J Stat Phys (2007) 128: 1177–1188
DOI 10.1007/s10955-007-9349-2

Notes on Superadditivity of Wigner–Yanase–Dyson Information

Shunlong Luo

Received: 5 August 2006 / Accepted: 23 May 2007 / Published online: 27 June 2007

Abstract The classical Fisher information is superadditive in the sense that the Fisher information of a bivariate probability density is always not less than the sum of those of the marginals. The longstanding conjecture concerning the superadditivity of the Wigner–Yanase–Dyson information is a quantum analogue of this property. It is remarkable that Hansen constructed a numerical counterexample to the quantum case (J. Stat. Phys. 126: 643–648, 2007). However, the requirement of superadditivity of an information-theoretic quantity such as the Wigner–Yanase–Dyson information seems so intuitive, it is desirable to identify conditions as general as possible such that the superadditivity holds. In this paper, we establish the superadditivity in several physically significant cases.

Keywords Fisher information · Wigner–Yanase–Dyson information · Superadditivity · Density operator · Spectral decomposition

1 Introduction

Apart from the Shannon entropy (and the von Neumann entropy in the quantum case), the Fisher information is another well known information concept [2]. This notion has its origin in statistical inference [3], and now has even found deep and wide applications in studying probability limit theorems [7] and in an informational approach to physics [4].

Recall that the Fisher information of a probability density $p(x) = p(x_1, x_2, \dots, x_n)$ (with respect to the location parameters) is defined as [1]

$$I_{\mathrm{F}}(p) := 4 \int_{R^n} |\nabla \sqrt{p(x)}|^2 dx \tag{1}$$

provided the integral exists. Here ∇ denotes the gradient and $|\cdot|$ denotes the usual Euclidean norm in R^n.

S. Luo (✉)
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, 100080 Beijing, China
e-mail: luosl@amt.ac.cn

More generally, the Fisher information *matrix* of a parametric probability densities $p_\theta(x)$ on R^n with parameter $\theta = (\theta_1, \theta_2, \ldots, \theta_m) \in R^m$ is the $m \times m$ matrix

$$\mathbf{I}_F(p_\theta) = (I_{ij})$$

defined as [2]

$$I_{ij} := 4 \int_{R^n} \frac{\partial \sqrt{p_\theta(x)}}{\partial \theta_i} \frac{\partial \sqrt{p_\theta(x)}}{\partial \theta_j} dx, \quad i, j = 1, 2, \ldots, m$$

provided the integrals exist. In particular, when $n = m$ and $p_\theta(x) = p(x - \theta)$ is a translation family, $\mathbf{I}_F(p_\theta) = (I_{ij})$ is independent of the parameter θ, and moreover

$$I_{ij} = 4 \int_{R^n} \frac{\partial \sqrt{p(x)}}{\partial x_i} \frac{\partial \sqrt{p(x)}}{\partial x_j} dx, \quad i, j = 1, 2, \ldots, n. \tag{2}$$

In this case, we may simply denote $\mathbf{I}_{\mathrm{F}}(p_\theta)$ by $\mathbf{I}_{\mathrm{F}}(p)$. We see that the Fisher information defined by (1) is the trace of the Fisher information matrix determined by (2), that is,

$$I_{\mathrm{F}}(p) = \mathrm{tr}\, \mathbf{I}_{\mathrm{F}}(p).$$

Some fundamental properties of the Fisher information are as follows.

(a) The Fisher information is convex in the sense that [6] (p. 79)

$$I_{\mathrm{F}}(\lambda_1 p_1 + \lambda_2 p_2) \le \lambda_1 I_{\mathrm{F}}(p_1) + \lambda_2 I_{\mathrm{F}}(p_2).$$

Here p_1 and p_2 are two probability densities and $\lambda_1 + \lambda_2 = 1,\ \lambda_j \ge 0,\ j = 1, 2$.

(b) The Fisher information is additive in the sense that

$$I_{\mathrm{F}}(p_1 \otimes p_2) = I_{\mathrm{F}}(p_1) + I_{\mathrm{F}}(p_2).$$

Here p_1 and p_2 are two probability densities, and $p_1 \otimes p_2(x, y) := p_1(x) p_2(y)$ is the independent product density (which is a kind of tensor product).

(c) The Fisher information is invariant under location translation, that is, for any fixed $y \in R^n$, if we put $p_y(x) := p(x - y)$, then $I_{\mathrm{F}}(p_y) = I_{\mathrm{F}}(p)$.

(d) The Fisher information $I_{\mathrm{F}}(p)$ is superadditive in the sense that [1]

$$I_{\mathrm{F}}(p) \ge I_{\mathrm{F}}(p_1) + I_{\mathrm{F}}(p_2). \tag{3}$$

Here $p(x) = p(x_1, x_2)$ is a bivariate probability density with marginal densities p_1 and p_2, respectively.

Although the Fisher information was introduced as early as 1925 by Fisher [3], it is amusing and remarkable that the above superadditive property (d) was discovered as late as 1991 by Carlen [1]. Kagan and Landsman gave an intuitive statistical proof of this inequality [8], while Carlen's original derivation is analytical. The statistical meaning of the above inequality is as follows: When a composite system is decomposed into two subsystems, the correlation information between them is missing, and thus the Fisher information decreases.

In this paper, we are interested in a quantum analogue of the above scenario. In particular, we are concerned with a longstanding open problem related to a quantum generalization of the Fisher information, viz., the superadditivity conjecture for the Wigner–Yanase–Dyson information, whose mathematical definition will be given shortly. First of all, this open problem is resolved by Hansen negatively who provides a numerical counterexample quite recently [5]. However, the superaditivity is such an intuitive requirement for quantum Fisher information, it is desirable to identify further conditions under which the superadditivity survives. We will establish this conjecture under some natural conditions with physical significance.

To establish notion and notation, and to put our discussion in a statistical perspective, let us first briefly review some mathematical aspects of quantum mechanics from the quantum probability viewpoint.

The state space of a quantum system is described by a *complex* Hilbert space $\mathcal{H}$. We will use Dirac's notation which is particularly convenient in manipulating matrix elements of quantum mechanical operators and is universally employed in physics literature [14]. Thus a generic vector in $\mathcal{H}$ is denoted by a so-called ket $|\psi\rangle$, which represents a pure state of the system. The corresponding linear functional on $\mathcal{H}$ is denoted by the so-called bra $\langle\psi|$. The inner product (complex linear in the second variable) between two vectors $|\psi\rangle, |\phi\rangle$ is denoted by $\langle\psi|\phi\rangle$ (which is the pairing between the bra $\langle\psi|$ and the ket $|\phi\rangle$, thus also the name bracket). On the other hand, $|\psi\rangle\langle\psi|$ denotes the orthogonal projection onto the vector $|\psi\rangle$. A mixed state is represented by a non-negative operator on $\mathcal{H}$ with unit trace, this is the quantum analogue of a probability density, and is usually called a density operator. In particular, if $|\psi\rangle \in \mathcal{H}$ is normalized, then $|\psi\rangle\langle\psi|$ is a degenerate density operator. An observable, which is the quantum analogue of a classical real-valued random variable, is represented by a self-adjoint operator on $\mathcal{H}$. For any self-adjoint operator A, the inner product between $|\psi\rangle$ and $A|\phi\rangle$ is denoted by $\langle\psi|A|\phi\rangle$.

Now we review a remarkable quantum extension of the classical Fisher information. In 1963, when studying the theory of quantum measurement, Wigner and Yanase [17] proposed to use the quantity

$$I(\rho, H) := -\frac{1}{2}\,\mathrm{tr}[\rho^{1/2}, H]^2$$

as the amount of information on the values of observables *not* commuting with H. Here tr denotes trace, the square bracket denotes commutator (that is, $[A, B] = AB - BA$ for any two operators), ρ is a density operator, and H is a fixed self-adjoint operator serving as a conserved observable. They called this quantity skew information, and proved and conjectured that it satisfies many desirable intuitive requirements for an information-theoretic measure such as those quantum analogues of (a)–(d) of the classical Fisher information.

As mentioned by Wigner and Yanase [17], Dyson suggested a more general quantity

$$I_\alpha(\rho, H) := -\frac{1}{2}\,\mathrm{tr}[\rho^\alpha, H][\rho^{1-\alpha}, H], \quad 0 < \alpha < 1 \tag{4}$$

to measure information content of ρ with respect to H (we have put the constant factor $\frac{1}{2}$ before the trace for convenience of comparison with the skew information). This quantity, now called the *Wigner–Yanase–Dyson information*, has many similar properties as the skew information $I(\rho, H)$. Its convexity with respect to ρ has been studied extensively by Lieb [10], Uhlmann [15], and Kosaki [9], among others. Clearly, when $\alpha = \frac{1}{2}$, $I_\alpha(\rho, H)$ is the skew information.

In order for $I_\alpha(\rho, H)$ to be a reasonable information-theoretic measure, the following superadditive property seems a necessary requirement, as advocated by Wigner and Yanase [17].

Conjecture (Superadditivity of the Wigner–Yanase–Dyson Information) For any bipartite density operator ρ, it holds that

$$I_\alpha(\rho, H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2) \geq I_\alpha(\rho_1, H_1) + I_\alpha(\rho_2, H_2). \tag{5}$$

Here $\rho_1 = \mathrm{tr}_2\, \rho$, $\rho_2 = \mathrm{tr}_1\, \rho$ are the partial trace of ρ over the first and the second system, respectively, H_1 and H_2 are observables over the first and the second system, respectively. $\mathbf{1}$ is the identity operator (depending on the system), and $\otimes$ denotes tensor product of operators. This conjecture was reviewed by Lieb [10]. The only non-trivial confirmed case is for pure states with $\alpha = \frac{1}{2}$ [17].

Quite remarkably, this conjecture cannot be true in its most general form as demonstrated recently by Hansen [5] who provides a numerical counterexample! However, this does not nullify the significant requirement of superadditivity, and in contrast to Hansen's counterexample, we review an intuitive relation between the Wigner–Yanase–Dyson information and the Fisher information in order to reveal some insight into the conjecture, and prove the conjecture in the following three particular cases:

(1) The density operator ρ is pure. This generalizes the result of Wigner and Yanase [17].
(2) The density operator ρ is diagonal in a particular representation.
(3) The density operator ρ commutes with either $H_1 \otimes \mathbf{1}$ or $\mathbf{1} \otimes H_2$.

The remaining part of this paper is structured as follows. In Sect. 2, we review the analogy between the Fisher information and the Wigner–Yanase–Dyson information. In Sect. 3, we prove the superadditivity in the three mentioned cases. Finally, Sect. 4 is devoted to discussions.

2 Analogy with Fisher Information

In order to gain some intuition about the superadditivity of the Wigner–Yanase–Dyson information, let us first illustrate how the Wigner–Yanase–Dyson information can be heuristically interpreted as a quantum analogue of the classical Fisher information [11–13], and compare its properties with the corresponding properties of the Fisher information.

To see how Wigner–Yanase–Dyson information arises naturally when considering quantum analogues of the classical Fisher information, we rewrite the classical Fisher information defined by (1) as

$$I_{\mathrm{F}}(p) = \frac{1}{\alpha(1-\alpha)} \int_{R^n} \nabla p^\alpha(x) \cdot \nabla p^{1-\alpha}(x)dx, \quad \alpha \in (0,1).$$

Clearly, the above equation is equivalent to (1).

Following the standard procedure of passing from classical to quantum and formally replacing the integration by the trace, the probability density p by a density operator ρ, the ordinary differentiation ∇ by the quantum (inner) differentiation $D_H(\cdot) = i[\cdot, H]$ (that is, for any operator A, $D_H A = i[A, H]$, here i is the imaginary unit), we come to the following quantum analogue

$$\frac{1}{\alpha(1-\alpha)} \mathrm{tr}\, D_H \rho^\alpha \cdot D_H \rho^{1-\alpha} = -\frac{1}{\alpha(1-\alpha)} \mathrm{tr}[\rho^\alpha, H][\rho^{1-\alpha}, H],$$

which is, up to a constant factor, essentially the Wigner–Yanase–Dyson information defined by (4). Accordingly, the Wigner–Yanase–Dyson information may be regarded as a natural generalization of the classical Fisher information.

More generally, for a parametric family of density operators ρ_θ, $\theta \in R$, we may define

$$I_\alpha(\rho_\theta) = \frac{1}{2}\operatorname{tr}\left(\frac{\partial \rho_\theta^\alpha}{\partial\theta}\frac{\partial \rho_\theta^{1-\alpha}}{\partial\theta}\right), \quad \alpha \in (0,1)$$

as a generalization of the Wigner–Yanase–Dyson information. In particular, when ρ_θ is determined by the Landau-von Neumann equation $\frac{\partial\rho_\theta}{\partial\theta} = D_H\rho_\theta$, $\rho_0 = \rho$, that is, $\rho_\theta = e^{-i\theta H}\rho e^{i\theta H}$, then $I_\alpha(\rho_\theta) = I_\alpha(\rho, H)$.

Following Wigner and Yanase [17], and Lieb [10], we know that the Wigner–Yanase–Dyson information has the following properties which are quantum (non-commutative) analogues of the properties (a)–(d) of the classical Fisher information.

(A) The Wigner–Yanase–Dyson information decreases when two different density operators are mixed, since by mixing, one "forgets" from which density operator a particular sample stems. In mathematical terms,

$$I_\alpha(\lambda_1\rho_1 + \lambda_2\rho_2, H) \le \lambda_1 I_\alpha(\rho_1, H) + \lambda_2 I_\alpha(\rho_2, H).$$

Here ρ_1 and ρ_2 are two density operators, and $\lambda_1 + \lambda_2 = 1$, $\lambda_1, \lambda_2 \ge 0$.

(B) The Wigner–Yanase–Dyson information is additive in the sense that the information content of a system composed of two independent parts is the sum of information of the parts. More precisely,

$$I_\alpha(\rho_1 \otimes \rho_2, H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2) = I_\alpha(\rho_1, H_1) + I_\alpha(\rho_2, H_2).$$

Here ρ_1 and ρ_2 are density operators for the first and the second system, respectively, H_1 and H_2 are observables for the first and the second system, respectively.

(C) The Wigner–Yanase–Dyson information remains constant for isolated systems, that is, it is invariant as long as the state changes according to the Landau-von Neumann equation: if $i\frac{\partial}{\partial\theta}\rho_\theta = [H, \rho_\theta]$, $\rho_0 = \rho$, $\theta \in R$, then $I_\alpha(\rho_\theta, H) = I_\alpha(\rho, H)$. More generally, $I_\alpha(\rho, H) = I_\alpha(U\rho U^\dagger, H)$ whenever the unitary operator U commutes with H. Here $\dagger$ denotes adjoint of an operator.

(D) For any density operator ρ which is *pure* (that is, $\rho = |\Psi\rangle\langle\Psi|$ for some vector $|\Psi\rangle$ in the composite Hilbert space), it holds that

$$I(\rho, H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2) \ge I(\rho_1, H_1) + I(\rho_2, H_2).$$

Here $I(\rho, H) = I_{1/2}(\rho, H)$ is the skew information, $\rho_1 = \operatorname{tr}_2 \rho$, $\rho_2 = \operatorname{tr}_1 \rho$ are the partial trace of ρ over the first and the second system, respectively, H_1 and H_2 are observables over the first and the second system, respectively. Recall that the partial trace $\rho_1 = \operatorname{tr}_2 \rho$ is defined via the following equation

$$\langle\psi|\rho_1|\phi\rangle = \sum_j \langle\psi| \otimes \langle h_j|\rho|\phi\rangle \otimes |h_j\rangle.$$

Here $|\psi\rangle$, $|\phi\rangle$ are any two vectors in the Hilbert space of the first system, and $\{|h_j\rangle\}$ is any orthonormal base for the Hilbert space of the second system. This definition is independent of the choice $\{|h_j\rangle\}$. The partial trace $\rho_1 = \operatorname{tr}_2 \rho$ is defined similarly.

3 Superadditivity under Further Conditions

We will work exclusively in finite dimensional complex spaces for simplification. For later applications, we evaluate explicitly the expression of $I_\alpha(\rho, H)$ in a representation when the density operator ρ is diagonal.

Lemma 1 *If ρ has eigenvalues $\{\lambda_j\}$ with corresponding normalized eigenvectors $\{|\psi_j\rangle\}$ (which, without loss of generality, may be assumed to constitute an orthonormal base. Otherwise, we can add vectors corresponding to the zero eigenvalue to form a base), then*

(1) $I_\alpha(\rho, H) = \sum_{i,j}(\lambda_i - \lambda_i^\alpha \lambda_j^{1-\alpha})|\langle\psi_i|H|\psi_j\rangle|^2$.
(2) $I_\alpha(\rho, H) = \frac{1}{2}\sum_{i,j}(\lambda_i^\alpha - \lambda_j^\alpha)(\lambda_i^{1-\alpha} - \lambda_j^{1-\alpha})|\langle\psi_i|H|\psi_j\rangle|^2$.

Proof (1) Clearly,

$$I_\alpha(\rho, H) = \operatorname{tr}\rho H^2 - \operatorname{tr}\rho^\alpha H \rho^{1-\alpha} H.$$

Since $\{|\psi_j\rangle\}$ is an orthonormal base, we have the resolution of identity

$$\sum_j |\psi_j\rangle\langle\psi_j| = \mathbf{1}.$$

By using this identity repeatedly, we have

$$\begin{aligned}
\operatorname{tr}\rho H^2 &= \sum_i \langle\psi_i|\rho H^2|\psi_i\rangle = \sum_{i,j}\langle\psi_i|\rho|\psi_j\rangle\langle\psi_j|H^2|\psi_i\rangle \\
&= \sum_{i,j}\lambda_j\langle\psi_i|\psi_j\rangle\langle\psi_j|H^2|\psi_i\rangle \\
&= \sum_i \lambda_i\langle\psi_i|H^2|\psi_i\rangle \quad (\text{since } \langle\psi_i|\psi_j\rangle = \delta_{ij}) \\
&= \sum_{i,j}\lambda_i\langle\psi_i|H|\psi_j\rangle\langle\psi_j|H|\psi_i\rangle = \sum_{i,j}\lambda_i|\langle\psi_i|H|\psi_j\rangle|^2,
\end{aligned}$$

and similar manipulation yields

$$\operatorname{tr}\rho^\alpha H\rho^{1-\alpha}H = \sum_{i,j}\lambda_i^\alpha\lambda_j^{1-\alpha}|\langle\psi_i|H|\psi_j\rangle|^2.$$

The conclusion follows. □

(2) This is readily derived from (1) by a symmetry argument.
The next lemma follows from Corollary 1.3 in Lieb [10].

Lemma 2 *The Wigner–Yanase–Dyson information decreases under the operation of partial trace in the following sense*

$$I_\alpha(\rho, H_1 \otimes \mathbf{1}) \geq I_\alpha(\rho_1, H_1).$$

Here ρ is a density operator of a composite quantum system and $\rho_1 = \operatorname{tr}_2\rho$ is the partial trace of ρ over the second system, and H_1 is an observable for the first system, $\mathbf{1}$ is the identity operator for the second system.

The following statement, which establishes the superadditivity of the Wigner–Yanase–Dyson information in an important case, is a generalization of the case $\alpha = \frac{1}{2}$ proved by Wigner and Yanase [17].

Proposition 1 *Let $H = H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2$. If $\rho = |\Psi\rangle\langle\Psi|$ is a pure state on the composite quantum system, then inequality* (5) *holds true, that is*

$$I_\alpha(\rho, H) \geq I_\alpha(\rho_1, H_1) + I_\alpha(\rho_2, H_2).$$

Proof Suppose that $|\Psi\rangle$ has the following Schmidt decomposition

$$|\Psi\rangle = \sum_i \sqrt{\lambda_i}|\psi_i^{(1)}\rangle \otimes |\psi_i^{(2)}\rangle.$$

Here $\{|\psi_i^{(1)}\rangle\}$ and $\{|\psi_i^{(2)}\rangle\}$ are orthonormal bases for the first and the second system, respectively, and $\sum_i \lambda_i = 1$, $\lambda_i \geq 0$, $i = 1, 2, \ldots$. If the two component systems are of different dimensions, we add zero vectors to equal the index. Then

$$\rho_1 = \operatorname{tr}_2 \rho = \sum_i \lambda_i |\psi_i^{(1)}\rangle\langle\psi_i^{(1)}|, \qquad \rho_2 = \operatorname{tr}_1 \rho = \sum_i \lambda_i |\psi_i^{(2)}\rangle\langle\psi_i^{(2)}|.$$

Since $\rho = |\Psi\rangle\langle\Psi|$ is a pure state, we have $\sqrt{\rho} = \rho$, and thus

$$I_\alpha(\rho, H) = \langle\Psi|H^2|\Psi\rangle - \langle\Psi|H|\Psi\rangle^2.$$

Now put

$$a_{ij} = \langle\psi_i^{(1)}|H_1|\psi_j^{(1)}\rangle, \qquad b_{ij} = \langle\psi_i^{(2)}|H_2|\psi_j^{(2)}\rangle,$$

then

$$\begin{aligned}
&\langle\Psi|(H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2)^2|\Psi\rangle \\
&\quad = \sum_{i,j} \sqrt{\lambda_i\lambda_j}\langle\psi_i^{(1)}| \otimes \langle\psi_i^{(2)}|H_1^2 \otimes \mathbf{1} + \mathbf{1} \otimes H_2^2 + 2H_1 \otimes H_2|\psi_j^{(1)}\rangle \otimes |\psi_j^{(2)}\rangle \\
&\quad = \sum_{i,j} \sqrt{\lambda_i\lambda_j}(\langle\psi_i^{(1)}|H_1^2|\psi_j^{(1)}\rangle\langle\psi_i^{(2)}|\mathbf{1}|\psi_j^{(2)}\rangle + \langle\psi_i^{(1)}|\mathbf{1}|\psi_j^{(1)}\rangle\langle\psi_i^{(2)}|H_2^2|\psi_j^{(2)}\rangle \\
&\qquad + 2\langle\psi_i^{(1)}|H_1|\psi_j^{(1)}\rangle\langle\psi_i^{(2)}|H_2|\psi_j^{(2)}\rangle) \\
&\quad = \sum_i \lambda_i\langle\psi_i^{(1)}|H_1^2|\psi_i^{(1)}\rangle + \sum_j \lambda_j\langle\psi_j^{(2)}|H_2^2|\psi_j^{(2)}\rangle + 2\sum_{i,j}\sqrt{\lambda_i\lambda_j}a_{ij}b_{ij} \\
&\qquad + \sum_{i,j}\lambda_i a_{ij}a_{ji} + \sum_{i,j}\lambda_j b_{ij}b_{ji} + 2\sum_{i,j}\sqrt{\lambda_i\lambda_j}a_{ij}b_{ij},
\end{aligned}$$

and

$$\begin{aligned}
&\langle\Psi|H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2|\Psi\rangle^2 \\
&\quad = \left(\sum_{i,j}\sqrt{\lambda_i\lambda_j}\langle\psi_i^{(1)}| \otimes \langle\psi_i^{(2)}|H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2|\psi_j^{(1)}\rangle \otimes |\psi_j^{(2)}\rangle\right)^2
\end{aligned}$$

$$= \left(\sum_{i,j} \sqrt{\lambda_i \lambda_j} \langle \psi_i^{(1)} | H_1 | \psi_j^{(1)} \rangle \langle \psi_i^{(2)} | \mathbf{1} | \psi_j^{(2)} \rangle + \langle \psi_i^{(1)} | \mathbf{1} | \psi_j^{(1)} \rangle \langle \psi_i^{(2)} | H_2 | \psi_j^{(2)} \rangle \right)^2$$

$$= \left(\sum_i \lambda_i (a_{ii} + b_{ii})\right)^2.$$

Therefore

$$I_\alpha(\rho, H) = \sum_{i,j} \lambda_i a_{ij} a_{ji} + \sum_{i,j} \lambda_j b_{ij} b_{ji} + 2 \sum_{i,j} \sqrt{\lambda_i \lambda_j} a_{ij} b_{ij} - \left(\sum_i \lambda_i (a_{ii} + b_{ii})\right)^2.$$

On the other hand, by Lemma 1, we have

$$I_\alpha(\rho_1, H_1) = \sum_{i,j} (\lambda_i - \lambda_i^\alpha \lambda_j^{1-\alpha}) a_{ij} a_{ji},$$

$$I_\alpha(\rho_2, H_2) = \sum_{i,j} (\lambda_i - \lambda_i^\alpha \lambda_j^{1-\alpha}) b_{ij} b_{ji}.$$

Thus in order to prove the desired conclusion, it suffices to show that

$$2 \sum_{ij} \sqrt{\lambda_i \lambda_j} a_{ij} b_{ij} - \left(\sum_i \lambda_i (a_{ii} + b_{ii})\right)^2 \geq - \sum_{i,j} \lambda_i^\alpha \lambda_j^{1-\alpha} (a_{ij} a_{ji} + b_{ij} b_{ji}).$$

Writing

$$2 \sum_{ij} \sqrt{\lambda_i \lambda_j} a_{ij} b_{ij} = 2 \sum_{i \neq j} \sqrt{\lambda_i \lambda_j} a_{ij} b_{ij} + 2 \sum_i \lambda_i a_{ii} b_{ii}$$

$$\sum_{i,j} \lambda_i^\alpha \lambda_j^{1-\alpha} (a_{ij} a_{ji} + b_{ij} b_{ji}) = \frac{1}{2} \sum_{i \neq j} (\lambda_i^\alpha \lambda_j^{1-\alpha} + \lambda_i^{1-\alpha} \lambda_j^\alpha)(a_{ij} a_{ji} + b_{ij} b_{ji}) + \sum_i \lambda_i (a_{ii}^2 + b_{ii}^2),$$

the desired inequality follows from

$$\left(\sum_i \lambda_i (a_{ii} + b_{ii})\right)^2 \leq \sum_i \lambda_i (a_{ii} + b_{ii})^2,$$

and

$$\lambda_i^\alpha \lambda_j^{1-\alpha} + \lambda_i^{1-\alpha} \lambda_j^\alpha \geq 2\sqrt{\lambda_i \lambda_j}, \qquad a_{ij} a_{ji} + b_{ij} b_{ji} \geq 2\sqrt{|a_{ij} b_{ij}|^2} = 2|a_{ij} b_{ij}|. \qquad \square$$

In the following result, we assume that $H_1 = (a_{ij})$, $H_2 = (b_{kl})$ and ρ have already been represented as matrices in a fixed representation. That is, the (arbitrary) bases for the two constituent Hilbert spaces have already been chosen, and the joint state ρ is diagonal in the tensor product of the Hilbert spaces with respect to this choice (this implies in particular that the basis vectors in the tensor product are simple tensors).

Proposition 2 *Let $H = H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2$, and ρ be a diagonal density matrix. Then inequality* (5) *holds true, that is*

$$I_\alpha(\rho, H) \geq I_\alpha(\rho_1, H_1) + I_\alpha(\rho_2, H_2).$$

Proof Suppose that H_1 and H_2 act on m-dimensional and n-dimensional spaces, respectively. Let $\rho = \mathrm{diag}\{\Lambda_1, \Lambda_2, \ldots, \Lambda_m\}$, with $\Lambda_i = \mathrm{diag}\{\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{in}\}$, $i = 1, 2, \ldots, m$. Then

$$\rho_1 = \mathrm{diag}\{\mathrm{tr}\,\Lambda_1, \mathrm{tr}\,\Lambda_2, \ldots, \mathrm{tr}\,\Lambda_m\} = \mathrm{diag}\left\{\sum_j \lambda_{1j}, \sum_j \lambda_{2j}, \ldots, \sum_j \lambda_{mj}\right\},$$

and

$$\rho_2 = \mathrm{diag}\left\{\sum_i \lambda_{i1}, \sum_i \lambda_{i2}, \ldots, \sum_i \lambda_{in}\right\}.$$

Since H_1 has the matrix form (a_{ij}), we have

$$\begin{aligned} H &= H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2 \\ &= \begin{pmatrix} a_{11}\mathbf{1} & a_{12}\mathbf{1} & \cdots & a_{1m}\mathbf{1} \\ a_{21}\mathbf{1} & a_{22}\mathbf{1} & \cdots & a_{2m}\mathbf{1} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}\mathbf{1} & a_{m2}\mathbf{1} & \cdots & a_{mm}\mathbf{1} \end{pmatrix} + \begin{pmatrix} H_2 & 0 & \cdots & 0 \\ 0 & H_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & H_2 \end{pmatrix} \\ &= \begin{pmatrix} a_{11}\mathbf{1} + H_2 & a_{12}\mathbf{1} & \cdots & a_{1m}\mathbf{1} \\ a_{21}\mathbf{1} & a_{22}\mathbf{1} + H_2 & \cdots & a_{2m}\mathbf{1} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}\mathbf{1} & a_{m2}\mathbf{1} & \cdots & a_{mm}\mathbf{1} + H_2 \end{pmatrix}. \end{aligned}$$

Therefore

$$\rho^\alpha H = \begin{pmatrix} \Lambda_1^\alpha(a_{11}\mathbf{1} + H_2) & a_{12}\Lambda_1^\alpha & \cdots & a_{1m}\Lambda_1^\alpha \\ a_{21}\Lambda_2^\alpha & \Lambda_2^\alpha(a_{22}\mathbf{1} + H_2) & \cdots & a_{2m}\Lambda_2^\alpha \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}\Lambda_m^\alpha & a_{m2}\Lambda_m^\alpha & \cdots & \Lambda_m^\alpha(a_{mm}\mathbf{1} + H_2) \end{pmatrix},$$

and

$$[\rho^\alpha, H] = \begin{pmatrix} [\Lambda_1^\alpha, H_2] & a_{12}(\Lambda_1^\alpha - \Lambda_2^\alpha) & \cdots & a_{1m}(\Lambda_1^\alpha - \Lambda_m^\alpha) \\ a_{21}(\Lambda_2^\alpha - \Lambda_1^\alpha) & [\Lambda_2^\alpha, H_2] & \cdots & a_{2m}(\Lambda_2^\alpha - \Lambda_m^\alpha) \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}(\Lambda_m^\alpha - \Lambda_1^\alpha) & a_{m2}(\Lambda_m^\alpha - \Lambda_2^\alpha) & \cdots & [\Lambda_m^\alpha, H_2] \end{pmatrix}.$$

Similarly,

$$[\rho^{1-\alpha}, H] = \begin{pmatrix} [\Lambda_1^{1-\alpha}, H_2] & a_{12}(\Lambda_1^{1-\alpha} - \Lambda_2^{1-\alpha}) & \cdots & a_{1m}(\Lambda_1^{1-\alpha} - \Lambda_m^{1-\alpha}) \\ a_{21}(\Lambda_2^{1-\alpha} - \Lambda_1^{1-\alpha}) & [\Lambda_2^{1-\alpha}, H_2] & \cdots & a_{2m}(\Lambda_2^{1-\alpha} - \Lambda_m^{1-\alpha}) \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}(\Lambda_m^{1-\alpha} - \Lambda_1^{1-\alpha}) & a_{m2}(\Lambda_m^{1-\alpha} - \Lambda_2^{1-\alpha}) & \cdots & [\Lambda_m^{1-\alpha}, H_2] \end{pmatrix}.$$

Consequently

$$
\begin{aligned}
I_\alpha(\rho, H) &= -\frac{1}{2}\,\mathrm{tr}[\rho^\alpha, H][\rho^{1-\alpha}, H] \\
&= -\frac{1}{2}\sum_i \mathrm{tr}[\Lambda_i^\alpha, H_2][\Lambda_i^{1-\alpha}, H_2] + \sum_{i<j} a_{ij}a_{ji}\,\mathrm{tr}(\Lambda_i^\alpha - \Lambda_j^\alpha)(\Lambda_i^{1-\alpha} - \Lambda_j^{1-\alpha}) \\
&= \sum_i \sum_{k,l} b_{kl}b_{lk}(\lambda_{ik} - \lambda_{ik}^\alpha \lambda_{il}^{1-\alpha}) + \sum_{i,j} a_{ij}a_{ji}(\mathrm{tr}\,\Lambda_i - \mathrm{tr}\,\Lambda_i^\alpha \Lambda_j^{1-\alpha}).
\end{aligned}
$$

On the other hand, by Lemma 1, we have

$$
I_\alpha(\rho_1, H_1) = \sum_{i,j} a_{ij}a_{ji}(\mathrm{tr}\,\Lambda_i - (\mathrm{tr}\,\Lambda_i)^\alpha (\mathrm{tr}\,\Lambda_j)^{1-\alpha})
$$

$$
I_\alpha(\rho_2, H_2) = \sum_{k,l} b_{kl}b_{lk}\left(\sum_i \lambda_{ik} - \left(\sum_i \lambda_{ik}\right)^\alpha \left(\sum_i \lambda_{il}\right)^{1-\alpha}\right).
$$

Therefore, in order to prove the proposition, it suffices to show that

$$
\sum_i \sum_{k,l} b_{kl}b_{lk}(\lambda_{ik} - \lambda_{ik}^\alpha \lambda_{il}^{1-\alpha}) \geq \sum_{k,l} b_{kl}b_{lk}\left(\sum_i \lambda_{ik} - \left(\sum_i \lambda_{ik}\right)^\alpha \left(\sum_i \lambda_{il}\right)^{1-\alpha}\right)
$$

and

$$
\sum_{i,j} a_{ij}a_{ji}(\mathrm{tr}\,\Lambda_i - \mathrm{tr}\,\Lambda_i^\alpha \Lambda_j^{1-\alpha}) \geq \sum_{i,j} a_{ij}a_{ji}(\mathrm{tr}\,\Lambda_i - (\mathrm{tr}\,\Lambda_i)^\alpha (\mathrm{tr}\,\Lambda_j)^{1-\alpha}).
$$

Both of the above inequalities follow directly from the Hölder inequality. The proof is complete. □

Proposition 3 *Let $H = H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2$. If ρ commutes with either $H_1 \otimes \mathbf{1}$ or $\mathbf{1} \otimes H_2$, then inequality* (5) *holds true, that is,*

$$
I_\alpha(\rho, H) \geq I_\alpha(\rho_1, H_1) + I_\alpha(\rho_2, H_2).
$$

Proof Let $|\psi\rangle, |\phi\rangle$ be any two vectors for the second system, and $\{h_j\}$ be an orthonormal base for the first system, then by the definition of partial trace, we have

$$
\begin{aligned}
\langle \psi | \,\mathrm{tr}_1(\rho(\mathbf{1} \otimes H_2))|\phi\rangle &= \sum_j \langle h_j| \otimes \langle \psi|\rho(\mathbf{1} \otimes H_2)|h_j\rangle \otimes |\phi\rangle \\
&= \sum_j \langle h_j| \otimes \langle \psi|\rho|\mathbf{1}h_j\rangle \otimes |H_2\phi\rangle \\
&= \langle \psi|\,\mathrm{tr}_1\,\rho|H_2\phi\rangle = \langle \psi|\rho_2 H_2|\phi\rangle.
\end{aligned}
$$

Since the above equations hold for any $|\psi\rangle, |\phi\rangle$, we obtain

$$
\rho_2 H_2 = \mathrm{tr}_1(\rho(\mathbf{1} \otimes H_2)).
$$

Similarly,

$$
H_2 \rho_2 = \mathrm{tr}_1((\mathbf{1} \otimes H_2)\rho).
$$

Suppose that ρ commutes with $\mathbf{1}\otimes H_2$, that is, $\rho(\mathbf{1}\otimes H_2)=(\mathbf{1}\otimes H_2)\rho$, then

$$I_\alpha(\rho,H)=I_\alpha(\rho,H_1\otimes\mathbf{1}),$$

and $\rho_2 H_2=H_2\rho_2$ which implies

$$I_\alpha(\rho_2,H_2)=0.$$

Now the desired superadditivity inequality follows form Lemma 2. The case when ρ commutes with $H_1\otimes\mathbf{1}$ can be proved similarly. □

4 Discussions

The Wigner–Yanase–Dyson information can be interpreted as a formal quantum analogue of the classical Fisher information. This is probably the intuitive underlying reason that the Wigner–Yanase–Dyson information has so many desirable properties as required for an information-theoretic measure. In particular, Wigner and Yanase proposed the superadditivity (5) as a postulate for the skew information should satisfy [17], and Lieb emphasized, apart from the convexity, the superadditivity as another "absolute requirement" [10] in order that the Wigner–Yanase–Dyson information be a sensible definition of information.

The physical significance of superadditivity lies in that when a composite system is separated into two parts, the information content should, in general, drop, because the correlations between the two subsystems are lost.

Unfortunately, Hansen's counterexample illustrated that superadditivity cannot be true for the skew information in the most general form [5], and at present we have no idea about the deep reason for this unexpected and quite counterintuitive phenomenon. Since superadditivity holds for the classical Fisher information, and fails for the quantum case, further investigation of superadditivity in quantum case may reveal some intrinsic mechanism underlying the difference between the classical and the quantum.

Despite the negative result of Hansen, we still have superadditivity in several important cases. Proposition 2 may be of particular interest since any density operator can be diagonalized. Thus in matrix representation of operators, let $U\rho U^\dagger=D$, where U is unitary and D is diagonal, then

$$I_\alpha(\rho,H_1\otimes\mathbf{1}+\mathbf{1}\otimes H_2)=I_\alpha(D,U(H_1\otimes\mathbf{1}+\mathbf{1}\otimes H_2)U^\dagger).$$

If it happens that $U(H_1\otimes\mathbf{1}+\mathbf{1}\otimes H_2)U^\dagger$ can be written as (of course, this may not be possible)

$$U(H_1\otimes\mathbf{1}+\mathbf{1}\otimes H_2)U^\dagger=H_1'\otimes\mathbf{1}+\mathbf{1}\otimes H_2',$$

then by Proposition 2, we have

$$I_\alpha(\rho,H_1\otimes\mathbf{1}+\mathbf{1}\otimes H_2)=I_\alpha(D,H_1'\otimes\mathbf{1}+\mathbf{1}\otimes H_2')\ge I_\alpha(D_1,H_1')+I(D_2,H_2'),$$

which may be interpreted as a modified version of superadditivity. Here $D_1=\mathrm{tr}_2\,D$, $D_2=\mathrm{tr}_1\,D$, while H_1' and H_2' are observables for the first and second system, respectively. We emphasize that Proposition 2 does not imply that superadditivity always holds true, because when we diagonalize the density operator ρ, the simple tensor additivity structure $H_1\otimes\mathbf{1}+\mathbf{1}\otimes H_2$ may be lost in the new representation.

It is also interesting to note that if we define an information measure

$$J(\sigma) := \operatorname{tr} \sigma \log \sigma$$

as the negative of the von Neumann (quantum entropy) for any density operator σ, then from subadditivity of the von Neumann entropy [16], this information measure has the following superadditivity property:

$$J(\rho) \geq J(\rho_1) + J(\rho_2).$$

Here ρ is a density operator for a composite quantum system, ρ_1 and ρ_2 are the reduced density operator (partial trace) for the first and the second system, respectively. The above inequality is in the same spirit as inequality (3) and the superadditivity requirement (5).

Acknowledgements The author is grateful to the referees for pointing out Hansen's counterexample, and for constructive suggestions. This work is supported by NSFC, Grant. No. 10571166.

References

1. Carlen, E.A.: Superadditivity of Fisher's information and logarithmic Sobolev inequalities. J. Funct. Anal. **101**, 194–211 (1991)
2. Cramér, H.: Mathematical Methods of Statistics. Princeton University Press, Princeton (1974)
3. Fisher, R.A.: Theory of statistical estimation. Proc. Camb. Philos. Soc. **22**, 700–725 (1925)
4. Frieden, B.R.: Science from Fisher Information, a Unification. Cambridge University Press, Cambridge (2004)
5. Hansen, F.: The Wigner–Yanase entropy is not subadditive. J. Stat. Phys. **126**, 643–648 (2007)
6. Huber, P.J.: Robust Statistics. Wiley, New York (1981)
7. Johnson, O.: Information Theory and the Central Limit Theorems. Imperial College Press, London (2004)
8. Kagan, A., Landsman, Z.: Statistical meaning of Carlen's superadditivity of the Fisher information. Stat. Probab. Lett. **32**, 175–179 (1997)
9. Kosaki, H.: Interpolation theory and the Wigner–Yanase–Dyson–Lieb concavity. Commun. Math. Phys. **87**, 315–329 (1982)
10. Lieb, E.H.: Convex trace functions and the Wigner–Yanase–Dyson conjecture. Adv. Math. **11**, 267–288 (1973)
11. Luo, S.L.: Wigner–Yanase skew information and uncertainty relations. Phys. Rev. Lett. **91**, 180403 (2003)
12. Luo, S.L.: Wigner–Yanase skew information vs. quantum Fisher information. Proc. Am. Math. Soc. **132**, 885–890 (2004)
13. Luo, S.L.: Quantum versus classical uncertainties. Theor. Math. Phys. **143**, 681–688 (2005)
14. Sakurai, J.J.: Modern Quantum Mechanics. Addison–Wesley, Massachusetts (1994)
15. Uhlmann, A.: Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory. Commun. Math. Phys. **54**, 21–32 (1977)
16. Wehrl, A.: General properties of entropy. Rev. Mod. Phys. **50**, 221–260 (1978)
17. Wigner, E.P., Yanase, M.M.: Information contents of distributions. Proc. Natl. Acad. Sci. USA **49**, 910–918 (1963)