Tsallis divergence and superadditivity

Thomas Oikonomou and G. Baris Bagci

College of Engineering and Computer Science, VinUniversity, Hanoi 10000, Vietnam and
Department of Physics, Mersin University, Mersin 33343, Turkey

(Dated: May 25, 2021)

The well-known Kullback-Leibler divergence is known to be superadditive, enabling its use in many information-theoretic applications. However, such a proof is still missing for Tsallis divergence to the best of our knowledge. In this work, we first prove that Tsallis divergence is $\alpha$-superadditive for $\alpha > 1$. Then, we show that this fact entails ordinary superadditivity in the same aforementioned interval. This opens up a novel route to many future applications of Tsallis divergence in information theory.

PACS numbers:
Keywords: Kullback-Leibler divergence; Tsallis divergence; superadditivity; nonadditivity; deformed logarithm

I. INTRODUCTION

Shannon entropy as an information measure has widespread applications in many diverse fields. However, it cannot be used in the continuum case, since it leads to erroneous results. To this aim, one generalizes Shannon entropy so as to obtain Kullback-Leibler (K-L) divergence which is valid for both discrete and continuous cases [1]. Apart from this, K-L divergence can also be used as a second law in phase space. The quantum divergence, on the other hand, measures the distinguishability of quantum states where these states are given by their concomitant density matrices.

K-L divergence is inherently based on the logarithmic function just as Shannon entropy. Recent progress in the generalized entropies showed that one can use deformed logarithms as the underlying fundamental structure. This change yields the Tsallis entropy and divergence [2–5]. K-L and Tsallis divergences have non negativity, monotonicity, and joint convexity as their common features [6, 7]. However, K-L divergence is additive whereas Tsallis divergence is not. Another important issue is superadditivity: K-L divergence is proven to be superadditive while there is no such proof for Tsallis case. Since superadditivity is used in many applications of information entropy, it is of great need to assess whether Tsallis divergence has this property or not.

In Section II, we first provide the rudiments necessary for the proof of superadditivity. Section III is devoted to the proof of superadditivity of Tsallis divergence. Finally, concluding remarks are presented in section IV.

II. DEFINITIONS AND PROPERTIES

In this section, we provide some definitions and relations which will be useful in proving the superadditivity of Tsallis divergence. In this regard, the deformed $\alpha$-logarithmic function reads [3]

$$\ln_\alpha (x) := \frac{x^{1-\alpha} - 1}{1-\alpha},$$

(1)

which is concave for $\alpha > 0$ and convex for $\alpha < 0$. For $\alpha = 0$ is both convex and concave, since $\ln_0(x) = x - 1$ is a linear function. In particular, the deformed $\alpha$-logarithm satisfies the following:

$$\ln_\alpha (xy) = \ln_\alpha (x) + \ln_\alpha (y) + (1-\alpha) \ln_\alpha (x) \ln_\alpha (y), \quad \alpha \in \mathbb{R}, \quad x, y > 0,$$

(2a)

$$\ln_\alpha (x/y) = y^{\alpha-1} \left[ \ln_\alpha (x) - \ln_\alpha (y) \right], \quad \alpha \in \mathbb{R}, \quad x, y > 0,$$

(2b)

$$\ln_\alpha (x) = -x^{1-\alpha} \ln_\alpha (1/x), \quad \alpha \in \mathbb{R}, \quad x > 0,$$

(2c)

$$\ln_\alpha (x) \leq \ln(x) \leq x - 1, \quad \alpha \geq 1, \quad x > 0.$$  

(2d)
Tsallis relative entropy $S^\alpha_p (p|q)$ and the concomitant divergence $D^\alpha_p (p||q)$ are given by [5]

$$
S^\alpha_p (p|q) = \sum_{i=1}^{n} p_i \ln\left( \frac{q_i}{p_i} \right), \quad
D^\alpha_p (p||q) = \frac{\text{sgn}(\alpha)}{\alpha - 1} \left[ \sum_{i=1}^{n} \left( \frac{p_i}{q_i} \right)^\alpha q_i - 1 \right] = -\text{sgn}(\alpha) S^\alpha_p (p||q).
$$

When one has product states as the arguments with the probability vectors $p, q \in \mathbb{R}^n$ and $r, s \in \mathbb{R}^m$, Tsallis divergence yields the following relation

$$
D^\alpha_p (p \otimes r||q \otimes s) = D^\alpha_p (p||q) + D^\alpha_r (r||s) + \text{sgn}(\alpha)(\alpha - 1)D^\alpha_p (p||q)D^\alpha_r (r||s)
$$

for $\alpha \in \mathbb{R}$ which can be shown by using Eq. (3). We also note the following relations regarding Tsallis divergence

Invoking the Jensen’s inequality for convex or concave functions, i.e., $\mathbb{E}\{f(X)\} \geq f(\mathbb{E}\{X\})$ or $\mathbb{E}\{f(X)\} \leq f(\mathbb{E}\{X\})$, respectively, one can prove

$$
D^\alpha_p (p||q) \geq 0, \quad \forall \alpha \in \mathbb{R}.
$$

Then, from Eqs. (4) and (5) we read

$$
D^\alpha_p (p||q) \in [0, \infty), \quad \alpha \in (-\infty, 0) \cup [1, \infty),
$$

$$
D^\alpha_p (p||q) \in \left[0, \frac{1}{1-\alpha}\right], \quad \alpha \in [0, 1).
$$

III. $\alpha$-SUPERADDITIVITY OF TSALLIS DIVERGENCE

We now consider the probability vectors $q_A = \{q_i\}_{i=1}^{n}, q_B = \{q_j\}_{j=1}^{m}, p_A = \{p_i\}_{i=1}^{n}, p_B = \{p_j\}_{j=1}^{m}$. We also have $p_{AB} = \{p_{ij}\}_{i=1,j=1}^{n,m}$ with $p_i = \sum_j p_{ij}$ and $p_j = \sum_i p_{ij}$ while $p_{ij} \neq p_i p_j$ in general. Before proceeding further, we need to determine the sign of the following expression for later use

$$
\mathcal{A} := -\sum_{ij} (p_{ij} - p_i p_j) \ln\left( \frac{q_i q_j}{p_i p_j} \right).
$$

The sign of the term $(p_{ij} - p_i p_j)$ for some values of $(i, j)$ maybe positive, negative or zero. Therefore, we can split the above summation into two parts, the negative and the non-negative values of $(p_{ij} - p_i p_j)$, $p_{ij} < p_i p_j$ and $p_{ij} \geq p_i p_j$, denoted by $(p_{ij} - p_i p_j)_{-}$ and $(p_{ij} - p_i p_j)_{+}$, respectively. Then, we can write $\mathcal{A}$ as

$$
\mathcal{A} = -\sum_{ij} (p_{ij} - p_i p_j)_{+} \ln\left( \frac{q_i q_j}{p_i p_j} \right) - \sum_{ij} (p_{ij} - p_i p_j)_{-} \ln\left( \frac{q_i q_j}{p_i p_j} \right)
$$

However, we cannot yet determine the sign of $\mathcal{A}$, since the $\alpha$-logarithm admits values in $\mathbb{R}$. Therefore, we introduce two non-negative functions,

$$
f_\alpha(x) := \ln(x) - \ln\left( x \right), \quad \alpha \geq 1, \quad x > 0,
$$

$$
g_\alpha(x,y) := \ln\left( xy \right) - x^{1-\alpha} \ln\left( x \right) - y^{1-\alpha} \ln\left( y \right), \quad \alpha \geq 1, \quad x, y > 0.
$$

The non-negativity of $f_\alpha$ can be easily proven through the well-known inequality $\ln(z) \leq z - 1$ for $z > 0$ with $z = x^{1-\alpha}$. To prove non-negativity of $g_\alpha$, we fix $y = y_0$ and calculate the critical points of $g_\alpha(x,y_0)$. Doing so we obtain a single critical point $x_0 = [(1 + y_0^{-1-\alpha})/2]^{1/(1-\alpha)}$. Calculating the second derivative we observe that $g_\alpha''(x_0,y_0) \geq 0$, which means that $x_0$ is a global minimum. Since $y_0$ can be any point in the $y$-domain, we have $g_\alpha(x,y) \geq g_\alpha(x_0,y_0) \geq 0$ where the equality holds for $\alpha = 1$. Then, using Eq. (9) we can rewrite $\mathcal{A}$ as

$$
\mathcal{A} = \sum_{ij} (p_{ij} - p_i p_j)_{+} f\left( \frac{q_i q_j}{p_i p_j} \right) - \sum_{ij} (p_{ij} - p_i p_j)_{-} g\left( \frac{q_i}{p_i}, \frac{q_j}{p_j} \right) \geq 0.
$$

Eqs. (8) and (10) are equivalent because of the relation $\sum_{ij} (p_{ij} - p_i p_j)(x_i - y_j) = 0$ for any $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. Hence, we proved that $\mathcal{A}$ is non-negative.
Having determined the sign of $\mathcal{A}$, we now proceed to study Tsallis divergence for correlated states, given by

$$D^\alpha_{\mathcal{A}}(p_A \otimes q_A) = - \sum_{ij} p_{ij} \ln_\alpha \left( \frac{q_{ij}}{p_{ij}} \right).$$

Then, we have

$$D^\alpha_{\mathcal{A}}(p_A \otimes q_A \otimes q_B) = - \sum_{ij} p_{ij} \ln_\alpha \left( \frac{q_{ij}}{p_{ij}} \right) + \sum_{ij} p_{ij} \left[ \ln_\alpha \left( \frac{q_{ij}}{p_{ij}} \right) - \ln_\alpha \left( \frac{q_{ij}}{p_{ij}} \right) \right]$$

$$= - \sum_{ij} p_{ij} \ln_\alpha \left( \frac{q_{ij}}{p_{ij}} \right) + \sum_{ij} p_{ij} \left( \frac{q_{ij}}{p_{ij}} \right)^{1-\alpha} \ln_\alpha \left( \frac{p_{ij}}{p_{ij}} \right)$$

$$= - \sum_{ij} p_{ij} \ln_\alpha \left( \frac{q_{ij}}{p_{ij}} \right) - \sum_{ij} p_{ij} \left( \frac{q_{ij}}{p_{ij}} \right)^{1-\alpha} \ln_\alpha \left( \frac{p_{ij}}{p_{ij}} \right)$$

$$\geq - \sum_{ij} p_{ij} \ln_\alpha \left( \frac{q_{ij}}{p_{ij}} \right) + \sum_{ij} p_{ij} \left( \frac{q_{ij}}{p_{ij}} \right)^{1-\alpha} \left( 1 - \frac{p_{ij}}{p_{ij}} \right)$$

$$= - \sum_{ij} p_{ij} \ln_\alpha \left( \frac{q_{ij}}{p_{ij}} \right) + (1 - \alpha) \sum_{ij} p_{ij} \ln_\alpha \left( \frac{q_{ij}}{p_{ij}} \right) - (1 - \alpha) \sum_{ij} p_{ij} \ln_\alpha \left( \frac{q_{ij}}{p_{ij}} \right)$$

$$= - \sum_{ij} p_{ij} \ln_\alpha \left( \frac{q_{ij}}{p_{ij}} \right) - \alpha \sum_{ij} (p_{ij} - p_{ij}) \ln_\alpha \left( \frac{q_{ij}}{p_{ij}} \right)$$

$$\geq - \sum_{ij} p_{ij} \ln_\alpha \left( \frac{q_{ij}}{p_{ij}} \right) = D^\alpha_{\mathcal{A}}(p_A \otimes p_B \| q_A \otimes q_B)$$

for $\alpha \geq 1$. Note that we have used Eqs. (2b), (2c) and (2d) in Eqs. (11b), (11c) and (11e), respectively. Hence, we have shown $D^\alpha_{\mathcal{A}}(p_A \otimes q_A \otimes q_B) \geq D^\alpha_{\mathcal{A}}(p_A \otimes p_B \| q_A \otimes q_B)$ for $\alpha \geq 1$. We call this relation as $\alpha$-superadditivity. Note that it also entails usual superadditivity, since we have

$$D^\alpha_{\mathcal{A}}(p_A \otimes q_A \otimes q_B) \geq D^\alpha_{\mathcal{A}}(p_A \otimes p_B \| q_A \otimes q_B)$$

$$= D^\alpha_{\mathcal{A}}(p_A \| q_A) + D^\alpha_{\mathcal{A}}(p_B \| q_B) + (\alpha - 1)D^\alpha_{\mathcal{A}}(p_A \| q_A)D^\alpha_{\mathcal{A}}(p_B \| q_B)$$

$$\geq D^\alpha_{\mathcal{A}}(p_A \| q_A) + D^\alpha_{\mathcal{A}}(p_B \| q_B)$$

for $\alpha \geq 1$. It is worth noting that we have used Eq. (4) in Eq. (12b).

### IV. CONCLUSIONS

Superadditivity is important for many applications in information theory. So far, quantum version of K-L divergence has been known to be superadditive so that it has been the sole candidate for such applications in general. In this work, we have proven that Tsallis divergence is superadditive for $\alpha \geq 1$. Note that this result also includes the K-L divergence in the limit $\alpha \to 1$. Hence, one can consider Tsallis divergence as a suitable generalization whenever the superadditivity property is required for a particular application.

[1] E. T. Jaynes, *Probability Theory: The Logic of Science*, (Cambridge University Press, New York, 2003).
[2] C. Tsallis, *Possible Generalization of Boltzmann-Gibbs Statistics*, J. Stat. Phys. **52** (1988) 479.
[3] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*, (Springer, New York, 2009).
[4] S. Abe and G. B. Bagci, *Necessity of $q$-expectation value in nonextensive statistical mechanics*, Phys. Rev. E **71** (2005) 016139.
[5] T. Oikonomou and G. B. Bagci, *Route from discreteness to the continuum for the Tsallis q-entropy*, Phys. Rev. E **97** (2018) 012104.

[6] R. F. Vigelas, L. H. F. de Andrade and C. C. Cavalcante, *Properties of a Generalized Divergence Related to Tsallis Generalized Divergence*, IEEE Trans. on Inf. Th. **66** (2020) 2891.

[7] Á. Capel, A. Lucia, and D. Pérez-García, *Superadditivity of Quantum Relative Entropy for General States*, IEEE Trans. on Inf. Th. **64** (2018) 4758.