On Exponentially Concave Functions and Their Impact in Information Theory

Gholamreza Alirezaei and Rudolf Mathar
Institute for Theoretical Information Technology
RWTH Aachen University, D-52056 Aachen, Germany
{alirezaei, mathar}@ti.rwth-aachen.de

***Abstract*—Concave functions play a central role in optimization. So-called exponentially concave functions are of similar importance in information theory. In this paper, we comprehensively discuss mathematical properties of the class of exponentially concave functions, like closedness under linear and convex combination and relations to quasi-, Jensen- and Schur-concavity. Information theoretic quantities such as self-information and (scaled) entropy are shown to be exponentially concave. Furthermore, new inequalities for the Kullback-Leibler divergence, for the entropy of mixture distributions, and for mutual information are derived.**

***Index Terms*—Inequalities, convex optimization, Schur-concave, quasiconcave, Shannon theory, self-information, entropy, mutual information, divergence, mixture distribution.**

I. Introduction

More than 100 years ago, Johan Jensen published his seminal paper [1] in which the fundamentals of convexity are investigated. Nowadays, convex and concave functions are central components in both optimization theory and set theory. Furthermore, the concept of convexity and concavity of functions is extended in various directions to tackle the needs of physicists, engineers and mathematicians. For example, the class of *logarithmically convex* functions provides more accurate bounds and inequalities than the ones derived from convex functions, see [2], [3]. Moreover, convexity and concavity allow for very elegant proofs in inequality theory, cf. [4].

In contrast to logarithmically convex (log-convex) functions, their counterpart, the so-called *exponentially concave* (exp-concave) functions, are rarely discussed in the literature, probably due to their intricate structure. Exponentially concave functions play an important role in information theory, as we will see later. For instance, the (scaled) discrete entropy, as a convex combination of logarithms, is exponentially concave. It is exactly this type of combination that makes the mathematical investigation of exponentially concave functions cumbersome. Besides the application in information theory, a thorough investigation of exponentially concave functions is of general mathematical interest.

The growth of research on big data analytics and deep learning has recently increased the interest in exponentially concave functions. In [5], the smoothness of exponentially concave functions is exploited for statistical learning, sequential prediction and stochastic optimization, which are important topics in machine learning. Lower and upper bounds on solutions of stochastic exponentially concave optimization problems are discussed in [6]. Empirical risk minimization, which is a general optimization framework that captures several important learning problems including linear and logistic regression, learning support vector machines (SVMs) with the squared hinge-loss, and portfolio selection, see [7], is investigated for exponentially concave loss functions in [8]. To the best of our knowledge, a comprehensive investigation of exponentially concave functions as in the present paper is new.

In the present paper, we present basic properties of exponentially concave functions and inequalities. Afterwards, information theoretic quantities are considered. We especially show that both self-information and the (scaled) discrete entropy are exponentially concave functions. Finally, these results are applied to deduce a new lower bound on the Kullback-Leibler divergence, new entropy inequalities for mixtures of distributions, and inequalities for the mutual information of discrete channels.

II. Preliminaries

Let[1] $\mathcal{D}$ be a convex[2] subset[3] of $\mathcal{R}^n$. A function $g : \mathcal{D} \to \mathcal{R}$ is called *concave*, if

$$\beta g(\boldsymbol{x}) + \bar{\beta} g(\boldsymbol{y}) \leq g(\beta \boldsymbol{x} + \bar{\beta} \boldsymbol{y}) \tag{1}$$

holds for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$ and for all $\beta \in [0,1]$ with $\bar{\beta} = 1 - \beta$. Function g is called *strictly* concave, if strict inequality holds in (1) for all $\boldsymbol{x} \neq \boldsymbol{y}$ and for all $\beta \in (0,1)$. A function is called (strictly) *convex*, if its negative is (strictly) concave, cf. [2]. Note that convex and concave functions are necessarily continuous in finite spaces. If in addition g is differentiable, then the inequality

$$g(\boldsymbol{y}) - g(\boldsymbol{x}) \leq \nabla g(\boldsymbol{x})^{\mathrm{T}} (\boldsymbol{y} - \boldsymbol{x}) \tag{2}$$

holds for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$. For a twice differentiable function g, an equivalent definition of concavity is given by

$$\boldsymbol{v}^{\mathrm{T}} \nabla^2 g(\boldsymbol{x}) \, \boldsymbol{v} \leq 0 \tag{3}$$

for any vector $\boldsymbol{v} \in \mathcal{R}^n$ and for all $\boldsymbol{x} \in \mathcal{D}$.

[1]The set of natural, real, and nonnegative real numbers are denoted by $\mathcal{N}$, $\mathcal{R}$ and $\mathcal{R}_+$, respectively. The sets (a,b), $(a,b]$, and $[a,b]$ are the open, the half-open, and the closed interval, respectively. Vectors are written in boldface and their transpose is indicated by the superscript T. Gradient and Hessian of a function f are shown by ∇f and $\nabla^2 f$, respectively.

[2]A set $\mathcal{D}$ is convex if the line segment joining any two points in $\mathcal{D}$ is part of $\mathcal{D}$.

[3]Unless otherwise stated, we consider finite-dimensional (sub-)spaces and finite (sub-)sets with cardinalities denoted by $m \in \mathcal{N}$ and $n \in \mathcal{N}$ throughout the present paper. In few cases we also allow countably infinite spaces.

The concept of convexity and concavity can be extended in various directions. A crucial and well-known representative is the class of *logarithmically convex* functions, from which tighter bounds in optimization theory evolve. As we will see later, its counterpart, the class of *exponentially concave* functions plays an important role in information theory. The definition is analogous to the one of logarithmically convex functions.

A function $f : \mathcal{D} \to \mathcal{R}$ is exponentially (strictly) concave in $\boldsymbol{x} \in \mathcal{D}$, if the function $\exp\big(f(\boldsymbol{x})\big)$ is (strictly) concave. Hence, equivalent definitions for exponentially concave functions are obtained by applying (1), (2) and (3) to $\exp\big(f(\boldsymbol{x})\big)$. Straightforward algebra yields the following:

a) (equivalent definition[4]) $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{D},\ \forall \beta \in [0, 1]$:

$$\log\big(\beta \exp[f(\boldsymbol{x})] + \bar{\beta} \exp[f(\boldsymbol{y})]\big) \leq f(\beta \boldsymbol{x} + \bar{\beta} \boldsymbol{y})\,, \tag{4}$$

b) (assuming differentiability) $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$:

$$f(\boldsymbol{y}) - f(\boldsymbol{x}) \leq \log\big(1 + \nabla f(\boldsymbol{x})^{\mathrm{T}} (\boldsymbol{y} - \boldsymbol{x})\big)\,, \tag{5}$$

c) (assuming twice-differentiability) $\forall \boldsymbol{x} \in \mathcal{D},\ \forall \boldsymbol{v} \in \mathcal{R}^n$:

$$\boldsymbol{v}^{\mathrm{T}} \left(\nabla^2 f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})\, \nabla f(\boldsymbol{x})^{\mathrm{T}}\right) \boldsymbol{v} \leq 0\,. \tag{6}$$

Comparing (1)-(3) with (4)-(6) shows that tighter inequalities for exponentially concave functions exist. In particular, we have

$$\begin{aligned} &\beta f(\boldsymbol{x}) + \bar{\beta} f(\boldsymbol{y}) \\ &\quad \leq \log\big(\beta \exp[f(\boldsymbol{x})] + \bar{\beta} \exp[f(\boldsymbol{y})]\big) \leq f(\beta \boldsymbol{x} + \bar{\beta} \boldsymbol{y})\,, \end{aligned} \tag{7}$$

$$\begin{aligned} f(\boldsymbol{y}) - f(\boldsymbol{x}) &\leq \log\big(1 + \nabla f(\boldsymbol{x})^{\mathrm{T}} (\boldsymbol{y} - \boldsymbol{x})\big) \\ &\leq \nabla f(\boldsymbol{x})^{\mathrm{T}} (\boldsymbol{y} - \boldsymbol{x})\,, \end{aligned} \tag{8}$$

and

$$\boldsymbol{v}^{\mathrm{T}}\, \nabla^2 f(\boldsymbol{x})\, \boldsymbol{v} \leq \boldsymbol{v}^{\mathrm{T}} \left(\nabla^2 f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})\, \nabla f(\boldsymbol{x})^{\mathrm{T}}\right) \boldsymbol{v} \leq 0\,. \tag{9}$$

Inequalities (7)-(9) also show that every exponentially concave function is a concave function, too.

By using mathematical induction, we can generalize inequality (4) as follows.

Corollary 1. Let $f : \mathcal{D} \to \mathcal{R}$ be an exponentially concave function on the convex set $\mathcal{D} \subseteq \mathcal{R}^n$. Then for all $w_i \in \mathcal{R}_+$ with $\sum_i w_i = 1$ the inequalities

$$\sum\nolimits_i w_i f(\boldsymbol{x}_i) \leq \log\Big(\sum\nolimits_i w_i \exp\big(f(\boldsymbol{x}_i)\big)\Big) \leq f\Big(\sum\nolimits_i w_i \boldsymbol{x}_i\Big) \tag{10}$$

hold. On the right side, equality is attained for $f(\boldsymbol{x}) = \log\big(c_0 + \sum_j c_j x_j\big)$ with properly chosen constants $c_0, c_1, \ldots$.

Note that the left inequality in (10) is a simple consequence of Jensen's inequality. However, we have observed that especially for exponentially concave functions the left inequality usually achieves smaller gaps compared to the right one which is interesting in its own.

[4]The base of the logarithm is the Euler's number e while exp refers to the natural exponential function. Whenever the base becomes important, we write $\log_a(x)$ and a^x instead of $\log(x)$ and $\exp(x)$, respectively, to highlight the corresponding base a of both the logarithm and the exponential function.

A generalization of (5) yields a squeeze-inequality for the difference $f\big(\sum_j w_j \boldsymbol{x}_j\big) - \sum_i w_i f(\boldsymbol{x}_i)$, which is the content of the next two corollaries.

By choosing $\boldsymbol{y} = \boldsymbol{x}_i$ and $\boldsymbol{x} = \sum_j w_j \boldsymbol{x}_j$, and multiplying both sides of (5) with w_i and summing up over all i, the following lower bound is derived.

Corollary 2. Let $f : \mathcal{D} \to \mathcal{R}$ be an exponentially concave function on the convex set $\mathcal{D} \subseteq \mathcal{R}^n$. Then for all $w_i \in \mathcal{R}_+$ with $\sum_i w_i = 1$ the double-inequality

$$\begin{aligned} 0 \leq &- \sum\nolimits_i w_i \log\Big(1 + \nabla f\Big(\sum\nolimits_j w_j \boldsymbol{x}_j\Big)^{\mathrm{T}} \Big(\boldsymbol{x}_i - \sum\nolimits_j w_j \boldsymbol{x}_j\Big)\Big) \\ &\leq f\Big(\sum\nolimits_j w_j \boldsymbol{x}_j\Big) - \sum\nolimits_i w_i f(\boldsymbol{x}_i) \end{aligned} \tag{11}$$

holds.

The logarithmic terms on the lefthand side of (11) are always negative, as can be seen by enlarging their argument by the aid of AM-GM inequality, cf. [4].

By replacing $\boldsymbol{x} = \boldsymbol{x}_i$ and $\boldsymbol{y} = \sum_j w_j \boldsymbol{x}_j$ in (5) and proceeding analogously to the above, we obtain the following upper bound.

Corollary 3. Let $f : \mathcal{D} \to \mathcal{R}$ be an exponentially concave function on the convex set $\mathcal{D} \subseteq \mathcal{R}^n$. Then for all $w_i \in \mathcal{R}_+$ with $\sum_i w_i = 1$ the inequality

$$\begin{aligned} &f\Big(\sum\nolimits_j w_j \boldsymbol{x}_j\Big) - \sum\nolimits_i w_i f(\boldsymbol{x}_i) \\ &\quad \leq \sum\nolimits_i w_i \log\Big(1 - \nabla f(\boldsymbol{x}_i)^{\mathrm{T}} \Big(\boldsymbol{x}_i - \sum\nolimits_j w_j \boldsymbol{x}_j\Big)\Big) \end{aligned} \tag{12}$$

holds.

Integral versions of (10)-(12) can be derived under additional constraints on f, e.g., boundedness and Riemann integrability.

We will now discuss mathematical properties of exponentially concave functions.

III. Mathematical Attributes

In the following we will show that under certain assumptions the shift, scale, combination, and composition of exponentially concave functions preserve exponential concavity.

Proposition 4. Let f be an exponentially concave function. Adding a constant $c_1 \in \mathcal{R}$ to f or multiplying f by a factor $c_2 \in [0, 1]$ preserves its exponential concavity.

Proof. The proof is easy, since $\exp\big(c_1 + f(\boldsymbol{x})\big) = \exp(c_1) \exp\big(f(\boldsymbol{x})\big)$ and $\exp\big(c_2 f(\boldsymbol{x})\big) = \big[\exp\big(f(\boldsymbol{x})\big)\big]^{c_2}$ are concave functions for any $c_1 \in \mathcal{R}$ and any $c_2 \in [0, 1]$ whenever $\exp\big(f(\boldsymbol{x})\big)$ is concave. ∎

For certain functions the multiplicative constant may be greater than one, since for a twice-differentiable and exponentially concave function $f : \mathcal{D} \subseteq \mathcal{R} \to \mathcal{R}$ we obtain for $cf(x)$ the largest possible constant by $c_{\text{max}} = \inf_{x \in \mathcal{D}} \frac{-f''(x)}{[f'(x)]^2}$ from (6). For exponentially concave functions c_{max} is necessarily greater than or equal to one. Hence, functions with $c_{\text{max}} < 1$ cannot be exponentially concave.

Proposition 5. The convex combination $\sum_i w_i f_i(\boldsymbol{x}_i)$ of exponentially concave functions $f_i : \mathcal{D}_i \to \mathcal{R}$ defined on convex sets $\mathcal{D}_i \subseteq \mathcal{R}^n$ with weights $w_i \in \mathcal{R}_+$, $\sum_i w_i = 1$, is exponentially concave on $\bigcap_i \mathcal{D}_i$.

Proof. We show that (4) holds for $\sum_i w_i f_i(\boldsymbol{x}_i)$. By using the generalized Hölder's inequality, cf. [4], we obtain the following chain of inequalities

$$\begin{aligned}
&\log\Big(\beta \exp\Big[\sum\nolimits_i w_i f_i(\boldsymbol{x}_i)\Big] + \bar{\beta}\exp\Big[\sum\nolimits_i w_i f_i(\boldsymbol{y}_i)\Big]\Big)\\
&\quad = \log\Big(\beta \prod\nolimits_i \exp[w_i f_i(\boldsymbol{x}_i)] + \bar{\beta}\prod\nolimits_j \exp[w_j f_j(\boldsymbol{y}_i)]\Big)\\
&\quad \le \log\Big(\prod\nolimits_i \big(\beta\exp[f_i(\boldsymbol{x}_i)] + \bar{\beta}\exp[f_i(\boldsymbol{y}_i)]\big)^{w_i}\Big)\\
&\quad = \sum\nolimits_i w_i \log\big(\beta\exp[f_i(\boldsymbol{x}_i)] + \bar{\beta}\exp[f_i(\boldsymbol{y}_i)]\big)\\
&\quad \le \sum\nolimits_i w_i\, f_i(\beta\boldsymbol{x}_i + \bar{\beta}\boldsymbol{y}_i)\,,
\end{aligned}$$

which completes the proof. ∎

Proposition 6. The sum $f_1(\boldsymbol{x}) + f_2(\boldsymbol{x})$ of exponentially concave functions $f_1, f_2 : \mathcal{D} \to \mathcal{R}$, on the convex set $\mathcal{D} \subseteq \mathcal{R}^n$, is exponentially concave, if

$$[f_1(\boldsymbol{x}) - f_1(\boldsymbol{y})][f_2(\boldsymbol{x}) - f_2(\boldsymbol{y})] \le 0 \tag{13}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$ holds. For $n = 1$ relation (13) holds if f_1 and f_2 are contra-monotonic.

Proof. Since the exponential function is monotonic, relation (13) is equivalent to

$$\begin{aligned}
&\exp\big(f_1(\boldsymbol{x}) + f_2(\boldsymbol{x})\big) + \exp\big(f_1(\boldsymbol{y}) + f_2(\boldsymbol{y})\big)\\
&\quad \le \exp\big(f_1(\boldsymbol{x}) + f_2(\boldsymbol{y})\big) + \exp\big(f_1(\boldsymbol{y}) + f_2(\boldsymbol{x})\big)\,.
\end{aligned} \tag{14}$$

By using the exponential concavity of f_1 and f_2, and applying (14) afterwards, we infer

$$\begin{aligned}
&\exp\big(f_1(\beta\boldsymbol{x} + \bar{\beta}\boldsymbol{y}) + f_2(\beta\boldsymbol{x} + \bar{\beta}\boldsymbol{y})\big)\\
&\quad \ge \big[\beta\exp\big(f_1(\boldsymbol{x})\big) + \bar{\beta}\exp\big(f_1(\boldsymbol{y})\big)\big]\\
&\quad\quad \cdot \big[\beta\exp\big(f_2(\boldsymbol{x})\big) + \bar{\beta}\exp\big(f_2(\boldsymbol{y})\big)\big]\\
&\quad = \beta^2\exp\big(f_1(\boldsymbol{x}) + f_2(\boldsymbol{x})\big) + \bar{\beta}^2\exp\big(f_1(\boldsymbol{y}) + f_2(\boldsymbol{y})\big)\\
&\quad\quad + \beta\bar{\beta}\big[\exp\big(f_1(\boldsymbol{x}) + f_2(\boldsymbol{y})\big) + \exp\big(f_1(\boldsymbol{y}) + f_2(\boldsymbol{x})\big)\big]\\
&\quad \ge \beta\exp\big(f_1(\boldsymbol{x}) + f_2(\boldsymbol{x})\big) + \bar{\beta}\exp\big(f_1(\boldsymbol{y}) + f_2(\boldsymbol{y})\big)\,,
\end{aligned}$$

which is equivalent to (4) and thus completes the proof. ∎

Proposition 7. Let $f : \mathcal{D}_f \to \mathcal{R}$ be an exponentially concave function on the convex set $\mathcal{D}_f \subseteq \mathcal{R}^m$. Let $\boldsymbol{g} : \mathcal{D}_{\boldsymbol{g}} \to \mathcal{D}_f$ have components g_i, $1 \le i \le m$, where $\mathcal{D}_{\boldsymbol{g}} \subseteq \mathcal{R}^n$ is a convex set. If f is non-decreasing in each argument and g_i concave for all i, or f is non-increasing in each argument and g_i convex for all i, then the composition $f\big(\boldsymbol{g}(\boldsymbol{x})\big)$ is exponentially concave.

Proof. By applying inequality (4) to f, then using the monotonicity of f in connection with concavity or convexity of each g_i, we obtain

$$\begin{aligned}
&\log\big(\beta\exp\big[f\big(\boldsymbol{g}(\boldsymbol{x})\big)\big] + \bar{\beta}\exp\big[f\big(\boldsymbol{g}(\boldsymbol{y})\big)\big]\big)\\
&\quad \le f\big(\beta\boldsymbol{g}(\boldsymbol{x}) + \bar{\beta}\boldsymbol{g}(\boldsymbol{y})\big)\\
&\quad \le f\big(\boldsymbol{g}[\beta\boldsymbol{x} + \bar{\beta}\boldsymbol{y}]\big)\,,
\end{aligned}$$

which proves the assertion. ∎

An obvious consequence of Proposition 7 is that $\log\big(\sum_i w_i g_i(\boldsymbol{x}_i)\big)$ is exponentially concave for any convex combination $\sum_i w_i g_i(\boldsymbol{x}_i)$ of concave functions g_i.

A degenerate case of Proposition 7 is when each g_i becomes the identity function. In this case we obtain the following statement.

Proposition 8. Let $f : \mathcal{D}_f \to \mathcal{R}$ be an exponentially concave function on the convex set $\mathcal{D}_f \subseteq \mathcal{R}^m$. Then the composition $f\big(\sum_i w_i \boldsymbol{x}_i\big)$, with $\boldsymbol{x}_i \in \mathcal{D}_f$, $w_i \ge 0$ and $\sum_i w_i = 1$, is an exponentially concave function of $(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots)$.

Proof. By applying inequality (4) to f and using the linearity of $\sum_i w_i \boldsymbol{x}_i$, we obtain

$$\begin{aligned}
&\log\Big(\beta\exp\Big[f\Big(\sum\nolimits_i w_i\boldsymbol{x}_i\Big)\Big] + \bar{\beta}\exp\Big[f\Big(\sum\nolimits_i w_i\boldsymbol{y}_i\Big)\Big]\Big)\\
&\quad \le f\Big(\beta\sum\nolimits_i w_i\boldsymbol{x}_i + \bar{\beta}\sum\nolimits_i w_i\boldsymbol{y}_i\Big)\\
&\quad \le f\Big(\sum\nolimits_i w_i[\beta\boldsymbol{x}_i + \bar{\beta}\boldsymbol{y}_i]\Big)\,,
\end{aligned}$$

which shows the assertion. ∎

We will now investigate the relation between exponentially concave and *Schur-concave* functions. It is well-known that symmetric concave functions are Schur-concave, see [4] with many extensions in [9]. Thus, for symmetric and exponentially concave functions one can easily show the Schur-concavity. The converse, that a Schur-concave function is exponentially concave, is not true in general and needs additional assumptions, one of which is introduced in the next proposition.

Proposition 9. Let $\mathcal{D}$ be an interval and f a continuous real function on $\mathcal{D}^n$. If the function $\phi(\boldsymbol{x}_1, \boldsymbol{x}_2) = \exp\big(f(\boldsymbol{x}_1)\big) + \exp\big(f(\boldsymbol{x}_2)\big)$ is Schur-concave on $\mathcal{D}^{n\times 2}$, i.e., $\phi(\boldsymbol{y}_1, \boldsymbol{y}_2) \le \phi(\boldsymbol{x}_1, \boldsymbol{x}_2)$ for all matrices[5] $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ that are row majorized[6] by $(\boldsymbol{y}_1, \boldsymbol{y}_2)$, then the function f is exponentially concave on $\mathcal{D}^n$.

Proof. Since ϕ is Schur-concave, it holds that $\phi(\boldsymbol{y}_1, \boldsymbol{y}_2) \le \phi(\boldsymbol{x}_1, \boldsymbol{x}_2)$ for the specific choice $\boldsymbol{x}_1 = \boldsymbol{x}_2 = (\boldsymbol{y}_1 + \boldsymbol{y}_2)/2$. This yields $\exp\big(f(\boldsymbol{y}_1)\big) + \exp\big(f(\boldsymbol{y}_2)\big) \le 2\exp\big(f(\boldsymbol{y}_1/2 + \boldsymbol{y}_2/2)\big)$, which shows that $\exp\big(f(\boldsymbol{x})\big)$ is a midconcave function of $\boldsymbol{x}$. Since f is continuous, the concavity of $\exp\big(f(\boldsymbol{x})\big)$ follows from Jensen's theorem, see [3, p. 215]. Hence, f is exponentially concave. ∎

A simple connection to *quasiconcave* functions is considered next, which shows that exponentially concave functions are as well quasiconcave.

Proposition 10. The exponential function of any concave function is quasiconcave.

Proof. Let $f : \mathcal{D} \to \mathcal{R}$ be a concave function on the convex set $\mathcal{D} \in \mathcal{R}^n$. Then the inequality $\beta f(\boldsymbol{x}) + \bar{\beta} f(\boldsymbol{y}) \le$

[5] Note that $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are column vectors such that $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$ and $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$ are $n \times 2$ matrices.

[6] The statement in Proposition 9 remains valid for different multivariate majorization techniques, e.g., the chain majorization or the ordinary majorization, whenever the row majorization is implied by them, cf. [9, p. 620].

$f(\beta \boldsymbol{x} + \bar\beta \boldsymbol{y})$ holds for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$, $\beta \in [0,1]$. For $g(\boldsymbol{x}) = \exp\big(f(\boldsymbol{x})\big)$ to be a quasiconcave function of $\boldsymbol{x} \in \mathcal{D}$ we have to show the inequality

$$\min\{g(\boldsymbol{x}), g(\boldsymbol{y})\} \le g(\beta \boldsymbol{x} + \bar\beta \boldsymbol{y})$$

for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$, $\beta \in [0,1]$. Assuming w.l.o.g. that $g(\boldsymbol{x}) \le g(\boldsymbol{y})$, which is equivalent to $f(\boldsymbol{x}) \le f(\boldsymbol{y})$, we conclude

$$\begin{aligned} g(\boldsymbol{x}) &= \exp\big(\beta f(\boldsymbol{x}) + \bar\beta f(\boldsymbol{x})\big) \le \exp\big(\beta f(\boldsymbol{x}) + \bar\beta f(\boldsymbol{y})\big) \\ &\le \exp\big(f(\beta \boldsymbol{x} + \bar\beta \boldsymbol{y})\big) = g(\beta \boldsymbol{x} + \bar\beta \boldsymbol{y}) \,, \end{aligned}$$

which completes the proof. ∎

The converse of the above statement does not hold as can be seen from simple examples.

The *perspective* of exponentially concave functions can be defined analogously to [10, p. 89]. Unfortunately, the corresponding perspective is not homogeneous.

Proposition 11. Let $f : \mathcal{D} \to \mathcal{R}$ be an exponentially concave function on the convex set $\mathcal{D} \in \mathcal{R}^n$. Then its perspective $g(y, \boldsymbol{x}) = \log(y) + f\big(\frac{\boldsymbol{x}}{y}\big)$ is also exponentially concave on $\mathcal{R}_+ \times \mathcal{D}$.

Proof. The assertion is proven by the following chain of equations

$$\begin{aligned} &\exp\big[g(\beta y_1 + \bar\beta y_2, \beta \boldsymbol{x}_1 + \bar\beta \boldsymbol{x}_2)\big] \\ &\quad = \exp\Big[\log(\beta y_1 + \bar\beta y_2) + f\Big(\frac{\beta \boldsymbol{x}_1 + \bar\beta \boldsymbol{x}_2}{\beta y_1 + \bar\beta y_2}\Big)\Big] \\ &\quad = (\beta y_1 + \bar\beta y_2) \exp\Big[f\Big(\frac{\beta y_1}{\beta y_1 + \bar\beta y_2}\frac{\boldsymbol{x}_1}{y_1} + \frac{\bar\beta y_2}{\beta y_1 + \bar\beta y_2}\frac{\boldsymbol{x}_2}{y_2}\Big)\Big] \\ &\quad \ge \beta y_1 \exp\Big[f\Big(\frac{\boldsymbol{x}_1}{y_1}\Big)\Big] + \bar\beta y_2 \exp\Big[f\Big(\frac{\boldsymbol{x}_2}{y_2}\Big)\Big] \\ &\quad = \beta \exp\big[g(y_1, \boldsymbol{x}_1)\big] + \exp\big[g(y_2, \boldsymbol{x}_2)\big] \,, \end{aligned}$$

which satisfies (4). ∎

Since exponentially concave functions are concave, maximization of exponentially concave functions has the advantage that every local optimum is also a global maximum. Moreover, exponentially concave functions can be upper-bounded by concave and sufficiently smooth hypersurfaces.

Proposition 12. Let $f : \mathcal{D} \to \mathcal{R}$ be an exponentially concave function on the convex set $\mathcal{D} \in \mathcal{R}^n$. Then the inequality $f(\boldsymbol{y}) \le f(\boldsymbol{x}) + \log\big(1 + \boldsymbol{v}^{\mathrm{T}} (\boldsymbol{y} - \boldsymbol{x})\big)$ holds for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$ and proper vector $\boldsymbol{v}$.

Proof. It is well-known that for any concave function g at each point $\boldsymbol{x} \in \mathcal{D}$ there exists some $\tilde{\boldsymbol{v}}$ such that $g(\boldsymbol{y}) \le g(\boldsymbol{x}) + \tilde{\boldsymbol{v}}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x})$ for all $\boldsymbol{y} \in \mathcal{D}$. If g is differentiable at $\boldsymbol{x}$ then $\tilde{\boldsymbol{v}} = \nabla g(\boldsymbol{x})$ may be chosen. Replacing g by $\exp(f)$ and applying the logarithm on both sides yields $f(\boldsymbol{y}) \le f(\boldsymbol{x}) + \log\big(1 + \exp(-f(\boldsymbol{x}))\tilde{\boldsymbol{v}}^{\mathrm{T}} (\boldsymbol{y} - \boldsymbol{x})\big)$ with $\boldsymbol{v} = \exp(-f(\boldsymbol{x}))\tilde{\boldsymbol{v}}$, which proves the statement, c.f. (5). ∎

Considering the argument of the logarithm in equation (5), a new necessary condition for exponentially concave and differentiable functions is obtained.

Corollary 13. For an exponentially concave and differentiable function $f : \mathcal{D} \to \mathcal{R}$ the inequality

$$\nabla f(\boldsymbol{y})^{\mathrm{T}} (\boldsymbol{y} - \boldsymbol{x}) \le 1 \tag{15}$$

holds for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$. Particularly for $\boldsymbol{x} = \boldsymbol{0}$, the relation $\nabla f(\boldsymbol{y})^{\mathrm{T}} \boldsymbol{y} \le 1$ limits the inner product between the slope and the position of any point $\boldsymbol{y}$.

Corollary 13 provides a quick test for excluding functions from the class of exponentially concave functions. For example, the function $\log(1 + x)$ is exponentially concave, but the sum $\log(1 + x) + x$ is not. Since multiplying the derivative $[\log(1 + x) + x]' = \frac{2+x}{1+x}$ by $y - x$ and choosing $y = 2x + 1$ yields $2 + x > 1$ for all $x > -1$. Hence, inequality (15) is violated, which means that $\log(1+x)+x$ is not exponentially concave.

The converse of Corollary 13 is not valid in general as simple examples show.

IV. Exponential Concavity of Information Theoretic Quantities

In this section, we show exponential concavity of information theoretic quantities like self-information and entropy. We start with the self-information and will consequently show that the (scaled) entropy is exponentially concave which results in an inequality for the entropy of mixtures of discrete distributions.

Theorem 14. Self-information $\rho(x) = -x \log x$ is exponentially concave for $x \in [0, 1]$.

Proof. Since ρ is differentiable, we consider the second derivative $\exp\big(\rho(x)\big)'' = \exp\big(\rho(x)\big)v(x)$ with $v(x) = \log^2(\mathrm{e}\, x) - x^{-1}$ as well as $v'(x) = x^{-1}\big(x^{-1} + 2\log(\mathrm{e}\, x)\big)$ and $v''(x) = 2x^{-2}\big(1 - x^{-1} - \log(\mathrm{e}\, x)\big)$. Due to the inequality $\log y \ge 1 - y^{-1}$, we can enlarge $v''(x)$ to obtain $v''(x) \le 2x^{-3}(e^{-1} - 1) < 0$. Since $v''(x)$ is negative, $v(x)$ is concave and $v'(x)$ is decreasing. With $v'(1) = 3 > 0$ it follows that $v'(x)$ is positive and hence $v(x)$ is increasing. With $v(1) = 0$ we deduce that $v(x)$ is non-positive. Thus $\exp\big(\rho(x)\big)'' \le 0$ yields exponential concavity of $\rho(x)$ for $x \in [0, 1]$. ∎

Since self-information is exponentially concave, we can simply use Proposition 5 to show the following inequality.

Corollary 15. The *weighted* entropy [11], defined by $\mathcal{H}_n(\boldsymbol{x}, \boldsymbol{u}) = -\sum_{i=1}^n u_i x_i \log x_i$, with $u_i, x_i \in [0, 1]$ and $\sum_{i=1}^n u_i = \sum_{i=1}^n x_i = 1$, is exponentially concave in $\boldsymbol{x}$ for all $n \ge 1$, i.e.,

$$\begin{aligned} \beta \exp\Big(\sum_{i=1}^n u_i \rho(x_i)\Big) + \bar\beta \exp\Big(\sum_{i=1}^n u_i \rho(y_i)\Big) \\ \le \exp\Big(\sum_{i=1}^n u_i \rho\big(\beta x_i + \bar\beta y_i\big)\Big) \end{aligned} \tag{16}$$

holds for any $\beta \in [0, 1]$.

The compositions $\exp\big(H_2(x, 1-x)\big)$ and $\exp\big(\rho(x)\big)$ of the exponential function with both the binary entropy and the self-information are depicted in Figure 1.

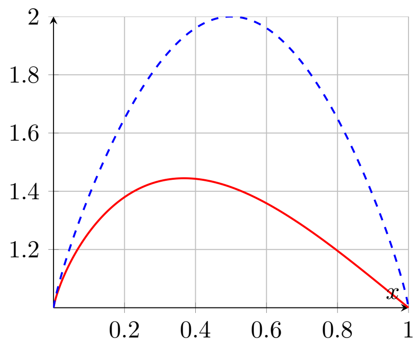

Fig. 1: The functions $\exp\big(H_2(x, 1-x)\big)$ and $\exp\big(\rho(x)\big)$ are shown by a blue dashed and a red solid curve, respectively.

In addition to the above relationship, we can deduce similar inequalities for the ordinary entropy for a limited number of dimensions. In particular, we show that only the binary and the ternary entropies are exponentially concave.

Theorem 16. The entropy $H_n(\boldsymbol{x}) = -\sum_{i=1}^n x_i \log x_i$, with $x_i \in [0,1]$ and $\sum_{i=1}^n x_i = 1$, is exponentially concave in $\boldsymbol{x}$ only for $n \in \{2,3\}$.

Proof. Since ρ, and consequently H_n, is differentiable, we consider the second derivative of

$$g(\beta) = \exp\left(\sum_{i=1}^n \rho\big(\beta x_i + \bar{\beta} y_i\big)\right),$$

in order to show that g is concave between any two feasible points $\boldsymbol{x}$ and $\boldsymbol{y}$ as long as the number n of elements is less than four.

The second derivative of g reads as $g''(\beta) = g(\beta)v(\beta)$ with $v(\beta)$ given by

$$\left(\sum_{i=1}^n \rho'\left(\beta x_i + \bar{\beta} y_i\right)(x_i - y_i)\right)^2 + \sum_{i=1}^n \rho''\left(\beta x_i + \bar{\beta} y_i\right)(x_i - y_i)^2 .$$

Since g is nonnegative, we only need to check the sign of v for the proof. By the substitutions $z_i = \beta x_i + \bar{\beta} y_i$ and $d_i = x_i - y_i$ we obtain

$$v_n(\boldsymbol{d}, \boldsymbol{z}) = \left(\sum_{i=1}^n d_i \log(z_i)\right)^2 - \sum_{i=1}^n \frac{d_i^2}{z_i} .$$

Note that $\boldsymbol{z}$ has to fulfill $\sum_{i=1}^n z_i = 1$ and $z_i \geq 0$ for all i while $\sum_{i=1}^n d_i = 0$ and $-1 \leq d_i \leq 1$ for all i have to be taken into account.

First we show that for all $n \geq 4$ the function $v_n(\boldsymbol{d}, \boldsymbol{z})$ can be positive for particularly chosen $\boldsymbol{d}$ and $\boldsymbol{z}$ which disproves the exponential concavity of entropy H_n for all $n \geq 4$. Consider the choice $\boldsymbol{x}_0 = \big(1 - (n-1)\frac{\varepsilon}{2}, \frac{\varepsilon}{2}, \frac{\varepsilon}{2}, \dots, \frac{\varepsilon}{2}\big)$ and $\boldsymbol{y}_0 = \big((n-1)\frac{\varepsilon}{2}, \frac{1}{n-1} - \frac{\varepsilon}{2}, \frac{1}{n-1} - \frac{\varepsilon}{2}, \dots, \frac{1}{n-1} - \frac{\varepsilon}{2}\big)$ with a sufficiently small $\varepsilon > 0$ with which we determine $\boldsymbol{d}_0 = \big(1 - (n-1)\varepsilon, \varepsilon - \frac{1}{n-1}, \varepsilon - \frac{1}{n-1}, \dots, \varepsilon - \frac{1}{n-1}\big)$ and $\boldsymbol{z}_0 = \big(z, \frac{1-z}{n-1}, \frac{1-z}{n-1}, \dots, \frac{1-z}{n-1}\big)$ with $z = \beta + (1-2\beta)(n-1)\frac{\varepsilon}{2}$. For $0 < \beta < \frac{1}{2}$, corresponding to $0 < z < 1$, it leads to the quantity

$$\nu(z, n) = \frac{v_n(\boldsymbol{d}_0, \boldsymbol{z}_0)}{\big(1 - (n-1)\varepsilon\big)^2} = \log^2\left(z\,\frac{n-1}{1-z}\right) - \frac{1}{z} - \frac{1}{1-z} ,$$

which is a function of z and n, and is increasing in n. Selecting $z = \frac{85}{100}$ yields $\nu\big(\frac{85}{100}, n\big) = \log^2\big(17\,\frac{n-1}{3}\big) - \frac{400}{51} \geq \log^2(17) - \frac{400}{51} > \log^2\big(\mathrm{e}^{\sqrt{8}}\big) - 8 = 0$, where the first inequality arises from the monotonicity in n and the second from modification of the constants. Hence, we obtain that $\nu(z, n)$, and consequently $v_n(\boldsymbol{d}, \boldsymbol{z})$, can be positive for all $n \geq 4$ which disproves the exponential concavity of entropy for $n \geq 4$.

Now, we show that $v_3(\boldsymbol{d}, \boldsymbol{z})$ is always negative in order to prove the exponential concavity of the ternary entropy. Therefore we apply the Cauchy-Bunyakovsky-Schwarz inequality, cf. [4], on the first term in $v_3(\boldsymbol{d}, \boldsymbol{z})$ to obtain the inequality

$$\begin{aligned}
\left(\sum_{i=1}^3 d_i \log(z_i)\right)^2 &= \left(\sum_{i=1}^3 d_i \log\left(\frac{z_i}{\lambda}\right)\right)^2 \\
&\leq \left(\sum_{i=1}^3 \frac{|d_i|}{\sqrt{z_i}} \cdot \sqrt{z_i}\left|\log\left(\frac{z_i}{\lambda}\right)\right|\right)^2 \\
&\leq \sum_{i=1}^3 \frac{d_i^2}{z_i} \cdot \sum_{i=1}^3 z_i \log^2\left(\frac{z_i}{\lambda}\right),
\end{aligned}$$

for any positive λ. Rearranging $\boldsymbol{z}$ such that $z_1 \leq z_2 \leq z_3$ holds and choosing $\lambda = z_3$, it leads to

$$\sum_{i=1}^3 z_i \log^2\left(\frac{z_i}{\lambda}\right) = z_3\left[\frac{z_1}{z_3}\log^2\left(\frac{z_1}{z_3}\right) + \frac{z_2}{z_3}\log^2\left(\frac{z_2}{z_3}\right)\right].$$

Note that the function $y_1 \log^2(y_1) + y_2 \log^2(y_2)$ is Schur-concave in (y_1, y_2) and hence we obtain the upper bound

$$\begin{aligned}
z_3&\left[\frac{z_1}{z_3}\log^2\left(\frac{z_1}{z_3}\right) + \frac{z_2}{z_3}\log^2\left(\frac{z_2}{z_3}\right)\right] \\
&\leq z_3\left[2\frac{1-z_3}{2z_3}\log^2\left(\frac{1-z_3}{2z_3}\right)\right] \\
&\leq (1-z_3)\log^2\left(\frac{1-z_3}{2z_3}\right),
\end{aligned}$$

for all $\frac{1}{3} \leq z_3 \leq 1$. The last quantity can simply be maximized by numerical methods due to its quasiconcavity. This leads to an upper bound of $0.88474 < 1$ at $z_3^\star = 0.84287$. Hence, the inequality

$$\begin{aligned}
\left(\sum_{i=1}^3 d_i \log(z_i)\right)^2 &\leq \sum_{i=1}^3 \frac{d_i^2}{z_i} \cdot \sum_{i=1}^3 z_i \log^2\left(\frac{z_i}{a}\right) \\
&\leq \sum_{i=1}^3 \frac{d_i^2}{z_i}
\end{aligned}$$

holds, which proves $v_3(\boldsymbol{d}, \boldsymbol{z}) \leq 0$. Thus, the ternary entropy is exponentially concave.

It remains to show the exponential concavity of the binary entropy, which can easily be done by forcing $d_1 = 0$ and then following the same steps as for the proof of ternary entropy. ∎

As can be seen from Corollary 15 and Theorem 16 there must exist a value $c_n < n$, depending on the number n of dimensions, such that $\frac{1}{c_n} H_n(\boldsymbol{x})$ becomes exponentially concave for all $n > 1$. The next theorem uses this principle and hence improves and consolidates Corollary 15 and Theorem 16.

Theorem 17. Let H_n be the entropy as defined in Theorem 16. Then the function $\frac{1}{c_n}H_n(\boldsymbol{x})$ is exponentially concave in $\boldsymbol{x}$ for all $n>1$ with $c_n \geq \max_{1/n<z<1} z(1-z)\log^2\big(\frac{1-z}{z(n-1)}\big)$.

Proof. Analogously to the proof of the ternary entropy from Theorem 16, we have to show that the function

$$g(\beta)=\exp\bigg(\frac{1}{c_n}\sum_{i=1}^{n}\rho\big(\beta x_i+\bar{\beta}y_i\big)\bigg)$$

is concave in β, or equivalently we can show that

$$v_n(\boldsymbol{d},\boldsymbol{z})=\frac{1}{c_n}\bigg(\sum_{i=1}^{n}d_i\log(z_i)\bigg)^2-\sum_{i=1}^{n}\frac{d_i^2}{z_i}$$

is non-positive. Again we enlarge $v_n(\boldsymbol{d},\boldsymbol{z})$ by applying the Cauchy-Bunyakovsky-Schwarz inequality to obtain

$$\begin{aligned}v_n(\boldsymbol{d},\boldsymbol{z})&=\frac{1}{c_n}\bigg(\sum_{i=1}^{n}d_i\log\Big(\frac{z_i}{\lambda}\Big)\bigg)^2-\sum_{i=1}^{n}\frac{d_i^2}{z_i}\\&\leq\frac{1}{c_n}\sum_{i=1}^{n}\frac{d_i^2}{z_i}\sum_{i=1}^{n}z_i\log^2\Big(\frac{z_i}{\lambda}\Big)-\sum_{i=1}^{n}\frac{d_i^2}{z_i}\\&=\sum_{i=1}^{n}\frac{d_i^2}{z_i}\bigg(\frac{1}{c_n}\sum_{i=1}^{n}z_i\log^2\Big(\frac{z_i}{\lambda}\Big)-1\bigg).\end{aligned}$$

By optimizing λ the gap of the above inequality can be reduced. We first consider the case $\lambda=1$ which yields a weak lower bound for c_n. Afterwards we optimize λ to obtain a sharper lower bound as stated in the assertion. W.l.o.g. we assume $z_1\leq z_2\leq\cdots\leq z_n$ in what follows.

For $\lambda=1$ the sum $\sum_{i=1}^{n}z_i\log^2(z_i)$ is Schur-concave in $\boldsymbol{z}$ which yields $\frac{1}{c_n}\sum_{i=1}^{n}z_i\log^2(z_i)\leq\frac{1}{c_n}\log^2(n)$ at $z_i^\star=\frac{1}{n}\,\forall i$. From $v_n(\boldsymbol{d},\boldsymbol{z})\leq 0$ we hence obtain $\frac{1}{c_n}\log^2(n)-1\leq 0$ and in turn $c_n\geq\log^2(n)$, which can be improved as follows.

For $\lambda\leq z_n$ the function $y\log^2\big(\frac{y}{\lambda}\big)$ is quasiconcave in y for $y\leq\lambda$ and convex for $y\geq\lambda$. Thus, maximizing $\sum_{i=1}^{n}z_i\log^2\big(\frac{z_i}{\lambda}\big)$ subject to $\sum_{i=1}^{n}z_i=1$, $z_i\geq 0$, will attain its maximum at $z_1^\star=z_2^\star=\cdots=z_{n-1}^\star=\frac{1-z_n^\star}{n-1}\leq\lambda\leq z_n^\star$, i.e., $\sum_{i=1}^{n}z_i\log^2\big(\frac{z_i}{\lambda}\big)\leq(1-z_n^\star)\log^2\big(\frac{1-z_n^\star}{\lambda(n-1)}\big)+z_n^\star\log^2\big(\frac{z_n^\star}{\lambda}\big)$. Since the last term is quasiconvex in λ we can reduce the gap by minimizing over λ. With $\lambda^\star=z_n^{\star z_n^\star}\big(\frac{1-z_n^\star}{n-1}\big)^{1-z_n^\star}$ we end up in $(1-z_n^\star)\log^2\big(\frac{1-z_n^\star}{\lambda^\star(n-1)}\big)+z_n^\star\log^2\big(\frac{z_n^\star}{\lambda^\star}\big)=z_n^\star(1-z_n^\star)\log^2\big(\frac{1-z_n^\star}{z_n^\star(n-1)}\big)$. From $v_n(\boldsymbol{d},\boldsymbol{z})\leq 0$ we hence obtain $\frac{1}{c_n}z_n^\star(1-z_n^\star)\log^2\big(\frac{1-z_n^\star}{z_n^\star(n-1)}\big)-1\leq 0$ and in turn $c_n\geq\max_{1/n<z<1}z(1-z)\log^2\big(\frac{1-z}{z(n-1)}\big)$.

Note that by using the relation between geometric and logarithmic means one can simply show $z^\star(1-z^\star)\log^2\big(\frac{1-z^\star}{z^\star(n-1)}\big)\leq\log^2(n-1)\leq\log^2(n)$. Table I lists some examples for comparison. ∎

Analogously to Theorem 17, the exponential concavity of Rényi entropy [12], [13], scaled by a number c_n, can also be shown with some more effort.

The following functional structures are key elements for dealing with the mutual information of certain fundamental channel models, see [14]–[17].

TABLE I: Multiplicative factors of the entropy for becoming exponentially concave.

n	$\log^2(n)$	$\max_{1/n<z<1} z(1-z)\log^2\big(\frac{1-z}{z(n-1)}\big)$
2	0.480453	$0.439229=\frac{1}{2.27672}$
3	1.20695	$0.761802=\frac{1}{1.31268}$
4	1.92181	$1.02349=\frac{1}{0.977049}$
5	2.59029	$1.24645=\frac{1}{0.80228}$
10	5.3019	$2.05839=\frac{1}{0.485817}$
20	8.97441	$3.06501=\frac{1}{0.326263}$
50	15.3039	$4.7187=\frac{1}{0.211923}$
100	21.2076	$6.22687=\frac{1}{0.160594}$
1000	47.7171	$12.9004=\frac{1}{0.0775172}$
10000	84.8304	$22.1922=\frac{1}{0.0450608}$

Theorem 18. The difference $\rho\big(\beta\gamma_1+\bar{\beta}\gamma_2\big)-\beta\rho(\gamma_1)-\bar{\beta}\rho(\gamma_2)$ is exponentially concave in $\beta\in[0,1]$ for all $\gamma_1,\gamma_2\in[0,1]$.

Proof. Since self-information is differentiable, we show that inequality (6) holds. Let $f(\beta)=\rho\big(\beta\gamma_1+\bar{\beta}\gamma_2\big)-\beta\rho(\gamma_1)-\bar{\beta}\rho(\gamma_2)$ and $v(\beta)=f''(\beta)+[f'(\beta)]^2$. If $f(\beta)$ is exponentially concave in β, then $v(\beta)\leq 0$ must hold. By simple calculation we obtain $f'(\beta)=(\gamma_1-\gamma_2)\rho'\big(\beta\gamma_1+\bar{\beta}\gamma_2\big)-\big(\rho(\gamma_1)-\rho(\gamma_2)\big)$ and $f''(\beta)=(\gamma_1-\gamma_2)^2\rho''\big(\beta\gamma_1+\bar{\beta}\gamma_2\big)$ with $\rho'(\beta)=-1-\log\beta$ and $\rho''(\beta)=-\beta^{-1}$. It is easy to check that $v(\beta)=0$ for $\gamma_1=\gamma_2$, so no further investigation is needed for this case. Since $f''(\beta)$ is non-positive, we can represent $v(\beta)$ in the form $\big(f'(\beta)-\sqrt{|f''(\beta)|}\big)\big(f'(\beta)+\sqrt{|f''(\beta)|}\big)$. Now we only consider the case $\gamma_1>\gamma_2$ in the following, since the opposite case can be treated analogously. Then we have

$$\begin{aligned}f'(\beta)-\sqrt{|f''(\beta)|}=&-\big(\gamma_1+\rho(\gamma_1)-\gamma_2-\rho(\gamma_2)\big)\\&-(\gamma_1-\gamma_2)\big(\tfrac{1}{\sqrt{\gamma_0}}+\log\gamma_0\big)\end{aligned}\tag{17}$$

and

$$\begin{aligned}f'(\beta)+\sqrt{|f''(\beta)|}=&-\big(\gamma_1+\rho(\gamma_1)-\gamma_2-\rho(\gamma_2)\big)\\&+(\gamma_1-\gamma_2)\big(\tfrac{1}{\sqrt{\gamma_0}}-\log\gamma_0\big)\end{aligned}\tag{18}$$

with $\gamma_0=\beta\gamma_1+\bar{\beta}\gamma_2$. Consider now the function $\gamma+\rho(\gamma)$ with $[\gamma+\rho(\gamma)]'=-\log\gamma\geq 0$ for all $\gamma\in(0,1]$. Since its derivative is non-negative, the function $\gamma+\rho(\gamma)$ is increasing in γ. This means that the quantity $\big(\gamma_1+\rho(\gamma_1)-\gamma_2-\rho(\gamma_2)\big)$ is non-negative for $\gamma_1>\gamma_2$. It is easy[7] to show that $\frac{1}{\sqrt{\gamma_0}}-\log\gamma_0\geq\frac{1}{\sqrt{\gamma_0}}+\log\gamma_0\geq 2(1-\log 2)\geq 0$ for $\gamma_0\in[0,1]$. Hence, the function (17) is negative. Since $\frac{1}{\sqrt{\gamma_0}}-\log\gamma_0$ is decreasing in

[7]The function $\frac{1}{\sqrt{\gamma}}+\log\gamma$ has the derivative $\frac{2\sqrt{\gamma}-1}{\sqrt{2}\gamma^{3/2}}$, which is positive for all $\gamma>1/4$ and non-positive otherwise. Hence, $\gamma=1/4$ is a global minimum, which leads to $\frac{1}{\sqrt{\gamma}}+\log\gamma\geq 2(1-\log 2)\geq 0$. In addition, since $\gamma\in[0,1]$, it holds $\frac{1}{\sqrt{\gamma}}-\log\gamma\geq\frac{1}{\sqrt{\gamma}}+\log\gamma$.

γ_0, we maximize γ_0 by replacing it with γ_1 to obtain

$$\begin{aligned} f'(\beta) + \sqrt{|f''(\beta)|} &\geq -\big(\gamma_1 + \rho(\gamma_1) - \gamma_2 - \rho(\gamma_2)\big) \\ &\quad + (\gamma_1 - \gamma_2)\big(\tfrac{1}{\sqrt{\gamma_1}} - \log \gamma_1\big) \\ &= (1 - \sqrt{\gamma_1})\big(1 - \tfrac{\gamma_2}{\gamma_1}\big)\sqrt{\gamma_1} + \gamma_2 \log\big(\tfrac{\gamma_1}{\gamma_2}\big) \\ &\geq 0\,. \end{aligned}$$

Hence, the product of $f'(\beta) - \sqrt{|f''(\beta)|}$ with $f'(\beta) + \sqrt{|f''(\beta)|}$ is negative, which shows the negativity of $v(\beta)$. This completes the proof. ∎

For the sake of compactness, we hereinafter denote the binary entropy $H_2(x, 1-x)$ by $H(x)$.

Theorem 19. The difference $H(\beta\gamma_1 + \bar{\beta}\gamma_2) - \beta H(\gamma_1) - \bar{\beta}H(\gamma_2)$ is exponentially concave in $\beta \in [0,1]$ for all $\gamma_1, \gamma_2 \in [0,1]$.

Proof. We consider the function $f(\beta) = H(\beta\gamma_1 + \bar{\beta}\gamma_2) - \beta H(\gamma_1) - \bar{\beta}H(\gamma_2)$ and its first and second derivatives w.r.t. β as given by

$$f'(\beta) = H'(\beta\gamma_1 + \bar{\beta}\gamma_2)(\gamma_1 - \gamma_2) - H(\gamma_1) + H(\gamma_2) \tag{19}$$

and

$$f''(\beta) = H''(\beta\gamma_1 + \bar{\beta}\gamma_2)(\gamma_1 - \gamma_2)^2\,, \tag{20}$$

where $H'(x) = \log\frac{1-x}{x}$ and $H''(x) = \frac{-1}{x(1-x)}$. Without loss of generality we assume $\gamma_2 < \gamma_1$, since for the case $\gamma_1 = \gamma_2$ the function f becomes zero. Similar to the above proof, we define $v(\beta) = f''(\beta) + [f'(\beta)]^2$ and show that v is non-positive to prove the exponential concavity of f. The decomposition of v yields

$$\begin{aligned} f'(\beta) - \sqrt{|f''(\beta)|} = &- \frac{\gamma_1 - \gamma_2}{\sqrt{\gamma_0(1-\gamma_0)}} - \big(H(\gamma_1) - H(\gamma_2)\big) \\ &+ (\gamma_1 - \gamma_2)\log\Big(\frac{1-\gamma_0}{\gamma_0}\Big) \end{aligned} \tag{21}$$

and

$$\begin{aligned} f'(\beta) + \sqrt{|f''(\beta)|} = &+ \frac{\gamma_1 - \gamma_2}{\sqrt{\gamma_0(1-\gamma_0)}} - \big(H(\gamma_1) - H(\gamma_2)\big) \\ &+ (\gamma_1 - \gamma_2)\log\Big(\frac{1-\gamma_0}{\gamma_0}\Big) \end{aligned} \tag{22}$$

with $\gamma_0 = \beta\gamma_1 + \bar{\beta}\gamma_2$. Note that both functions $\log\big(\frac{1-\gamma_0}{\gamma_0}\big) + \frac{1}{\sqrt{\gamma_0(1-\gamma_0)}}$ and $\log\big(\frac{1-\gamma_0}{\gamma_0}\big) - \frac{1}{\sqrt{\gamma_0(1-\gamma_0)}}$ are decreasing in γ_0 for $\gamma_0 \leq \frac{1}{2}$ and $\gamma_0 \geq \frac{1}{2}$, respectively, since the derivatives are non-positive, i.e.,

$$\begin{aligned} &\frac{\partial}{\partial\gamma_0}\log\Big(\frac{1-\gamma_0}{\gamma_0}\Big) \pm \frac{1}{\sqrt{\gamma_0(1-\gamma_0)}} \\ &\qquad = \frac{-1}{\gamma_0(1-\gamma_0)}\Big(1 \pm \frac{1-2\gamma_0}{2\sqrt{\gamma_0(1-\gamma_0)}}\Big) \leq 0\,. \end{aligned}$$

In the following we only consider the particular case $\gamma_2 < \gamma_1 < 1 - \gamma_2$ along with the subcases $\gamma_0 \geq \frac{1}{2}$ and $\gamma_0 \leq \frac{1}{2}$. The case $\gamma_2 < 1 - \gamma_2 < \gamma_1$ can be proven similarly, after multiplication of both (21) and (22) with minus one.

In the subcase $\gamma_2 < \gamma_1 < 1 - \gamma_2$ with $\gamma_0 \geq \frac{1}{2}$, we have $\log\big(\frac{1-\gamma_0}{\gamma_0}\big) \leq 0$, $\gamma_1 \geq \frac{1}{2}$, and $H(\gamma_1) > H(\gamma_2) \geq 0$ which lead to $f'(\beta) - \sqrt{|f''(\beta)|} \leq 0$. Using in addition the above monotonicity we obtain

$$\begin{aligned} f'(\beta) + \sqrt{|f''(\beta)|} &\geq + \frac{\gamma_1 - \gamma_2}{\sqrt{\gamma_1(1-\gamma_1)}} - \big(H(\gamma_1) - H(\gamma_2)\big) \\ &\quad + (\gamma_1 - \gamma_2)\log\Big(\frac{1-\gamma_1}{\gamma_1}\Big) \\ &= H(\gamma_2) - \gamma_2 \log\underbrace{\Big(\frac{1-\gamma_1}{\gamma_1}\Big)}_{\leq 1} + \frac{\gamma_1 - \gamma_2}{\sqrt{\gamma_1(1-\gamma_1)}} \geq 0\,. \end{aligned}$$

Hence, the product $v(\beta) = \big(f'(\beta) - \sqrt{|f''(\beta)|}\big)\big(f'(\beta) + \sqrt{|f''(\beta)|}\big)$ is non-positive.

In the subcase $\gamma_2 < \gamma_1 < 1 - \gamma_2$ with $\gamma_0 \leq \frac{1}{2}$, we have $\log\big(\frac{1-\gamma_0}{\gamma_0}\big) \geq 0$ and $H(\gamma_1) > H(\gamma_2)$ which lead to

$$\begin{aligned} f'(\beta) - \sqrt{|f''(\beta)|} &\leq - \frac{\gamma_1 - \gamma_2}{\sqrt{\gamma_0(1-\gamma_0)}} - \big(H(\gamma_1) - H(\gamma_2)\big) \\ &\quad + (\gamma_1 - \gamma_2)\frac{|1 - 2\gamma_0|}{\sqrt{\gamma_0(1-\gamma_0)}} \\ &\leq - \frac{\gamma_1 - \gamma_2}{\sqrt{\gamma_0(1-\gamma_0)}} - \big(H(\gamma_1) - H(\gamma_2)\big) \\ &\quad + \frac{\gamma_1 - \gamma_2}{\sqrt{\gamma_0(1-\gamma_0)}} = -\big(H(\gamma_1) - H(\gamma_2)\big) \leq 0\,, \end{aligned}$$

where we have used the inequality $\log^2 x \leq \frac{(1-x)^2}{x}$ from Proposition 26. With more effort we also deduce

$$\begin{aligned} f'(\beta) + \sqrt{|f''(\beta)|} &\geq + \frac{\gamma_1 - \gamma_2}{\sqrt{\gamma_2(1-\gamma_2)}} - \big(H(\gamma_1) - H(\gamma_2)\big) \\ &\quad + (\gamma_1 - \gamma_2)\log\Big(\frac{1-\gamma_2}{\gamma_2}\Big) \\ &= + \frac{\gamma_1 - \gamma_2}{\sqrt{\gamma_2(1-\gamma_2)}} - \log\underbrace{\Big(\frac{1-\gamma_2}{1-\gamma_1}\Big)}_{\geq 1} \\ &\quad + \gamma_1 \log\Big(\frac{1-\gamma_2}{1-\gamma_1}\frac{\gamma_1}{\gamma_2}\Big) \\ &\geq + \frac{\gamma_1 - \gamma_2}{\sqrt{\gamma_2(1-\gamma_2)}} - \frac{\gamma_1 - \gamma_2}{\sqrt{1-\gamma_1}\sqrt{1-\gamma_2}} \\ &\quad + \gamma_1 \log\Big(\frac{1-\gamma_2}{1-\gamma_1}\frac{\gamma_1}{\gamma_2}\Big) \\ &\geq + \frac{\gamma_1 - \gamma_2}{\sqrt{\gamma_2(1-\gamma_2)}} - \frac{\gamma_1 - \gamma_2}{\sqrt{\gamma_2(1-\gamma_2)}} \\ &\quad + \gamma_1 \log\Big(\frac{1-\gamma_2}{1-\gamma_1}\frac{\gamma_1}{\gamma_2}\Big) = \gamma_1 \log\Big(\underbrace{\frac{1-\gamma_2}{1-\gamma_1}}_{\geq 1}\underbrace{\frac{\gamma_1}{\gamma_2}}_{\geq 1}\Big) \geq 0\,, \end{aligned}$$

where we again have used the inequality $\log^2 x \leq \frac{(1-x)^2}{x}$ and $\gamma_2 < 1 - \gamma_1$, that follows from $\gamma_1 < 1 - \gamma_2$. Thus, the product $v(\beta) = \big(f'(\beta) - \sqrt{|f''(\beta)|}\big)\big(f'(\beta) + \sqrt{|f''(\beta)|}\big)$ is again non-positive.

In summary, $v(\beta)$ in non-positive, which proves the exponential concavity of $f(\beta)$. ∎

As can simply be shown by counterexamples, the differences $H\big(\sum_{i=1}^n \beta_i\gamma_i\big) - \sum_{i=1}^n \beta_i H(\gamma_i)$ under the constraints $\sum_{i=1}^n \beta_i = 1$ and $\beta_i \geq 0$ can never be exponentially concave in $\boldsymbol{\beta}$ for any $n \geq 3$. As an example we consider the function $\exp\big[H\big(\sum_{i=1}^3 \beta_i\gamma_i\big) - \sum_{i=1}^3 \beta_i H(\gamma_i)\big]$, which leads to the obviously convex function $\exp\big[(1-\beta_2)H\big(\frac{4}{5}\big)\big]$ for the particular choice $\gamma_1 = 0$, $\gamma_2 = \frac{4}{5}$, $\gamma_3 = 1$, $\beta_1 = \frac{1-\beta_2}{5}$, and $\beta_3 = 4\frac{1-\beta_2}{5}$.

Exponential concavity can be extended in many more directions, e.g., to Jensen-Steffensen or Hermite-Hadamard-like inequalities and even to differential entropy, which are devoted to future works.

V. Application of Exponential Concavity

Exponential concavity is a useful tool for proving information theoretic inequalities and bounds. This is demonstrated in the present section, where we derive nine new bounds.

Proposition 20. Let $H_n(\boldsymbol{x})$ be the entropy and c_n be as given in Theorem 17. Let $D_n(\boldsymbol{x}\|\boldsymbol{y}) = \sum_{i=1}^n x_i \log \frac{x_i}{y_i}$ be the Kullback-Leibler divergence, $x_i, y_i \in [0,1]$, $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i = 1$. It holds that

$$D_n(\boldsymbol{x}\|\boldsymbol{y}) \geq c_n \log\big(1+\tfrac{1}{c_n}\nabla H_n(\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y})\big) - H_n(\boldsymbol{x}) + H_n(\boldsymbol{y}) \geq 0\,. \tag{23}$$

Proof. By simple calculation, we observe the identity

$$D_n(\boldsymbol{x}\|\boldsymbol{y}) = c_n \frac{1}{c_n}\nabla H_n(\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y}) - H_n(\boldsymbol{x}) + H_n(\boldsymbol{y})\,.$$

Comparing the above identity with inequality (8) and recalling that $\frac{1}{c_n}H_n(\boldsymbol{x})$ is exponentially concave, yields the lower bound. ∎

The *entropy power inequality* is a well-known inequality to describe the relationship between the entropies of two independent random variables and their sum, cf. [18]–[20]. In the next two propositions, we carry over this principle to derive new inequalities between the weighted entropies of random variables and their *mixture* distributions.

Proposition 21. Consider the weighted entropies $\mathcal{H}_n(\boldsymbol{x}_i, \boldsymbol{u})$ of m probability vectors $\boldsymbol{x}_i \in \mathcal{R}_+^n$ and corresponding weights $\boldsymbol{u} \in \mathcal{R}_+^n$. Then the inequalities

$$\begin{aligned}\exp\Big(\sum_{i=1}^m w_i \mathcal{H}_n(\boldsymbol{x}_i, \boldsymbol{u})\Big) &\leq \sum_{i=1}^m w_i \exp\big(\mathcal{H}_n(\boldsymbol{x}_i, \boldsymbol{u})\big)\\ &\leq \exp\Big(\mathcal{H}_n\Big(\sum_{i=1}^m w_i \boldsymbol{x}_i, \boldsymbol{u}\Big)\Big)\end{aligned} \tag{24}$$

hold for all $w_i \in [0,1]$ with $\sum_{i=1}^m w_i = 1$.

Proof. Since the weighted entropy function is exponentially concave as stated in Corollary 15, we can use inequality (10) to deduce the assertion. ∎

Other useful inequalities for comparing the entropies of two distributions are the following ones.

Proposition 22. Consider the weighted entropies $\mathcal{H}_n(\boldsymbol{x}, \boldsymbol{u})$ and $\mathcal{H}_n(\boldsymbol{y}, \boldsymbol{u})$ of probability vectors $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u} \in \mathcal{R}_+^n$. Then the inequality

$$\begin{aligned}&\exp\big(\mathcal{H}_n(\boldsymbol{y},\boldsymbol{u})\big)\sum_{i=1}^n u_i(x_i - y_i)\log(y_i)\\ &\qquad\leq \exp\big(\mathcal{H}_n(\boldsymbol{x},\boldsymbol{u})\big)\sum_{i=1}^n u_i(x_i - y_i)\log(x_i)\end{aligned} \tag{25}$$

holds.

Proof. Since the weighted entropy function is exponentially concave as stated in Corollary 15, the first derivative of $g(\beta) = \exp\big(\mathcal{H}_n(\beta\boldsymbol{x} + \bar{\beta}\boldsymbol{y}, \boldsymbol{u})\big)$ w.r.t. β must be decreasing, i.e., the inequality $g'(1) \leq g'(0)$ holds. This completes the proof. ∎

Similar to the last statement, we can deduce the following assertion.

Proposition 23. Consider the weighted entropies $\mathcal{H}_n(\boldsymbol{x}, \boldsymbol{u})$ and $\mathcal{H}_n(\boldsymbol{y}, \boldsymbol{u})$ of probability vectors $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u} \in \mathcal{R}_+^n$. Then the inequality

$$\exp\big(\mathcal{H}_n(\boldsymbol{y}, \boldsymbol{u}) - \mathcal{H}_n(\boldsymbol{x}, \boldsymbol{u})\big) \leq 1 + \sum_{i=1}^n u_i(x_i - y_i)\log(x_i) \tag{26}$$

holds.

Proof. Since the weighted entropy function is differentiable and exponentially concave as stated in Corollary 15, the inequality in (5) holds which is equivalent to (26). ∎

We can further sharpen (24), (25) and (26) by exploiting the exponential concavity of $\frac{1}{c_n}H_n(\boldsymbol{x})$, as described in Theorem 17. Hence, we derive

$$\begin{aligned}\exp\Big(\tfrac{1}{c_n}\sum_{i=1}^m w_i H_n(\boldsymbol{x}_i)\Big) &\leq \sum_{i=1}^m w_i \exp\big(\tfrac{1}{c_n}H_n(\boldsymbol{x}_i)\big)\\ &\leq \exp\Big(\tfrac{1}{c_n}H_n\Big(\sum_{i=1}^m w_i \boldsymbol{x}_i\Big)\Big)\,,\end{aligned} \tag{27}$$

$$\begin{aligned}&\exp\big(\tfrac{1}{c_n}H_n(\boldsymbol{y})\big)\sum_{i=1}^n (x_i - y_i)\log(y_i)\\ &\qquad\leq \exp\big(\tfrac{1}{c_n}H_n(\boldsymbol{x})\big)\sum_{i=1}^n (x_i - y_i)\log(x_i)\end{aligned} \tag{28}$$

and

$$\exp\big(\tfrac{1}{c_n}H_n(\boldsymbol{y}) - \tfrac{1}{c_n}H_n(\boldsymbol{x})\big) \leq 1 + \tfrac{1}{c_n}\sum_{i=1}^n (x_i - y_i)\log(x_i)\,. \tag{29}$$

It is well-known that mutual information $I_2(\beta)$ of discrete binary channels is given by

$$I_2(z) = H\big(z\gamma_1 + \bar{z}\gamma_2\big) - zH(\gamma_1) - \bar{z}H(\gamma_2) \tag{30}$$

for the transition matrix $\left(\begin{smallmatrix}1-\gamma_1 & \gamma_1\\ 1-\gamma_2 & \gamma_2\end{smallmatrix}\right)$ with $\gamma_1, \gamma_2 \in [0,1]$. The entry in the x^{th} row and the y^{th} column of the matrix denotes the conditional probability that y is received when x is sent. The probability of the two input symbols are denoted by z and

$1-z$, cf. [21], [22]. Note that the mutual information of *binary symmetric channels* (BSC), *binary asymmetric channels* (BAC), and the one-bit quantizer are special cases of (30). The corresponding capacity is derived in [23] and extended in [14] by a decoder specific quantity. The next theorem helps to find proper inequalities for such classes of channels.

Proposition 24. For the mutual information I_2 of discrete binary channels, the inequality

$$\exp\Big(\sum_i w_i I_2(z_i)\Big) \leq \sum_i w_i \exp\big(I_2(z_i)\big) \leq \exp\Big(I_2\Big(\sum_i w_i z_i\Big)\Big) \tag{31}$$

holds for all $z_i, w_i \in [0,1]$ with $\sum_i w_i = 1$.

Proof. Due to Theorem 19, the mutual information $I_2(z)$ is exponentially concave. By the aid of (10) we infer the statement. ∎

Considering the generalized mutual information

$$I_n(\boldsymbol{z}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots) = H_n\Big(\sum_j z_j \boldsymbol{\gamma}_j\Big) - \sum_j z_j H_n(\boldsymbol{\gamma}_j) \tag{32}$$

depending on the input distribution $\boldsymbol{z}$ for given probability vectors $\boldsymbol{\gamma}_j$. Since $\frac{1}{c_n} H(\boldsymbol{\gamma})$ is exponentially concave as stated in Theorem 17, we can find lower and upper bounds for $I_n(\boldsymbol{z}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots)$ by the aid of Corollaries 2 and 3 as follows.

Corollary 25. For the mutual information I_n of discrete channels, the inequality chain

$$\begin{aligned} &-c_n \sum_i z_i \log\Big(1 + \tfrac{1}{c_n} \nabla H_n\Big(\sum_j z_j \boldsymbol{\gamma}_j\Big)^{\mathrm{T}} \Big(\boldsymbol{\gamma}_i - \sum_j z_j \boldsymbol{\gamma}_j\Big)\Big) \\ &\quad \leq I_n(\boldsymbol{z}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots) \\ &\quad \leq c_n \sum_i z_i \log\Big(1 - \tfrac{1}{c_n} \nabla H_n(\boldsymbol{\gamma}_i)^{\mathrm{T}} \Big(\boldsymbol{\gamma}_i - \sum_j z_j \boldsymbol{\gamma}_j\Big)\Big) \end{aligned} \tag{33}$$

holds for all probability vectors $\boldsymbol{z}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots$.

VI. Side Inequalities from Exponential Concavity

In the previous sections, we have investigated several inequalities and provided different proofs for them, from which we easily can infer the following inequalities.

Proposition 26. The inequality

$$\log^2(x) \leq \frac{(1-x)^2}{x} \tag{34}$$

holds for all $x > 0$ with equality at $x = 1$.

Proof. Since the binary entropy is exponentially concave, the quantity $v(x) = H''(x) + \big(H'(x)\big)^2 = \log^2\big(\frac{1-x}{x}\big) - x^{-1}(1-x)^{-1}$ is non-positive. Moreover, discussing $u(x) = 4 + v(x) \leq 0$ along with its first and second derivatives reveals that actually $u(x)$ is non-positive. Substituting $\frac{1-x}{x}$ by x in $u(x) \leq 0$ yields the inequality under consideration. ∎

Suppose that the weighted arithmetic, geometric, and harmonic means are defined by

$$\mathfrak{A}_{j=1}^n(x_j, w_j) = \sum_{j=1}^n w_j x_j\,, \tag{35}$$

$$\mathfrak{G}_{j=1}^n(x_j, w_j) = \prod_{j=1}^n x_j^{w_j} \tag{36}$$

and

$$\mathfrak{H}_{j=1}^n(x_j, w_j) = \left(\sum_{j=1}^n \frac{w_j}{x_j}\right)^{-1}, \tag{37}$$

for any $\boldsymbol{x} \in \mathcal{R}^n$ and $\boldsymbol{w} \in \mathcal{R}_+^n$ with $\sum_{j=1}^n w_j = 1$. Then the famous inequality chain $\mathfrak{H}_{j=1}^n(x_j, w_j) \leq \mathfrak{G}_{j=1}^n(x_j, w_j) \leq \mathfrak{A}_{j=1}^n(x_j, w_j)$ is well-known, cf. [24]. One can find many improvements for their relationship in the literature. Due to the exponential concavity we have achieved a new mixed mean inequality, which is in the vein of Henrici's, Nanjundiah's and Sierpinski's inequality [25], and is precisely stated in the next proposition.

Proposition 27. Let $\mathbf{X}$ be a real $m \times n$ matrix with nonnegative elements $x_{i,j}$ with $\sum_{j=1}^n x_{i,j} = 1$ for all i. Let $\boldsymbol{w} \in \mathcal{R}_+^m$ with $\sum_{i=1}^m w_i = 1$. Then with c_n from Theorem 17 we have

$$\begin{aligned} &\big[\mathfrak{G}_{j=1}^n\big(\mathfrak{A}_{i=1}^m(x_{i,j}, w_i), \mathfrak{A}_{i=1}^m(x_{i,j}, w_i)\big)\big]^{1/c_n} \\ &\qquad \leq \mathfrak{H}_{i=1}^m\big(\big[\mathfrak{G}_{j=1}^n(x_{i,j}, x_{i,j})\big]^{1/c_n}, w_i\big)\,. \end{aligned} \tag{38}$$

Proof. From Theorem 17, we know the exponential concavity of the scaled entropy. Reminding the relationship $\exp\Big(\sum_k \rho(y_k)\Big) = \prod_k y_k^{-y_k}$, we infer

$$\prod_{j=1}^n \left(\sum_{i=1}^m w_i x_{i,j}\right)^{\frac{1}{c_n} \sum_{i=1}^m w_i x_{i,j}} \leq \left(\sum_{i=1}^m w_i \prod_{j=1}^n x_{i,j}^{-\frac{x_{i,j}}{c_n}}\right)^{-1},$$

from (27) which is equal to (38). ∎

There are many exponentially concave functions, that are important in communication theory. One of the famous ones is the error-function

$$\operatorname{erf}(x) = \tfrac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-t^2}\, \mathrm{d}t, \tag{39}$$

that becomes exponentially concave in the form $\operatorname{erf}(\sqrt{x})$ for all $x \geq 0$. Its complementary usually describes the bit/symbol error probability of communication systems depending on the square root of the underlying signal-to-noise ratio. This example shows that exponentially concave functions can have a crucial role also in communication theory.

VII. Conclusion

Exponentially concave functions seem to play an important role in information theory. Since they are rarely discussed in the literature, we have investigated their mathematical properties along with their general applications. Especially the self-information and the (scaled) discrete entropy have been discussed and it has been shown that they are exponentially concave functions. In addition, we have derived new inequalities for the Kullback-Leibler divergence, the entropy of mixtures of distributions, and the mutual information of discrete channels.

REFERENCES

[1] J. L. W. V. Jensen, "Sur les fonctions convexes et les inégalités entre les valeurs moyennes," *Acta Mathematica*, vol. 30, no. 1, pp. 175–193, 1906.

[2] R. T. Rockafellar, *Convex Analysis*, ser. Princeton landmarks in mathematics and physics. Princeton University Press, 1997.

[3] A. W. Roberts and D. E. Varberg, *Convex Functions*, ser. Pure and Applied Mathematics; a Series of Monographs and Textbooks, 57. Academic Press, 1973.

[4] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, ser. Cambridge Mathematical Library. Cambridge University Press, 1952.

[5] M. Mahdavi, *Exploiting Smoothness in Statistical Learning, Sequential Prediction, and Stochastic Optimization*. East Lansing, MI, USA: Michigan State University, 2014.

[6] M. Mahdavi, L. Zhang, and R. Jin, "Lower and upper bounds on the generalization of stochastic exponentially concave optimization," in *Proceedings of The 28th Conference on Learning Theory*. Journal of Machine Learning Research (JMLR), 2015, pp. 1305–1320.

[7] S. Pal, "Exponentially concave functions and high dimensional stochastic portfolio theory," *ArXiv e-prints*, Mar. 2016.

[8] T. Koren and K. Y. Levy, "Fast rates for exp-concave empirical risk minimization," in *Proceedings of the 28th International Conference on Neural Information Processing Systems*, ser. NIPS'15. Cambridge, MA, USA: MIT Press, 2015, pp. 1477–1485.

[9] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, 2nd ed. New York: Springer, 2011.

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

[11] S. Guiaşu, "Weighted entropy," *Reports on Mathematical Physics*, vol. 2, no. 3, pp. 165–179, 1971.

[12] Z. Daróczy, "Über Mittelwerte und Entropien vollständiger Wahrscheinlichkeitsverteilungen," *Acta Mathematica Hungarica*, vol. 15, pp. 203–210, 1964.

[13] I. Csiszár and J. Körner, *Information Theory, Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[14] G. Alirezaei and R. Mathar, "Optimum one-bit quantization," in *The IEEE Information Theory Workshop (ITW'15)*, Jeju Island, Korea, Oct. 2015.

[15] ——, "An upper bound on the capacity of censored channels," in *The 9th International Conference on Signal Processing and Communication Systems (ICSPCS'15)*, Cairns (Barrier Reef), Australia, Dec. 2015.

[16] ——, "On the information capacity of hinge functions," in *The IEEE International Symposium on Information Theory and Its Applications (ISITA'16)*, Monterey, California, USA, 2016.

[17] A. Behboodi, G. Alirezaei, and R. Mathar, "On the capacity of censored channels," in *11th International ITG Conference on Systems, Communications and Coding (SCC'17)*, Hamburg, Germany, Feb. 2017.

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley-Interscience, 1991.

[19] V. Jog and V. Anantharam, "On the geometry of convex typical sets," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 829–833.

[20] T. A. Courtade, "Strengthening the entropy power inequality," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 2294–2298.

[21] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, July 1948.

[22] S. Muroga, "On the capacity of a discrete channel I," *Journal of the Physical Society of Japan*, vol. 8, no. 4, pp. 484–494, 1953.

[23] R. Silverman, "On binary channels and their cascades," *IRE Transactions on Information Theory*, vol. 1, no. 3, pp. 19–27, December 1955.

[24] P. S. Bullen, *Handbook of Means and Their Inequalities*, 2nd ed., ser. Mathematics and Its Applications. Springer Netherlands, 2003.

[25] ——, *Dictionary of Inequalities*, 2nd ed., ser. Chapman & Hall/CRC Monographs and Research Notes in Mathematics. CRC Press, 2015.