# ON THE STATE ALIGNMENT OF THE QUANTUM JENSEN SHANNON DIVERGENCE

ANDREA TORSELLO

Abstract. In this technical report we will prove some results of the behavior of the Quantum Jensen-Shannon divergence subject to an optimal state alignment of the density matrices. Namely, we will prove that the unitary transformation that minimizes the divergence aligns the eigenvectors of the two density matrices according to the magnitude of the corresponding eigenvalues

## 1. Preliminaries

In quantum mechanics a state is described by a unit complex vector $|\psi\rangle \in \mathcal{H}$ where $\mathcal{H}$ is a complex Hilbert space and $\langle \psi \,|\, \psi \rangle = 1$. Here, in accordance with Dirac notation, a ket $|\psi\rangle$ represent a vector, the corresponding bra $\langle\psi|$ represents the conjugate transpose, and as a consequence, $\langle \phi \,|\, \psi \rangle$ is the dot product between $\phi$ and $\psi$, and $|\phi\rangle \langle\psi|$ is a rank 1 operator.

The evolution of quantum states is governed by Schrödinger equation

$$i\hbar \frac{d}{dt} |\psi_t\rangle = H |\psi_t\rangle \tag{1}$$

where $H$ is the Hamiltonian of the system. As a consequence all state transformations are governed by unitary operators.

While a pure state can be naturally described using a single ket vector, in general a quantum system can be in a *mixed state*, i.e., a statistical ensemble of pure quantum states $|\psi_i\rangle$, each with probability $p_i$. The *density operator* (or *density matrix*) of such a system is defined as

$$\rho = \sum_j p_j |\psi_j\rangle \langle\psi_j| \,. \tag{2}$$

Density operators are positive-definite unit-trace Hermitian operators directly linked with the observables of the (mixed) quantum system. Let $O$ be an observable, i.e., an Hermitian operator acting on the quantum states and providing a measurement. Without loss of generality we have $O = \sum_j v_j P_j$, where $P_j$ is the orthogonal projector onto the $j$-th observation basis, and $v_j$ is the measurement value when the quantum state is observed to be in this observation basis. The expected value of the measurement $O$ over a mixed state can be calculated from the density matrix $\rho$: $\langle O \rangle = \mathrm{Tr}\,(\rho O)$, where Tr is the trace operator.

The Von Neumann entropy of a density operator $\rho$ is

$$H_N(\rho) = -Tr(\rho \log \rho) = -\sum_j \lambda_j \log \lambda_j \,, \tag{3}$$

where the $\lambda_j$s are the eigenvalues of $\rho$.

The quantum Jensen-Shannon divergence (QJSD) is a generalization of the classical Jensen Shannon divergence used as a measure of distinguishability between quantum states. Given two density operators $\rho$ and $\sigma$ is defined as

$$D_{\mathrm{JS}}(\rho, \sigma) = H_N\Big(\frac{\rho+\sigma}{2}\Big) - \frac{1}{2}\big(H_N(\rho) + H_N(\sigma)\big) \tag{4}$$

This quantity is symmetric, bounded between 0 and 1, and negative definite for pure states and is conjectured, with ample experimental evidence, to be negative definite for all states [2].

## 2. State-Aligned Quantum Jensen-Shannon Divergence

We define the *state-aligned* Quantum Jensen-Shannon Divergence as

$$D_{\mathrm{SAJS}}(\rho, \sigma) = \min_{U \in \mathcal{U}} D_{\mathrm{JS}}(\rho, U\sigma U^\dagger) = \min_{U \in \mathcal{U}} H_N\Big(\frac{\rho + U\sigma U^\dagger}{2}\Big) - \frac{1}{2}\big(H_N(\rho) + H_N(\sigma)\big) \tag{5}$$

where $\mathcal{U}$ is the unitary group over $\mathcal{H}$.

Let

$$\bar{H}_N(U) = H_N\left(\frac{\rho + U\sigma U^\dagger}{2}\right) = \mathrm{Tr}\left(-\frac{\rho + U\sigma U^\dagger}{2} \log\left(\frac{\rho + U\sigma U^\dagger}{2}\right)\right) \tag{6}$$

be the entropy of the aligned density matrix as a function of the state-alignment $O \in \mathcal{U}$. We can prove the following

**Lemma 1.**

$$\frac{\partial}{\partial U}_{U \in \mathcal{U}} \bar{H}_N(U) = \left[\frac{U\sigma U^\dagger}{2},\, \log\left(e\frac{\rho + U\sigma U^\dagger}{2}\right)^T\right]$$

*where $[\cdot,\, \cdot]$ is the commutator operator defined as $[A,\, B] = AB - BA$.*

*Proof* First note that

$$\frac{\partial}{\partial U}_{U \in \mathcal{U}} \bar{H}_N(U) = \frac{\partial}{\partial S}_{S \in \mathcal{U}} \bar{H}_N(SU)\bigg|_{S=I} \,. \tag{7}$$

Further, the unitary group $\mathcal{U}$ forms a a manifold in the space $\mathrm{End}(\mathcal{H})$ of endomorphisms of $\mathcal{H}$, so the derivative over $\mathcal{U}$ is the projection on the tangent space of $\mathcal{U}$ of the derivative over the whole of $\mathrm{End}(\mathcal{H})$:

$$\frac{\partial}{\partial S}_{S \in \mathcal{U}} f(S) = \mathcal{P}\left(\frac{\partial}{\partial S} f(S)\right) \,, \tag{8}$$

where $\mathcal{P}$ is the orthogonal projection on the tangent space of $\mathcal{U}$. For $S = I$ the tangent space is the space of anti-Hermitian operators, i.e., operators of the form $iA$ where $A = A^\dagger$ is Hermitian. Thus

$$\mathcal{P}(A) = \frac{A - A^\dagger}{2}\,. \tag{9}$$

Let $X = \frac{\rho + SU\sigma U^T S^T}{2}$, we have

$$\frac{\partial}{\partial S}\operatorname{Tr}\big(-X\log(X)\big)\Big|_{S=I} = \frac{\partial}{\partial X}\operatorname{Tr}\big(-X\log(X)\big)\frac{\partial X}{\partial S}\Big|_{S=I}\,. \tag{10}$$

For the first term of (11) we have

$$\frac{\partial}{\partial x_{ij}}\operatorname{Tr}\big(X\log(X)\big) = \operatorname{Tr}\left(\frac{\partial X}{\partial x_{ij}}\log(X)\right) + \operatorname{Tr}\left(X\frac{\partial}{\partial x_{ij}}\log(X)\right) \tag{11}$$

with

$$\operatorname{Tr}\left(\frac{\partial X}{\partial x_{ij}}\log(X)\right) = \sum_k\sum_l \frac{\partial x_{kl}}{\partial x_{ij}}\big(\log(X)\big)_{lk} = \big(\log(X)\big)_{ji}\,. \tag{12}$$

For the second term of (11) we use the following property of the exponential of a general time-dependent matrix $M(t)$ proved in [1]:

$$\frac{d}{dt}\exp(M(t)) = \int_0^1 \exp(\alpha M(t))\frac{dM(t)}{dt}\exp((1-\alpha)M(t))\,d\alpha\,. \tag{13}$$

Recall that $\frac{\partial}{\partial x_{ij}} f(X) = \frac{d}{dt} f(X + tE_{ij})$, where $E_{ij}$ is the matrix of all zeroes except the entry $(i,j)$ which is equal to one. Setting $M(t) = \log(X + tE_{ij})$ and evaluating for $t = 0$, we have:

$$\begin{aligned} E_{ij} &= \frac{d}{dt}(X + tE_{ij})\Big|_{t=0} \\ &= \frac{d}{dt}\exp\big(\log(X + tE_{ij})\big)\Big|_{t=0} \\ &= \int_0^1 X^\alpha\left(\frac{d}{dt}\log(X + tE_{ij})\Big|_{t=0}\right)X^{1-\alpha}\,d\alpha \\ &= \int_0^1 X^\alpha\frac{\partial}{\partial x_{ij}}\log(X)\,X^{1-\alpha}\,d\alpha\,. \end{aligned} \tag{14}$$

From which we get

$$
\begin{aligned}
\operatorname{Tr}\left(E_{ij}\right) &= \operatorname{Tr}\left(\int_0^1 X^\alpha \frac{\partial}{\partial x_{ij}} \log(X)\, X^{1-\alpha}\, d\alpha\right) \\
&= \int_0^1 \operatorname{Tr}\left(X^\alpha \frac{\partial}{\partial x_{ij}} \log(X) X^{1-\alpha}\right) d\alpha \\
&= \int_0^1 \operatorname{Tr}\left(X \frac{\partial}{\partial x_{ij}} \log(X)\right) d\alpha \\
&= \operatorname{Tr}\left(X \frac{\partial}{\partial x_{ij}} \log(X)\right).
\end{aligned} \tag{15}
$$

Putting the last results together, we obtain

$$
\frac{\partial}{\partial X} \operatorname{Tr}\big(-X\log(X)\big) = -\log(X)^T - I = -\log(eX^T)\,. \tag{16}
$$

On the other hand, we have

$$
\frac{\partial}{\partial S} \frac{A + SRBR^T S^T}{2}\bigg|_{S=I} = RBR^T\,. \tag{17}
$$

Finally, we have

$$
\begin{aligned}
&\frac{\partial}{\partial S}_{S\in\mathcal{U}} \operatorname{Tr}\left(-\frac{\rho + SU\sigma U^\dagger S^\dagger}{2} \log\left(\frac{\rho + SU\sigma U^\dagger S^\dagger}{2}\right)\right)\bigg|_{S=I} = \\
&\mathcal{P}\left(\frac{\partial}{\partial S} \operatorname{Tr}\left(-\frac{\rho + SU\sigma U^\dagger S^\dagger}{2} \log\left(\frac{\rho + SU\sigma U^\dagger S^\dagger}{2}\right)\right)\bigg|_{S=I}\right) = \\
&\mathcal{P}\left(-\log\left(e\frac{\rho + U\sigma U^\dagger}{2}\right)^T \left(U\sigma U^\dagger\right)\right) = \\
&\frac{-\log\left(e\frac{\rho+U\sigma U^\dagger}{2}\right)^T \left(U\sigma U^\dagger\right) + \left(U\sigma U^\dagger\right)^\dagger \log\left(e\frac{\rho+U\sigma U^\dagger}{2}\right)^{T\dagger}}{2} = \\
&\frac{U\sigma U^\dagger}{2} \log\left(e\frac{\rho + U\sigma U^\dagger}{2}\right)^T - \log\left(e\frac{\rho + U\sigma U^\dagger}{2}\right)^T \frac{U\sigma U^\dagger}{2} = \\
&\left[\frac{U\sigma U^\dagger}{2}, \log\left(e\frac{\rho + U\sigma U^\dagger}{2}\right)^T\right]
\end{aligned} \tag{18}
$$

where the second to last equality is due to the fact that $\rho$ and $\sigma$ are Hermitian. QED.

**Theorem 1.** *The extremants of $\bar{H}_N(U)$ are for unitary operators $U$ that align the eigenfunctions of $U\sigma U^\dagger$ with those of $\rho$.*

*Proof* for the previous lemma, the differential of $\bar{H}_N(U)$ with is null if and only if

$$\left[\frac{U\sigma U^\dagger}{2}, \log\left(e\frac{\rho + U\sigma U^\dagger}{2}\right)^T\right] = 0 \tag{19}$$

In fact

$$\frac{\partial}{\partial U}\bar{H}_N(U)\bigg|_{U=U_0} = \frac{\partial}{\partial S}\bar{H}_N(SU_0)\bigg|_{S=I} R_0^\dagger\,. \tag{20}$$

Since $R_0$ is unitary

$$\frac{\partial}{\partial U}\bar{H}_N(U)\bigg|_{U=U_0} = 0 \text{ if and only if } \frac{\partial}{\partial S}\bar{H}_N(SU_0)\bigg|_{S=I} = 0\,. \tag{21}$$

The commutator is zero if and only if the two matrices are co-diagonalizable. Hence, at points with zero gradient There must be a (countable) set of functions that are form a complete eigenbasis both for $U\sigma U^\dagger$ and $\log\left(e\frac{\rho+U\sigma U^\dagger}{2}\right)^T$, or, equivalently, of $\rho + U\sigma U^\dagger$. This happens if and only if the eigenfunctions of $U\sigma U^\dagger$ and those of $\rho$ are aligned. QED.

In the following we assume that the quantum system has a finite number of states, i.e., $\mathcal{H} = \mathbb{C}^n$.

**Theorem 2.** *Let $\rho = \Phi_\rho\Lambda_\rho\Phi_\rho^\dagger$ and $\sigma = \Phi_\sigma\Lambda_\sigma\Phi_\sigma^\dagger$ be the singular value decompositions of $\rho$ and $\sigma$ respectively, with the eigenvalues in ascending order in both $\Lambda_\rho$ and $\Lambda_\sigma$, i.e., $i > j \Rightarrow (\Lambda_\rho)_{ii} \geq (\Lambda_\rho)_{jj}$ and $(\Lambda_\sigma)_{ii} \geq (\Lambda_\sigma)_{jj}$. The global minimum of $\bar{H}_N(U)$ is attained by $U = \Phi_\rho\Phi_\sigma^\dagger$.*

*Proof* Since the gradient is zero only for rotations that align the eigenvalues, we can consider only those. There the eigenvalues of $\frac{1}{2}(\rho + U\sigma U^\dagger)$ assume the form $\frac{1}{2}(\lambda_j + \mu_{\sigma(j)})$ where $\lambda_j$ and $\mu_j$ for $j = 1\cdots n$ are the eigenvalues of $\rho$ and $\sigma$ respectively taken with their multiplicity in ascending order, and $\sigma : \Sigma_n$ is a permutation. Thus the problem of minimizing the Von Newman Entropy can be cast into the equivalent bipartite matching problem

$$\begin{aligned} \min_{\sigma} \quad & \sum_j -\frac{\lambda_i + \mu_{\sigma(i)}}{2}\log\left(\frac{\lambda_i + \mu_{\sigma(i)}}{2}\right) \\ \text{s.t.} \quad & \sigma \in \Sigma_n \end{aligned} \tag{22}$$

which in matrix notation becomes

$$\begin{aligned} \min_{P} \quad & \mathrm{Tr}\left(MP\right) \\ \text{s.t.} \quad & P \in \Sigma_n \end{aligned} \tag{23}$$

where $M = (m_{ij})$, $m_{ij} = -\frac{1}{2}(\lambda_i + \mu_j)\log\left(\frac{1}{2}(\lambda_i + \mu_j)\right)$. To prove the theorem we need to prove that the minimum is attained for $P = I$. To do this we prove that the identity matrix is the global minimum for the relaxed linear problem

$$
\begin{aligned}
\min_{P} \quad & \operatorname{Tr}\left(MP\right) \\
\text{s.t.} \quad & P1 = 1 \\
& P^T 1 = 1 \\
& P \geq 0\,.
\end{aligned} \tag{24}
$$

To prove the result we need only to show that all the vertices of the feasible polytope that are edge-adjacent to the identity matrix yield a higher value for the objective function. The adjacent vertices are of the form $I + SW_{ij}$ with

$$
SW_{ij} = E_{ij} + E_{ji} - E_{ii} - E_{jj}\,. \tag{25}
$$

However, we have

$$
\begin{aligned}
\operatorname{Tr}\left(M\, SW_{ij}\right) &= +M_{ji} + M_{ij} - M_{ii} - M_{jj} \\
&= -\frac{\lambda_j + \mu_i}{2} \log\left(\frac{\lambda_j + \mu_i}{2}\right) - \frac{\lambda_i + \mu_j}{2} \log\left(\frac{\lambda_i + \mu_j}{2}\right) + \\
&\quad + \frac{\lambda_i + \mu_i}{2} \log\left(\frac{\lambda_i + \mu_i}{2}\right) + \frac{\lambda_j + \mu_j}{2} \log\left(\frac{\lambda_j + \mu_j}{2}\right) \\
&= \frac{1}{2}\left[\lambda_i \log\left(\frac{\lambda_i + \mu_i}{\lambda_i + \mu_j}\right) - \lambda_j \log\left(\frac{\lambda_j + \mu_i}{\lambda_j + \mu_j}\right)\right] \\
&\quad + \frac{1}{2}\left[\mu_i \log\left(\frac{\lambda_i + \mu_i}{\lambda_j + \mu_i}\right) - \mu_j \log\left(\frac{\lambda_i + \mu_j}{\lambda_j + \mu_j}\right)\right] \geq 0
\end{aligned} \tag{26}
$$

where the inequality holds because, if $i > j$ then $\lambda_i \geq \lambda_j$, $\mu_i \geq \mu_j$, all the logarithms are positive, and the ones multiplied by $\lambda_i$ and $\mu_i$ are greater than the corresponding ones multiplied by $\lambda_j$ and $\mu_j$. On the other hand, if $i < j$ then $\lambda_i \leq \lambda_j$, $\mu_i \leq \mu_j$, all the logarithms are negative, and the ones multiplied by $\lambda_i$ and $\mu_i$ are in absolute value smaller than the corresponding ones multiplied by $\lambda_j$ and $\mu_j$. QED.

**Theorem 3.** *The orthogonal transformations minimizing the quantum Jensen Shannon divergence (QJSD) between pairs of density matrices in a set are transitive, i.e. let*

$$
\begin{aligned}
U_{AB} &= \operatorname*{argmin}_{U \in \mathcal{U}} \operatorname{QJSD}(A, RBR^T) \\
U_{BC} &= \operatorname*{argmin}_{U \in \mathcal{U}} \operatorname{QJSD}(B, RCR^T) \\
U_{AC} &= \operatorname*{argmin}_{U \in \mathcal{U}} \operatorname{QJSD}(A, RCR^T)
\end{aligned}
$$

*then*

$$
\operatorname{QJSD}\left(A, U_{AB} U_{BC} C U_{BC}^\dagger U_{AB}^\dagger\right) = \operatorname{QJSD}\left(A, U_{AC} C U_{AC}^\dagger\right)
$$

*Proof* The optimal transformation between two density matrices is completely determined by the relation $R^*_{AB} = \Phi_A \Phi_B^T$ up to a change of sign of the eigenvalue and a change of base for each eigenspace associated with an eigenvalue with multiplicity greater than one. In any case these changes do not affect the value of the divergence. but,

$$R^*_{AB}R^*_{BC} = \Phi_A\Phi_B^\dagger\Phi_B\Phi_C^\dagger = \Phi_A\Phi_C^\dagger = R^*_{AC} \tag{27}$$

QED.

**Theorem 4.** *The state-aligned QJSD kernel is negative definite.*

*Proof* Thanks to the previous theorems the value of the quantum Jensen Shannon divergence of the optimally aligned density matrices is equal to the normal Jensen Shannon divergence of sorted eigenvalues of the density matrices taken as probability distributions, and the Jensen Shannon divergence is negative definite [2]. QED.

## References


[1] R. M. Wilcox, Exponential Operators and Parameter Differentiation in Quantum Physics, *Journal of Mathematical Physics* **8**(4):962-982, 1967. doi:10.1063/1.1705306

[2] Briët, J., Harremoës, P., Properties of classical and quantum jensen-shannon divergence, *Physical review A* **79**, 052311, 2009.