_\rho, \boldsymbol{W}_\rho) \le \min_{\boldsymbol{Q}\in\mathcal{Q}_n(D_1)} \min_{0\le\rho\le 1} B(\rho, \boldsymbol{Q}, \boldsymbol{W}_\rho). \tag{26}$$

The left-most side of eq. (26) is lower bounded by

$$\min_{\boldsymbol{Q}\in\mathcal{Q}_n(D_1)} \max_{\boldsymbol{W}\in\mathcal{W}_n(D_2)} \min_{0\le\rho\le 1} B(\rho, \boldsymbol{Q}, \boldsymbol{W})$$

whereas the right-most side of eq. (26) is upper bounded by

$$\max_{\boldsymbol{W}\in\mathcal{W}_n(D_2)} \min_{\boldsymbol{Q}\in\mathcal{Q}_n(D_1)} \min_{0\le\rho\le 1} B(\rho, \boldsymbol{Q}, \boldsymbol{W}).$$

On the other hand, since minmax is never smaller than maxmin, all the inequalities must be equalities, i.e.,

$$\begin{aligned}\min_{\boldsymbol{Q}\in\mathcal{Q}_n(D_1)}\max_{\boldsymbol{W}\in\mathcal{W}_n(D_2)}\min_{0\le\rho\le 1} B(\rho,\boldsymbol{Q},\boldsymbol{W}) &= \min_{0\le\rho\le 1} B(\rho,\boldsymbol{Q}_\rho,\boldsymbol{W}_\rho)\\ &= \max_{\boldsymbol{W}\in\mathcal{W}_n(D_2)}\min_{\boldsymbol{Q}\in\mathcal{Q}_n(D_1)}\min_{0\le\rho\le 1} B(\rho,\boldsymbol{Q},\boldsymbol{W}),\end{aligned} \tag{27}$$

establishing the first equality in eq. (21). Moreover, eq. (27) tells us that the saddle point $(\boldsymbol{Q}_*,\boldsymbol{W}_*)$ is given by $(\boldsymbol{Q}_{\rho_0},\boldsymbol{W}_{\rho_0})$, where ρ_0 minimizes $B(\rho,\boldsymbol{Q}_\rho,\boldsymbol{W}_\rho)$. Let (Q_ρ,W_ρ) be the saddle point of $B(\rho,Q,W)$ corresponding to $n=1$. By Theorem 1, for every $\boldsymbol{W}\in\mathcal{W}_n(D_2)$,

$$\begin{aligned}B(\rho,Q_\rho^n,\boldsymbol{W}) &\le \max_{\boldsymbol{W}\in\mathcal{W}_n(D_2)} B(\rho,Q_\rho^n,\boldsymbol{W})\\ &= \left[\max_{W\in\mathcal{W}(D_2)} B(\rho,Q_\rho,W)\right]^n\\ &= B(\rho,Q_\rho,W_\rho)^n\\ &= B(\rho,Q_\rho^n,W_\rho^n).\end{aligned} \tag{28}$$

On the other hand, by Theorem 3, for every $\boldsymbol{Q}\in\mathcal{Q}_n(D_1)$,

$$\begin{aligned}B(\rho,\boldsymbol{Q},W_\rho^n) &\ge \min_{\boldsymbol{Q}\in\mathcal{Q}_n(D_1)} B(\rho,\boldsymbol{Q},W_\rho^n)\\ &= \left[\min_{Q\in\mathcal{Q}(D_1)} B(\rho,Q_\rho,W)\right]^n\\ &= B(\rho,Q_\rho,W_\rho)^n\\ &= B(\rho,Q_\rho^n,W_\rho^n).\end{aligned} \tag{29}$$

Thus, (Q_ρ^n,W_ρ^n) is a saddle point of $B(\rho,\boldsymbol{Q},\boldsymbol{W})$ for every ρ. In particular, for $\rho=\rho_0$, eq. (27) tells us that $(Q_{\rho_0}^n,W_{\rho_0}^n)$ achieves both the minimax and the maximin of $\min_{0\le\rho\le 1} B(\rho,\boldsymbol{Q},\boldsymbol{W})$. The

two expressions of $E(R, D_1, D_2)$ (the first one has been shown already in Theorem 1) follow now immediately from eq. (27) applied to the case $n = 1$. $\Box$

5 Proof of Theorem 1

The proof involves three auxiliary results stated below as lemmas. The first important property of F is concavity w.r.t. $\boldsymbol{W}$.

Lemma 1 *The functional $F(\rho, \boldsymbol{Q}, \boldsymbol{W})$ is concave in $\boldsymbol{W}$ for fixed $\boldsymbol{Q}$ and ρ.*

Proof. It would be sufficient to show that for every $(u, y) \in \mathcal{A} \times \mathcal{A}$, each summand of the r.h.s. of eq. (5) is concave in $\boldsymbol{W}$. Specifically, let $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ be two arbitrary channels and let α be an arbitrary number in $(0, 1)$. We wish to show then that

$$\begin{aligned}
& \left\{\sum_{\boldsymbol{x}} q(\boldsymbol{x}|\boldsymbol{u})\left[\alpha \boldsymbol{w}_1(\boldsymbol{y}|\boldsymbol{x}) + (1-\alpha)\boldsymbol{w}_2(\boldsymbol{y}|\boldsymbol{x})\right]^{1/(1+\rho)}\right\}^{1+\rho} \\
\geq\ & \alpha\left[\sum_{\boldsymbol{x}} \boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u})\boldsymbol{w}_1(\boldsymbol{y}|\boldsymbol{x})^{1/(1+\rho)}\right]^{1+\rho} + (1-\alpha)\left[\sum_{\boldsymbol{x}} \boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u})\boldsymbol{w}_2(\boldsymbol{y}|\boldsymbol{x})^{1/(1+\rho)}\right]^{1+\rho} \\
=\ & \left\{\sum_{\boldsymbol{x}} [\alpha \boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u})^{1+\rho}\boldsymbol{w}_1(\boldsymbol{y}|\boldsymbol{x})]^{1/(1+\rho)}\right\}^{1+\rho} + \left\{\sum_{x} [(1-\alpha)\boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u})^{1+\rho}\boldsymbol{w}_2(\boldsymbol{y}|\boldsymbol{x})]^{1/(1+\rho)}\right\}^{1+\rho}. \quad (30)
\end{aligned}$$

Now, Minkowski's inequality (see, e.g., [6, p. 523, Problem 4.15(g)]) states that for a set of non-negative numbers $\{a_{jk}\}$, $1 \leq j \leq J$, $1 \leq k \leq K$, and $r > 1$,

$$\left[\sum_j \left(\sum_k a_{jk}\right)^{1/r}\right]^r \geq \sum_k \left(\sum_j a_{jk}^{1/r}\right)^r. \quad (31)$$

It is easy then to see that eq. (30) is obtained as a special case of Minkowski's inequality with $r = 1 + \rho$, $K = 2$, $j = \boldsymbol{x}$, $a_{j1} = \alpha \boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u})^{1+\rho}\boldsymbol{w}_1(\boldsymbol{y}|\boldsymbol{x})$, and $a_{j2} = (1-\alpha)\boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u})^{1+\rho}\boldsymbol{w}_2(\boldsymbol{y}|\boldsymbol{x})$. $\Box$

Consider now an auxiliary maximization problem, henceforth referred to as *Problem A*, which is defined as follows. Let $\mathcal{N} = \{1, 2, \ldots, n\}$ and let S be a subset of $\mathcal{N}$, where the members are

listed in increasing order, e.g., $\{1,2,5\}$, $\{3,4,100,200\}$, etc. Let $n(S)$ denote the cardinality of S, and let $\mathcal{S}$ denote the collection of all 2^n-1 nonempty subsets of $\mathcal{N}$. Finally, let $\theta > 0$ be a given constant. Then, Problem A is defined as follows.

Problem A:

$$\max_{\boldsymbol{W}} F(\rho, \boldsymbol{Q}, \boldsymbol{W}) \tag{32}$$

subject to

$$\boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}) \geq 0 \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{A}^n \tag{33}$$

$$\sum_{\boldsymbol{y}} \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}) = 1 \qquad \forall \boldsymbol{x} \in \mathcal{A}^n \tag{34}$$

$$\sum_{\boldsymbol{y}} \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}) \exp[\theta \sum_{i \in S} d_2(x_i, y_i)] \leq \exp[\theta n(S) D_2], \quad \forall \boldsymbol{x} \in \mathcal{A}^n,\ S \in \mathcal{S}. \tag{35}$$

The reason for considering Problem A is the fact that it is easily shown to be solved by a product form channel. This will follow from a simple inspection of the Kuhn–Tucker conditions (similarly as in [5], [6, Theorems 4.5.1, 5.6.5]). The original problem of interest will be approached by letting $\theta \to 0$.

Lemma 2 *For a given P, Q, ρ, and D_2, let $W_* = \{w_*(y|x)\}$ maximize $F(\rho, Q, W)$ subject to the constraints*

$$w(y|x) \geq 0 \qquad \forall x, y \in \mathcal{A} \tag{36}$$

$$\sum_{y} w(y|x) = 1 \qquad \forall x \in \mathcal{A} \tag{37}$$

$$\sum_{y} w(y|x) e^{\theta d_2(x,y)} \leq e^{\theta D_2} \qquad \forall x \in \mathcal{A}. \tag{38}$$

Then, for $\boldsymbol{P} = P^n$ and $\boldsymbol{Q} = Q^n$, the channel $\boldsymbol{W} = W_*^n$ solves Problem A.

Proof. First observe that the solution to Problem A will be unaffected if the equality constraint (34) will be replaced by the inequality constraint $\sum_{\boldsymbol{y}} \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}) \leq 1$ for all $\boldsymbol{x}$. This is obvious because for any $\boldsymbol{W}$ with $\sum_{\boldsymbol{y}} \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}) < 1$ for some $\boldsymbol{x}$, the components $\boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x})$ can be scaled by a factor larger than one without violating the constraint, thereby increasing F. Therefore, the maximum of F must be attained for $\sum_{\boldsymbol{y}} \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}) = 1$ for all $\boldsymbol{x}$. Let us refer to this modification[3] of Problem A as Problem A'. Next observe that Problem A' is a convex program (see, e.g., [3, Sect. 4.5]) with inequality constraints only. Therefore, the Kuhn–Tucker conditions are both necessary and sufficient for optimality [3, Theorems 4.38, 4.39]. Specifically, let

$$\begin{aligned} V(\boldsymbol{Q}, \boldsymbol{W}, \boldsymbol{x}, \boldsymbol{y}) &\triangleq \frac{\partial F(\rho, \boldsymbol{Q}, \boldsymbol{W})}{\partial \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x})} \\ &= \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x})^{-\rho/(1+\rho)} \sum_{\boldsymbol{u}} \boldsymbol{p}(\boldsymbol{u}) \boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u}) \left[\sum_{\boldsymbol{x}'} \boldsymbol{q}(\boldsymbol{x}'|\boldsymbol{u}) \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}')^{1/(1+\rho)} \right]^{\rho}. \end{aligned} \tag{39}$$

The Kuhn–Tucker conditions for a channel $\boldsymbol{W}$ to solve Problem A' are that there exist non-negative constants $\{\boldsymbol{\mu}(\boldsymbol{x})\}$ and $\{\boldsymbol{\lambda}(\boldsymbol{x}, S)\}$, $\boldsymbol{x} \in \mathcal{A}^n$, $S \in \mathcal{S}$, such that for every $\boldsymbol{x}$ and $\boldsymbol{y}$,

$$V(\boldsymbol{Q}, \boldsymbol{W}, \boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\mu}(\boldsymbol{x}) + \sum_{S \in \mathcal{S}} \boldsymbol{\lambda}(\boldsymbol{x}, S) \exp[\theta \sum_{i \in S} d_2(x_i, y_i)], \qquad \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}) > 0 \tag{40}$$

$$V(\boldsymbol{Q}, \boldsymbol{W}, \boldsymbol{x}, \boldsymbol{y}) \leq \boldsymbol{\mu}(\boldsymbol{x}) + \sum_{S \in \mathcal{S}} \boldsymbol{\lambda}(\boldsymbol{x}, S) \exp[\theta \sum_{i \in S} d_2(x_i, y_i)], \qquad \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}) = 0. \tag{41}$$

Let $W_* = \{w_*(y|x),\ x, y \in \mathcal{A}\}$ be defined as in Lemma 2. Then, by the necessity of the Kuhn–Tucker conditions for $n = 1$, there must exist non-negative constants $\lambda(x)$, and $\mu(x)$, $x \in \mathcal{A}$, such that

$$V(Q, W_*, x_i, y_i) = \mu(x_i) + \lambda(x_i) \exp[\theta d_2(x_i, y_i)], \qquad w_*(y_i|x_i) > 0 \tag{42}$$

$$V(Q, W_*, x_i, y_i) \leq \mu(x_i) + \lambda(x_i) \exp[\theta d_2(x_i, y_i)], \qquad w_*(y_i|x_i) = 0. \tag{43}$$

[3] This modification guarantees that the Lagrange multiplier corresponding to the equality constraint is non-negative.

Taking the product of $\{V(Q, W_*, x_i, y_i)\}$ and $\{\mu(x_i) + \lambda(x_i)\exp[\theta d_2(x_i, y_i)]\}$ over $1 \le i \le n$, the former gives $V(\boldsymbol{Q}, W_*^n, \boldsymbol{x}, \boldsymbol{y})$, whereas and the latter is

$$\prod_i [\mu(x_i) + \lambda(x_i)\exp[\theta d_2(x_i, y_i)]] \stackrel{\Delta}{=} \boldsymbol{\mu}(\boldsymbol{x}) + \sum_{S\in\mathcal{S}} \boldsymbol{\lambda}(\boldsymbol{x}, S)\exp[\theta \sum_{i\in S} d_2(x_i, y_i)] \tag{44}$$

with

$$\boldsymbol{\mu}(\boldsymbol{x}) = \prod_{i=1}^{n} \mu(x_i) \tag{45}$$

and

$$\boldsymbol{\lambda}(\boldsymbol{x}, S) = \prod_{i\in S} \lambda(x_i) \prod_{i\in S^c} \mu(x_i). \tag{46}$$

Clearly, since $\mu_i(x)$ and $\lambda_i(x)$ are all non-negative, so are $\boldsymbol{\mu}(\boldsymbol{x})$ and $\boldsymbol{\lambda}(\boldsymbol{x}, S)$. In addition, if $w_*(y_i|x_i) > 0$ for all i (and hence also $\boldsymbol{w}_*(\boldsymbol{y}|\boldsymbol{x}) = \prod_i w_*(y_i|x_i) > 0$), then

$$V(\boldsymbol{Q}, \boldsymbol{W}, \boldsymbol{x}, \boldsymbol{y}) = \mu(\boldsymbol{x}) + \sum_S \lambda(\boldsymbol{x}, S)\exp[\theta \sum_{i\in S} d_2(x_i, y_i)],$$

otherwise

$$V(\boldsymbol{Q}, \boldsymbol{W}, \boldsymbol{x}, \boldsymbol{y}) \le \mu(\boldsymbol{x}) + \sum_S \lambda(\boldsymbol{x}, S)\exp[\theta \sum_{i\in S} d_2(x_i, y_i)].$$

The conclusion is, therefore, that the Lagrange multiplier values given in eqs. (45) and (46) satisfy the Kuhn–Tucker conditions together with the vector channel $\boldsymbol{W} = W_*^n$. Thus, by the sufficiency of the Kuhn–Tucker conditions, this product form channel solves Problem A', or equivalently, Problem A. □.

Let us now define Problem B similarly as Problem A, but with the the distortion constraints (35) replaced by

$$\sum_{\boldsymbol{y}} \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}) d_2(x_i, y_i) \le D_2, \qquad \forall \boldsymbol{x} \in \mathcal{A}^n, 1 \le i \le n. \tag{47}$$

In other words, Problem B is the original problem under discussion.

Lemma 3 *Let $\boldsymbol{P} = P^n$, Q, ρ, and D_2 be given, and define the function $F_\theta(D_2)$ as the maximum value of $F(\rho, Q^n, \boldsymbol{W})$ subject to the constraints of Problem A. Let $F_0(D_2)$ be defined as the maximum value of $F(\rho, Q^n, \boldsymbol{W})$ subject to the constraints of Problem B. Then,*

(a) The function $F_0(D_2)$ is concave.

(b) $\lim_{\theta \to 0} F_\theta(D_2) = F_0(D_2)$.

(c) The value $F_0(D_2)$ is attained by a memoryless channel W_*^n, where W^* solves Problem B for $n = 1$ w.r.t. the distortion constraints $\sum_y w(y|x) d_2(x,y) \leq D_2$ for all $x \in \mathcal{A}$.

Proof. Beginning from part (a), let D_2' and D_2'' denote two distortion levels and for $\alpha \in (0,1)$, let $D_2 = \alpha D_2' + (1-\alpha) D_2''$. Let $\boldsymbol{W}_*'$, $\boldsymbol{W}_*''$, and $\boldsymbol{W}_*$ solve Problem B for D_2', D_2'', and D_2, respectively. Now, we have

$$\begin{aligned} \alpha F_0(D_2') + (1-\alpha) F_0(D_2'') &= \alpha F(\rho, \boldsymbol{Q}, \boldsymbol{W}_*') + (1-\alpha) F(\rho, \boldsymbol{Q}, \boldsymbol{W}_*'') \\ &\leq F(\rho, \boldsymbol{Q}, \alpha \boldsymbol{W}_*' + (1-\alpha) \boldsymbol{W}_*'') \\ &\leq F(\rho, \boldsymbol{Q}, \boldsymbol{W}_*) \\ &= F_0(D_2), \end{aligned} \tag{48}$$

where the first inequality follows from the concavity of F (Lemma 1), and the second inequality results from the fact that $\alpha W_*' + (1-\alpha) W_*''$ satisfies the constraints of Problem B with distortion matrix D_2.

To prove (b), observe first that the set of individual coordinate distortion constraints given in eq. (47) is equivalent to the set of constraints

$$\sum_{\boldsymbol{y}} \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}) \left[\sum_{i \in S} d_2(x_i, y_i) \right] \leq n(S) D_2, \qquad \forall \boldsymbol{x} \in \mathcal{A}^n, \ S \in \mathcal{S}. \tag{49}$$

This is true because on the one hand, the subsets $\{1\}, \{2\}, ..., \{n\}$ are all in $\mathcal{S}$, and on the other hand, given the individual coordinate distortion constraints, the remaining distortion constraints are all redundant. Now, for every $(d_1, \ldots, d_n) \in [0, D_{\max}]^n$ and $S \in \mathcal{S}$, $e^{\theta \sum_{i \in S} d_i} \geq 1 + \theta \sum_{i \in S} d_i$, and for $0 < \theta \leq 1$, we also have

$$\begin{aligned} e^{\theta \sum_{i \in S} d_i} &\leq 1 + \theta \sum_{i \in S} d_i + \sum_{k \geq 2} \frac{(\theta n D_{\max})^k}{k!} \\ &= 1 + \theta \sum_{i \in S} d_i + \theta^2 \sum_{k \geq 2} \frac{\theta^{k-2} (n D_{\max})^k}{k!} \\ &\leq 1 + \theta \sum_{i \in S} d + \theta^2 \sum_{k \geq 2} \frac{(n D_{\max})^k}{k!} \\ &= 1 + \theta \sum_{i \in S} d_i + \theta^2 (e^{n D_{\max}} - n D_{\max} - 1) \\ &= 1 + \theta \sum_{i \in S} d_i + C\theta^2, \end{aligned} \tag{50}$$

where $C = e^{nD_{\max}} - nD_{\max} - 1$ is a constant (for a given n) that depends on neither θ nor $\{d_i\}$. Let $\mathcal{R}_\theta(D_2)$ denote the set of channels $\boldsymbol{W}$ that satisfy the constraints of Problem A and let $\mathcal{R}_0(D_2)$ denote the set of channels that satsify the constraints of Problem B. By using the above upper and lower bounds on the exponential function, it is readily seen that for $\theta \in (0, 1]$,

$$\mathcal{R}_0(D_2 - \theta C) \subseteq \mathcal{R}_\theta(D_2) \subseteq \mathcal{R}_0(D_2 + \theta C), \tag{51}$$

and therefore,

$$F_0(D_2 - \theta C) \leq F_\theta(D_2) \leq F_0(D_2 + \theta C). \tag{52}$$

From the concavity of F_0, proven already in part (a) above, it also follows that F_0 is continuous at any point in $(0, D_{\max})$. Thus, for such D_2, $\lim_{\theta \to 0} F_0(D_2 - \theta C) = \lim_{\theta \to 0} F_0(D_2 + \theta C) = F_0(D_2)$, and assertion (b) follows from eq. (52) by a sandwich argument.

As for part (c), let $\{\theta_k\}$ be a positive sequence that tends to zero, and consider the corresponding sequence of channels $\{\boldsymbol{W}_k\}$ that achieve $F_{\theta_k}(D_2)$, $k = 1, 2, \ldots$. According to Lemma 2, $\boldsymbol{W}_k = W_k^n$,

where W_k solves the single-letter version of Problem A (i.e., $\max_W F(\rho, Q, W)$ subject to the constraints of Lemma 2). Since the set of finite–alphabet channels is compact, there is a convergent subsequence $\{W_{k_j}\}_{j\geq 1}$. Let W_* denote the limit of that subsequence. Then, from the continuity of F w.r.t. $\boldsymbol{W}$,

$$\begin{aligned} F(\rho, \boldsymbol{Q}, W_*^n) &= \lim_{j\to\infty} F(\rho, \boldsymbol{Q}, W_{k_j}^n) \\ &= \lim_{\theta\to 0} F_\theta(D_2) \\ &= F_0(D_2), \end{aligned} \tag{53}$$

where the last step follows from part (b). Thus, Problem B is solved by W_*^n. Finally, the fact that W_* solves the single-letter version of Problem B, can be readily seen again from eq. (53), degenerated to the one-dimensional case. □

6 Proof of Theorem 3

The proof is completely analog to that of Theorem 1. Therefore, we give here a brief sketch only. The fact that $F(\rho, \boldsymbol{Q}, \boldsymbol{W})$ is convex w.r.t. $\boldsymbol{Q}$ is well-known [5, Theorem 4]. Next, define Problem A as

$$\min_{\boldsymbol{Q}} F(\rho, \boldsymbol{Q}, W^n) \tag{54}$$

subject to

$$\boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u}) \geq 0 \qquad \forall \boldsymbol{u}, \boldsymbol{x} \in \mathcal{A}^n \tag{55}$$

$$\sum_{\boldsymbol{x}} \boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u}) = 1 \qquad \forall \boldsymbol{u} \in \mathcal{A}^n \tag{56}$$

$$\sum_{\boldsymbol{y}} \boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u}) \exp[-\theta \sum_{i\in S} d_1(u_i, x_i)] \geq \exp[-\theta n(S) D_1], \quad \forall \boldsymbol{u} \in \mathcal{A}^n,\ S \in \mathcal{S}. \tag{57}$$

The first step is to show that Problem A is solved by a product form channel $\boldsymbol{Q} = Q^n$, where Q solves the one-dimensional version of Problem A corresponding to each coordinate. To see this, Problem A is first modified (without affecting the result) to Problem A', defined the same way as Problem A but with the equality constraint replaced by $\sum_{\boldsymbol{x}} \boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u}) \geq 1$ for all $\boldsymbol{u}$. This is again a convex program for which the Kuhn–Tucker (necessary and sufficient) conditions are the existence of non-negative constants $\{\boldsymbol{\mu}(\boldsymbol{u})\}$, $\{\boldsymbol{\lambda}(\boldsymbol{u}, S)\}$, $\boldsymbol{u} \in \mathcal{A}^n$, $S \in \mathcal{S}$, such that

$$U(\boldsymbol{Q}, \boldsymbol{W}, \boldsymbol{u}, \boldsymbol{x}) = \boldsymbol{\mu}(\boldsymbol{u}) + \sum_{S \in \mathcal{S}} \boldsymbol{\lambda}(\boldsymbol{u}, S) \exp[-\theta \sum_{i \in S} d_1(u_i, x_i)], \quad \boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u}) > 0 \tag{58}$$

$$U(\boldsymbol{Q}, \boldsymbol{W}, \boldsymbol{u}, \boldsymbol{x}) \geq \boldsymbol{\mu}(\boldsymbol{u}) + \sum_{S \in \mathcal{S}} \boldsymbol{\lambda}(\boldsymbol{u}, S) \exp[-\theta \sum_{i \in S} d_1(u_i, x_i)], \quad \boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u}) = 0 \tag{59}$$

where

$$U(\boldsymbol{Q}, \boldsymbol{W}, \boldsymbol{u}, \boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{u}) \sum_{\boldsymbol{y}} \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x})^{1/(1+\rho)} \left[\sum_{\boldsymbol{x}'} \boldsymbol{q}(\boldsymbol{x}'|\boldsymbol{u}) \boldsymbol{w}(\boldsymbol{y}|\boldsymbol{x}')^{1/(1+\rho)} \right]^{\rho} \tag{60}$$

is the partial derivative of F w.r.t. $\boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u})$. Similarly as in the proof of Lemma 2, these conditions give rise to the above-defined product form channel.

Next, define Problem B in the same way as Problem A, but with the set of distortion constraints replaced by

$$\sum_{\boldsymbol{x}} \boldsymbol{q}(\boldsymbol{x}|\boldsymbol{u}) d_1(u_i, x_i) \leq D_1, \qquad \forall \boldsymbol{u} \in \mathcal{A}^n, 1 \leq i \leq n. \tag{61}$$

Finally, an analogue to Lemma 3 (with the concavity property in part (a) replaced by a convexity property) is then proved in the same way, where the bounds on the exponential function to be used are $e^{-\theta \sum_{i \in S} d_i} \geq 1 - \theta \sum_{i \in S} d_i$ and $e^{-\theta \sum_{i \in S} d_i} \leq 1 - \theta \sum_{i \in S} d_i + \frac{1}{2}\theta^2 (n D_{\max})^2$.

Acknowledgement

Useful discussions with Ronny Roth are greatly appreciated.

References

[1] R. Anderson, "Stretching the limits of steganography," in *Information Hiding*, Springer Lecture Notes in Computer Science vol. 1174, pp. 39–48, 1996. Also, available at `http://www.bluespike.com/research/research.html`.

[2] R. J. Anderson and A. Petitcolas, "On the limits of steganography," *IEEE J. Select. Areas in Communications*, vol. 16, no. 4, pp. 463-473, May 1998.

[3] M. Avriel, *Nonlinear Programming Analysis and Methods.* Prentice-Hall, 1976.

[4] S. Craver, N. Memon, B.-L. Yeo, and M. M. Yeung, "Resolving rightful ownerships with invisible watermarking techniques: limitations, attacks, and implications," *IEEE J. Select. Areas in Commun.,* vol. 16, no. 4, pp. 573–586, May 1998.

[5] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. IT–11, no. 1, pp. 3–18, 1965.

[6] R. G. Gallager, *Information Theory and Reliable Communication*, J. Wiley & Sons, 1968.

[7] J. A. O'Sullivan, P. Moulin, and J. M. Ettinger, "Information theoretic analysis of Steganography," *Proc. ISIT '98*, p. 297, 1998.

[8] J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimensions I*, Springer–Verlag, New York, 1970.