Quantum correlations and distinguishability of quantum states

Dominique Spehner

Citation: Journal of Mathematical Physics **55**, 075211 (2014); doi: 10.1063/1.4885832
View online: http://dx.doi.org/10.1063/1.4885832
View Table of Contents: http://scitation.aip.org/content/aip/journal/jmp/55/7?ver=pdfcov
Published by the AIP Publishing

JOURNAL OF MATHEMATICAL PHYSICS **55**, 075211 (2014)

Quantum correlations and distinguishability of quantum states

Dominique Spehner
Université Grenoble Alpes and CNRS, Institut Fourier, F-38000 Grenoble, France and Laboratoire de Physique et Modélisation des Milieux Condensés, F-38000 Grenoble, France

(Received 14 June 2014; accepted 18 June 2014; published online 21 July 2014)

A survey of various concepts in quantum information is given, with a main emphasis on the distinguishability of quantum states and quantum correlations. Covered topics include generalized and least square measurements, state discrimination, quantum relative entropies, the Bures distance on the set of quantum states, the quantum Fisher information, the quantum Chernoff bound, bipartite entanglement, the quantum discord, and geometrical measures of quantum correlations. The article is intended both for physicists interested not only by collections of results but also by the mathematical methods justifying them, and for mathematicians looking for an up-to-date introductory course on these subjects, which are mainly developed in the physics literature. © *2014 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4885832]

I. INTRODUCTION

The fundamental role played by the theory of information in physics has been demonstrated in the last century along with the development of statistical physics.[14] More recently, it has been recognized that information is also at the heart of quantum physics, leading to the emergence of a new field called quantum information. In few words, quantum information theory is concerned with the use of quantum systems to accomplish information-processing tasks which are either not feasible classically or are done classically much less efficiently.[117] These tasks can be related to a computational problem or to communication, for instance, sending encrypted information in a secure way. Computational tasks are performed on a quantum computer made of qubits. Such qubits are two-level quantum systems in arbitrary superpositions of $|0\rangle$ and $|1\rangle$ instead of being either in state 0 or 1 as with classical bits. A quantum algorithm is a unitary quantum evolution on a set of qubits followed by a measurement, the outcomes of which should provide the solution of the problem. For example, the celebrated Shor algorithm factorizes an integer with N digits into prime numbers in a time $\mathcal{O}(N^2 \ln N \ln(\ln N))$,[145] instead of the exponential time required by all known classical algorithms. Quantum computers with a few qubits have been implemented in physics laboratories. There is still a lot of debate about whether we will be able in the future to manipulate coherently many qubits and address them locally during a sufficiently long computational time, and which quantum systems are the most promising.[117,28]

The fact that quantum algorithms and communication protocols can outperform their classical analogs is usually attributed to quantum correlations. Such correlations in composite quantum systems are at the origin of the violation of the Bell inequalities, which has been confirmed experimentally.[127] These quantum correlations are quite different in nature from classical correlations in stochastic processes. For a long time they have been identified with entanglement. However, there is now increasing evidence that other types of quantum correlations in mixed states, which may be present even in unentangled states and are captured notably by the quantum discord,[120,75] might be of relevance in certain quantum algorithms and communication protocols.[49,97,123,107,34,66,47]

In this survey article, we review the basic properties of the entanglement measures and quantum discord and present a geometrical description of these notions based on the Bures distance on the set of quantum states. In this approach, the quantum discord turns out to be related to the problem of discriminating non-orthogonal quantum states. Two central questions guide the discussion in this

article and can be formulated as follows. How well can one distinguish unknown quantum states pertaining to a given ensemble by performing a measurement on a system? If this system consists of several particles, does the amount of information one gets from measurements on a single particle tell us something about the way the particles are correlated? Quantum measurements and entropies obviously come into the game in these two questions. They constitute the subjects of Secs. III, IV, and VI. Some answers to the first question are given in Secs. V and VIII, devoted, respectively, to state discrimination and to related topics called hypothesis testing and parameter estimation. The Bures distance and Uhlmann fidelity are introduced in Sec. VII. A detailed account of their properties is given there. The remaining sections (Secs. IX, X, and XI) address the problem of quantifying quantum correlations and provide answers to the second question. It is neither our purpose to discuss thoroughly the (huge amount of) quantum correlation measures found in the literature nor to study how these correlations could explain the quantum efficiencies. Well-documented surveys on quantum entanglement already exist, see, e.g., Refs. 82 and 67, as well as on the quantum discord and related measures.[110] The precise role of entanglement as a resource in quantum computing and quantum communication is still not completely understood, in spite of recent progresses (such as the proof that, in order to offer an exponential speedup over classical algorithms, a quantum algorithm using pure states must produce entanglement which is not restricted to blocks of qubits of fixed size as the system size increases[92]). The role played by the discord as a quantum resource is, in turn, still poorly understood and constitutes a challenging issue (see Ref. 110).

We concentrate in our exposition on the mathematical and fundamental aspects of the theory. In particular, we will not investigate here the physical implementations and the system-dependent irreversible dynamical processes destroying (or sometimes producing) quantum correlations. We present the detailed proofs of some selected fundamental results, instead of relating all important achievements obtained so far. Most of these results have been published in physics journals, and are sometimes explained in the original papers without full mathematical rigor in their derivation. Others have been published in mathematical journals with full proofs, which are nevertheless given here for completeness. We try to emphasize how the results are connected between themselves and to stress the similarities in the arguments used to derive them. This sometimes leads to new proofs.

Quantum information is a rapidly growing field of research and the amount of articles and surveys devoted to it is already considerable. Researchers who got interested by this subject recently (including the author) may fear to have difficulties to form a clear opinion about the most pertinent open questions. Significant contributions have been made by physicists, mathematicians, and computer scientists, who constitute a broad community with different viewpoints. We hope that this article may be useful to mathematicians, by providing examples of interesting problems and explaining the mathematical tools used to tackle them. It may also be of help to physicists wishing to get acquainted with such tools, which could be useful to derive new results. The paper is written as an introductory course. Certain statements appear as remarks, which play the role of exercises, with the main arguments to justify them. We encourage the reader to complete these proofs by himself. This work is intended to be complementary to other surveys containing collections of results without explicit derivations and to more introductory monographs like Ref. 117, which do not include the most recent advances.

The following comments on the structure of the article might be helpful. The contents of Secs. V, VII, and IX–X are largely independent. On the other hand, Sec. V is partly related to Sec. IV, and Sec. VIII makes use of the results of Secs. V and VII E. The material of Secs. VI A and VI B is relevant for Secs. IX and X. Section XI needs more or less the knowledge of all previous sections. The main definitions and theorems presented in Secs. II and III are used in the whole article. Two appendices (Appendices A and B) contain textbook issues about operator convex functions and some less standard trace inequalities.

Before closing this Introduction, let us warn the reader that we will be exclusively concerned by quantum systems with *finite-dimensional Hilbert spaces*. This is motivated for two reasons. Firstly, this is the case of most systems in quantum information theory. Second, in this way one avoids the technical complications of infinite-dimensional spaces and concentrates oneself on the main ideas and concepts. Some of these concepts have been originally worked out in the general setting of C^*-algebras, but we shall present here simpler proofs applying to the finite-dimensional case only.

II. QUANTUM STATES

In this section, we review the basic definitions of pure and mixed states, entangled states, and the pure state decompositions and purifications of mixed states. Before that, we introduce in Sec. II A some notation and define a few mathematical objects from the theory of operator algebras, which will be used repeatedly in this article. In Sec. II B we discuss an extremely useful result from linear algebra, namely, the Schmidt decomposition.

In all what follows, capital letters A, B, etc., refer to quantum systems, $\mathcal{H}_\mathsf{A}$, $\mathcal{H}_\mathsf{B}$, etc., denote their Hilbert spaces, and $n_\mathsf{A} = \dim \mathcal{H}_\mathsf{A}$, $n_\mathsf{B} = \dim \mathcal{H}_\mathsf{B}$, etc., the dimensions of these spaces. These dimensions are always assumed to be finite. A bipartite system AB formed by putting together the systems A and B has Hilbert space given by the tensor product $\mathcal{H}_\mathsf{AB} = \mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$. For instance, if A and B are two qubits with Hilbert spaces $\mathcal{H}_\mathsf{A} \simeq \mathcal{H}_\mathsf{B} \simeq \mathbb{C}^2$, the space of these two qubits is $\mathcal{H}_\mathsf{AB} = \mathbb{C}^2 \otimes \mathbb{C}^2 \simeq \mathbb{C}^4$. Similarly, $\mathcal{H}_{\mathsf{A}_1 \ldots \mathsf{A}_k} = \mathcal{H}_{\mathsf{A}_1} \otimes \cdots \otimes \mathcal{H}_{\mathsf{A}_k}$ is the Hilbert space of the multipartite system formed by putting together the systems $\mathsf{A}_1, \ldots, \mathsf{A}_k$. The tensor product vectors $|\psi_\mathsf{A}\rangle \otimes |\phi_\mathsf{B}\rangle \in \mathcal{H}_\mathsf{AB}$ will be denoted either by $|\psi_\mathsf{A} \otimes \phi_\mathsf{B}\rangle$ or, more often, by $|\psi_\mathsf{A}\rangle|\phi_\mathsf{B}\rangle$ (as common in the physics literature we do not write the tensor product symbol $\otimes$ explicitly.).

A. Quantum states and observables

A *state* of a quantum system with Hilbert space $\mathcal{H}$ is given by a density matrix ρ, that is, a non-negative operator on $\mathcal{H}$ with unit trace $\mathrm{tr}\,\rho = 1$. We write $\mathcal{E}(\mathcal{H})$ the convex cone formed by all states on $\mathcal{H}$. States will always be denoted by the letters ρ, σ, or τ, with subscripts referring to the corresponding system if necessary. The extreme points of the cone $\mathcal{E}(\mathcal{H})$ are the *pure states* $\rho_\psi = |\psi\rangle\langle\psi|$, with $|\psi\rangle \in \mathcal{H}$, $\|\psi\| = 1$ (here $|\psi\rangle\langle\psi|$ designates the rank-one orthogonal projector onto $\mathbb{C}|\psi\rangle$). The pure states can be identified with elements of the projective space $P\mathcal{H}$, that is, the set of equivalence classes of normalized vectors in $\mathcal{H}$ modulo a phase factor. The vectors $e^{i\theta}|\psi\rangle \in \mathcal{H}$ with $0 \le \theta < 2\pi$ are called the representatives of $\rho_\psi \in P\mathcal{H}$. We will abusively write $|\psi\rangle$ instead of ρ_ψ, except when this may be a source of confusion. If ρ is a state of a bipartite system AB with Hilbert space $\mathcal{H}_\mathsf{AB} = \mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$, the *reduced states* of A and B are defined by partial tracing ρ over the other subsystem. They are denoted by $\rho_A = \mathrm{tr}_B(\rho) \in \mathcal{E}(\mathcal{H}_\mathsf{A})$ and $\rho_B = \mathrm{tr}_A(\rho) \in \mathcal{E}(\mathcal{H}_\mathsf{B})$. These reduced states correspond to the marginals of a joint probability in classical probability theory.

The C^*-algebra of bounded linear operators from $\mathcal{H}$ to $\mathcal{H}'$ is denoted by $\mathcal{B}(\mathcal{H}, \mathcal{H}')$, and we write $\mathcal{B}(\mathcal{H}) = \mathcal{B}(\mathcal{H}, \mathcal{H})$. In our finite-dimensional setting, $\mathcal{B}(\mathcal{H}, \mathcal{H}')$ is the algebra of all $n' \times n$ finite complex matrices, with $\dim \mathcal{H} = n$ and $\dim \mathcal{H}' = n'$. The Hilbert-Schmidt scalar product on $\mathcal{B}(\mathcal{H}, \mathcal{H}')$ is defined by

$$\langle X,\ Y\rangle = \mathrm{tr}(X^*Y), \tag{1}$$

where X^* denotes the adjoint operator of X. The associated norm is $\|X\|_2 = [\mathrm{tr}(X^*X)]^{\frac{1}{2}}$. The set of states $\mathcal{E}(\mathcal{H})$ can be endowed with the distances

$$d_p(\rho, \sigma) = \|\rho - \sigma\|_p = \left[\mathrm{tr}(|\rho - \sigma|^p)\right]^{\frac{1}{p}} \tag{2}$$

with $p \ge 1$. Here $|X|$ denotes the non-negative operator $|X| = \sqrt{X^*X}$. We shall see in Sec. VII that there are other more natural distances on $\mathcal{E}(\mathcal{H})$ from a quantum information point of view. When $p \to \infty$, $\|X\|_p$ converges to the operator norm $\|X\|_\infty = \|X\|$ of X, that is, the maximal eigenvalue of $|X|$. The Hölder inequality reads

$$\|X\|_p = \max_{Y, \|Y\|_q = 1} |\mathrm{tr}(XY)| \tag{3}$$

with $p > 1$ and $q = p/(p-1)$. This still holds for $p = 1$ and $q = \infty$, as can be shown by using the Cauchy-Schwarz inequality for the scalar product (1). In that case the maximum is achieved if and only if $YU|X|^{\frac{1}{2}} = e^{i\theta}|X|^{\frac{1}{2}}$ with $\theta \in [0, 2\pi)$ and U a unitary such that $X = U|X|$ (polar decomposition).

A self-adjoint operator $O \in \mathcal{B}(\mathcal{H})$ is called an *observable*. The real vector space of all observables on $\mathcal{H}$ is denoted by $\mathcal{B}(\mathcal{H})_{\mathrm{s.a.}}$. If AB is a bipartite system, one says that $O \in \mathcal{B}(\mathcal{H}_\mathsf{AB})_{\mathrm{s.a}}$ is a *local*

observable if either $O = A \otimes 1$ or $O = 1 \otimes B$, with $A \in \mathcal{B}(\mathcal{H}_\mathsf{A})_{\text{s.a.}}$ and $B \in \mathcal{B}(\mathcal{H}_\mathsf{B})_{\text{s.a.}}$. Here and in the following, 1 stands for the identity operator on $\mathcal{H}_\mathsf{A}$, $\mathcal{H}_\mathsf{B}$, or another space.

A linear map $\mathcal{M} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ is positive if it transforms a non-negative operator into a non-negative operator. It is completely positive (CP) if the map

$$\mathcal{M} \otimes 1 : X \in \mathcal{B}(\mathcal{H} \otimes \mathbb{C}^m) \ \mapsto \ \sum_{k,l=1}^{m} \mathcal{M}(X_{kl}) \otimes |k\rangle\langle l| \in \mathcal{B}(\mathcal{H}' \otimes \mathbb{C}^m) \tag{4}$$

is positive for any integer $m \geq 1$. Operators acting on the vector space of observables $\mathcal{B}(\mathcal{H})_{\text{s.a.}}$ or on the whole algebra $\mathcal{B}(\mathcal{H})$ are always denoted by calligraphic letters.

Given two orthonormal bases $\{|i\rangle\}_{i=1}^{n_\mathsf{A}}$ of $\mathcal{H}_\mathsf{A}$ and $\{|j\rangle\}_{j=1}^{n_\mathsf{B}}$ of $\mathcal{H}_\mathsf{B}$, one can identify any operator $O : \mathcal{H}_\mathsf{B} \to \mathcal{H}_\mathsf{A}$ with a vector $|\widetilde{\Psi}_O\rangle \in \mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$ thanks to the bijection

$$O \mapsto |\widetilde{\Psi}_O\rangle = \sum_{i,j} \langle i|O|j\rangle |i\rangle|j\rangle \ . \tag{5}$$

This bijection is an isomorphism between the Hilbert spaces $\mathcal{B}(\mathcal{H}_\mathsf{B}, \mathcal{H}_\mathsf{A})$ (endowed with the scalar product (1)) and $\mathcal{H}_{\mathsf{AB}}$. Similarly, one can represent the linear map $\mathcal{M} : \mathcal{B}(\mathcal{H}_\mathsf{B}) \to \mathcal{B}(\mathcal{H}_\mathsf{A})$ by an operator $O_\mathcal{M}$ acting on $\mathcal{H}_{\mathsf{BB}} = \mathcal{H}_\mathsf{B} \otimes \mathcal{H}_\mathsf{B}$ with values in $\mathcal{H}_{\mathsf{AA}} = \mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{A}$. The matrix elements of this operator in the product bases $\{|k\rangle|l\rangle\}_{k,l=1}^{n_\mathsf{B}}$ of $\mathcal{H}_{\mathsf{BB}}$ and $\{|i\rangle|j\rangle\}_{i,j=1}^{n_\mathsf{A}}$ of $\mathcal{H}_{\mathsf{AA}}$ are given by $(O_\mathcal{M})_{ij,kl} = \langle i|\mathcal{M}(|k\rangle\langle l|)|j\rangle$. This representation is an $*$-isomorphism between the C^*-algebras $\mathcal{B}(\mathcal{B}(\mathcal{H}_\mathsf{B}), \mathcal{B}(\mathcal{H}_\mathsf{A}))$, and $\mathcal{B}(\mathcal{H}_{\mathsf{BB}}, \mathcal{H}_{\mathsf{AA}})$. The so-called reshuffling operation[20] associates to $O_\mathcal{M}$ the operator $O_\mathcal{M}^\mathcal{R} \in \mathcal{B}(\mathcal{H}_{\mathsf{AB}})$ with matrix elements $(O_\mathcal{M}^\mathcal{R})_{ik,jl} = (O_\mathcal{M})_{ij,kl}$, which satisfies

$$\langle A \otimes B, \ O_\mathcal{M}^\mathcal{R} \rangle = \langle \widetilde{\Psi}_A | O_\mathcal{M} J | \widetilde{\Psi}_B \rangle = \langle A, \mathcal{M}(\overline{B}) \rangle \tag{6}$$

for any $A \in \mathcal{B}(\mathcal{H}_\mathsf{A})$ and $B \in \mathcal{B}(\mathcal{H}_\mathsf{B})$. Here J denotes the anti-unitary operator on $\mathcal{H}_{\mathsf{BB}}$ defined by $\langle k|\langle l|J|\Psi\rangle = \overline{\langle k|\langle l|\Psi\rangle}$ (complex conjugation in the canonical basis) and $\overline{B} = \sum_{k,l} \overline{\langle k|B|l\rangle} |k\rangle\langle l|$ is the operator associated to $J|\widetilde{\Psi}_B\rangle$ via the isomorphism (5). With these definitions, $\mathcal{M} : \mathcal{B}(\mathcal{H}_\mathsf{B}) \to \mathcal{B}(\mathcal{H}_\mathsf{A})$ is CP if and only if $O_\mathcal{M}^\mathcal{R} \geq 0$, that is, $O_\mathcal{M}^\mathcal{R}$ has non-negative eigenvalues. (Actually, $O_\mathcal{M}^\mathcal{R} \geq 0$ is equivalent to $O_\mathcal{M}^\mathcal{R} = A^* A$ for some $A \in \mathcal{B}(\mathcal{H}_{\mathsf{AB}})$, that is, to $(O_\mathcal{M}^\mathcal{R})_{ik,jl} = \langle i|\mathcal{M}(|k\rangle\langle l|)|j\rangle = \sum_{p,q} \overline{A_{pq,ik}} A_{pq,jl}$ for all $i, j = 1, \ldots, n_\mathsf{A}$ and $k, l = 1, \ldots, n_\mathsf{B}$. Setting $A_{pq} = \sum_{i,k} \overline{A_{pq,ik}} |i\rangle\langle k|$, it follows that $O_\mathcal{M}^\mathcal{R} \geq 0$ if and only if $\mathcal{M}(X) = \sum_{pq} A_{pq} X A_{pq}^*$ for all $X \in \mathcal{B}(\mathcal{H}_\mathsf{B})$, which is equivalent to $\mathcal{M}$ being CP by the Kraus representation theorem (Theorem 3.B.3 below).)

The left and right multiplications $\mathcal{L}_X$ and $\mathcal{R}_X$ by $X \in \mathcal{B}(\mathcal{H})$ are the operators from $\mathcal{B}(\mathcal{H})$ into itself defined by

$$\mathcal{L}_X(Y) = XY, \quad \mathcal{R}_X(Y) = YX, \quad \forall \ Y \in \mathcal{B}(\mathcal{H}). \tag{7}$$

They are represented on $\mathcal{B}(\mathcal{H} \otimes \mathcal{H})$ by local operators $X \otimes 1$ and $1 \otimes X^T$, respectively, where T stands for the transposition in the basis $\{|i\rangle\}$. In the C^*-algebra setting, the map $X \mapsto \mathcal{L}_X$ is the Gelfand-Neumark-Segal representation of the C^*-algebra.[29] Given two states ρ and $\sigma \in \mathcal{E}(\mathcal{H})$ with $\rho > 0$, the Araki relative modular operator $\Delta_{\sigma|\rho}$ is defined by[11]

$$\Delta_{\sigma|\rho}(Y) = \sigma Y \rho^{-1} = \mathcal{L}_\sigma \circ \mathcal{R}_{\rho^{-1}}(Y), \quad \forall \ Y \in \mathcal{B}(\mathcal{H}) \ . \tag{8}$$

It is a self-adjoint non-negative operator on the Hilbert space $\mathcal{B}(\mathcal{H})$ (for the scalar product (1)).

B. The Schmidt decomposition

The following standard result is very useful in quantum information theory.

Theorem 2.B.1. (Schmidt decomposition) *Any pure state* $|\Psi\rangle \in \mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$ *of a bipartite system admits a decomposition*

$$|\Psi\rangle = \sum_{i=1}^{n} \sqrt{\mu_i} |\alpha_i\rangle|\beta_i\rangle \tag{9}$$

where $n = \min\{n_{\mathsf{A}}, n_{\mathsf{B}}\}$, $\mu_i \geq 0$, *and* $\{|\alpha_i\rangle\}_{i=1}^{n}$ *(respectively,* $\{|\beta_i\rangle\}_{i=1}^{n}$*) is an orthonormal family of* $\mathcal{H}_{\mathsf{A}}$ *(respectively, of* $\mathcal{H}_{\mathsf{B}}$*). The* μ_i *and* $|\alpha_i\rangle$ *(respectively,* $|\beta_i\rangle$*) are the eigenvalues and eigenvectors of the reduced state* $\rho_{\mathsf{A}} = \mathrm{tr}_{\mathsf{B}}(|\Psi\rangle\langle\Psi|)$ *(respectively,* $\rho_{\mathsf{B}} = \mathrm{tr}_{\mathsf{A}}(|\Psi\rangle\langle\Psi|)$*). Thus, if the eigenvalues* μ_i *are non-degenerate then the decomposition (9) is unique.*

The non-negative numbers μ_i are called the *Schmidt coefficients* of $|\Psi\rangle$. They satisfy $\sum_i \mu_i = \|\Psi\|^2 = 1$.

Proof. Let $\{|i\rangle\}_{i=1}^{n_{\mathsf{A}}}$ and $\{|j\rangle\}_{j=1}^{n_{\mathsf{B}}}$ be some fixed orthonormal bases of $\mathcal{H}_{\mathsf{A}}$ and $\mathcal{H}_{\mathsf{B}}$. By using the isomorphism $|\Psi\rangle \mapsto O_\Psi = \sum_{i,j}\langle i \otimes j|\Psi\rangle|i\rangle\langle j|$ between $\mathcal{H}_{\mathsf{AB}}$ and the space of $n_{\mathsf{A}} \times n_{\mathsf{B}}$ matrices (see Sec. II A), we observe that the decomposition (9) corresponds to the singular value decomposition of O_Ψ, that is, $O_\Psi = U_{\mathsf{A}} \sum_i \sqrt{\mu_i}|i\rangle\langle i|U_{\mathsf{B}}^*$ with μ_i the eigenvalues of $O_\Psi^* O_\Psi$ and U_{A} and U_{B} unitaries on $\mathcal{H}_{\mathsf{A}}$ and $\mathcal{H}_{\mathsf{B}}$. Then $U_{\mathsf{A}}|i\rangle = |\alpha_i\rangle$ and $U_{\mathsf{B}}|i\rangle = |\beta_i^*\rangle$ are eigenvectors of $O_\Psi O_\Psi^*$ and $O_\Psi^* O_\Psi$, respectively. Denoting by J is the complex conjugation in the basis $\{|j\rangle\}$ (see above), one has $|\beta_i\rangle = J|\beta_i^*\rangle$. □

The Schmidt decomposition can be generalized to mixed states by considering ρ as a vector in the Hilbert space $\mathcal{B}(\mathcal{H}_{\mathsf{A}}) \otimes \mathcal{B}(\mathcal{H}_{\mathsf{B}})$. Any $\rho \in \mathcal{E}(\mathcal{H}_{\mathsf{AB}})$ can be written as

$$\rho = \sum_{m=1}^{n^2} \sqrt{\mu_m} X_m \otimes Y_m, \tag{10}$$

where $\{X_m\}_{m=1}^{n_{\mathsf{A}}^2}$ and $\{Y_m\}_{m=1}^{n_{\mathsf{B}}^2}$ are orthonormal bases of $\mathcal{B}(\mathcal{H}_{\mathsf{A}})$ and $\mathcal{B}(\mathcal{H}_{\mathsf{B}})$ for the scalar product (1) and μ_m are the eigenvalues of the $n_{\mathsf{A}}^2 \times n_{\mathsf{A}}^2$ matrix $R \geq 0$ defined by

$$R_{ij,i'j'} = \Big\langle \rho\,|i\rangle\langle i'| \otimes 1,\ |j\rangle\langle j'| \otimes 1\,\rho \Big\rangle \tag{11}$$

(the $R_{ij,i'j'}$ are the matrix elements in the orthonormal basis $\{|i\rangle\langle j|\}_{i,j=1}^{n_{\mathsf{A}}^2}$ of $\mathcal{B}(\mathcal{H}_{\mathsf{A}})$ of the operator playing the role of the reduced state in Theorem 2.B.1). Note that $\sum_m \mu_m = \mathrm{tr}(\rho^2) \leq 1$, with equality if and only if ρ is a pure state.

Remark 2.B.2. Alternatively, the μ_m *are the square roots of the singular values of* $\rho^{\mathcal{R}} \in \mathcal{B}(\mathcal{H}_{\mathsf{BB}}, \mathcal{H}_{\mathsf{AA}})$*, where* $\mathcal{R}$ *is the reshuffling operation (Sec. II A), and* X_m *and* Y_m *are given in terms of the eigenvectors* $|\chi_m\rangle$ *and* $|\psi_m\rangle$ *of* $\rho^{\mathcal{R}}(\rho^{\mathcal{R}})^*$ *and* $(\rho^{\mathcal{R}})^*\rho^{\mathcal{R}}$ *by* $X_m = \sum_{i,j}\langle i \otimes j|\chi_m\rangle|i\rangle\langle j|$ *and* $Y_m = \sum_{k,l}\overline{\langle k \otimes l|\psi_m\rangle}|k\rangle\langle l|$*, respectively.*

Proof. Considering ρ as a vector in $\mathcal{B}(\mathcal{H}_{\mathsf{A}}) \otimes \mathcal{B}(\mathcal{H}_{\mathsf{B}})$ and introducing two orthonormal bases $\{A_p\}$ of $\mathcal{B}(\mathcal{H}_{\mathsf{A}})$ and $\{B_q\}$ of $\mathcal{B}(\mathcal{H}_{\mathsf{B}})$, according to the proof of Theorem 2.B.1, $\sqrt{\mu_m}$ are the singular values of the $n_{\mathsf{A}}^2 \times n_{\mathsf{B}}^2$ matrix $(\langle A_p \otimes B_q,\ \rho\rangle)_{p,q}$. Denote by $\{|\alpha_p\rangle\}$ and $\{|\beta_q\rangle\}$ the orthonormal bases of $\mathcal{H}_{\mathsf{AA}}$ and $\mathcal{H}_{\mathsf{BB}}$ associated to $\{A_p\}$ and $\{B_q\}$ via the isomorphism (5). The statement follows by choosing $A_p = |i\rangle\langle j|$ and $B_q = |k\rangle\langle l|$ and using the identity $\langle\alpha_p|\rho^{\mathcal{R}}J|\beta_q\rangle = \langle A_p \otimes B_q, \rho\rangle$, see (6). □

C. Purifications and pure state decompositions of mixed states

Definition 2.C.1. *Let* ρ *be an arbitrary state on* $\mathcal{H}$ *and* $\mathcal{K}$ *be another Hilbert space. A pure state* $|\Psi\rangle \in \mathcal{H} \otimes \mathcal{K}$ *such that* $\rho = \mathrm{tr}_{\mathcal{K}}(|\Psi\rangle\langle\Psi|)$ *is called a purification of* ρ *on* $\mathcal{H} \otimes \mathcal{K}$*.*

In the language of C^*-algebras, a purification is an example of cyclic representation of a state.[29] An example of purification of ρ on $\mathcal{H} \otimes \mathcal{H}$ is

$$|\Psi\rangle = \sum_{k=1}^{n} \sqrt{p_k}|k\rangle|k\rangle, \tag{12}$$

where $\rho = \sum_k p_k |k\rangle\langle k|$ is a spectral decomposition of ρ. If $|\Psi\rangle$ and $|\Phi\rangle$ are two purifications of ρ on the same space $\mathcal{H}\otimes\mathcal{K}$, then there exists a unitary operator U acting on $\mathcal{K}$ such that $|\Phi\rangle = 1\otimes U|\Psi\rangle$. In fact, one infers from the Schmidt decomposition that $|\Psi\rangle = \sum_k \sqrt{p_k}|k\rangle|f_k\rangle$ and $|\Phi\rangle = \sum_k \sqrt{p_k}|k\rangle|g_k\rangle$, where $\{|f_k\rangle\}_{k=1}^n$ and $\{|g_k\rangle\}_{k=1}^n$ are two orthonormal families of $\mathcal{K}$. Thus $|g_k\rangle = U|f_k\rangle$ for some unitary U.

We will often be interested in the sequel by families of quantum states of a system S, $\rho_i \in \mathcal{E}(\mathcal{H}_\mathsf{S})$, $i = 1, \ldots, m$, to which we attach some probabilities $\eta_i \geq 0$, $\sum_i \eta_i = 1$. Following the terminology employed by physicists in statistical physics, we call $\{\rho_i, \eta_i\}_{i=1}^m$ an *ensemble of quantum states* (or more simply an *ensemble*). A *convex decomposition* of ρ is an ensemble $\{\rho_i, \eta_i\}_{i=1}^m$ such that $\rho = \sum_i \eta_i\rho_i$. A *pure state decomposition* of ρ is a convex decomposition in terms of finitely many pure states $\rho_i = |\psi_i\rangle\langle\psi_i|$, i.e.,

$$\rho = \sum_{i=1}^m \eta_i |\psi_i\rangle\langle\psi_i| \,. \tag{13}$$

If the vectors $|\psi_i\rangle$ are orthogonal, then (13) coincides with the spectral decomposition, but we will see that there are infinitely many other ways to decompose ρ. Physically, (13) describes a *state preparation*: it means that the system has been prepared in the pure state $|\psi_i\rangle$ with probability η_i. The non-uniqueness of the decomposition can be interpreted as follows. If a receiver is given two ensembles $\{|\psi_i\rangle, \eta_i\}_{i=1}^m$ and $\{|\phi_j\rangle, \xi_j\}_{j=1}^p$ corresponding to different state preparations of two identical systems in the same state ρ, then he cannot make any difference between them if he has no prior knowledge on the state preparation. Indeed, any measurement performed by him gives rise to the same distribution of outcomes for the two ensembles. In other words, the full information that the receiver can collect on the system via measurements is encoded in ρ, and *not* in the ensemble involved in the state preparation. This very important fact has consequences that are sometimes disconcerting to people unfamiliar with the conceptual aspects of quantum mechanics. For instance, if a preparer gives a maximally mixed state $\rho = 1/n$ to a receiver, the latter has no way to decide whether this state was prepared from n equiprobable orthonormal pure states (which are only known by the preparer) or if it was prepared by another procedure involving more than n states. It is also worth mentioning that the process transforming the ensemble $\{\rho_i, \eta_i\}_{i=1}^m$ into the average state $\rho = \sum_i \eta_i\rho_i$, which can be viewed as the inverse of a convex decomposition, corresponds physically to a loss of information about the state preparation.

Given a fixed orthonormal basis $\{|f_i\rangle\}_{i=1}^p$ of $\mathcal{K}$ with $p \geq \mathrm{ran}(\rho) = r$, there is a one-to-one correspondence between pure state decompositions of ρ containing at most p states and purifications of ρ on $\mathcal{H}\otimes\mathcal{K}$. Actually, given the pure state decomposition (13),

$$|\Psi\rangle = \sum_{i=1}^p \sqrt{\eta_i}|\psi_i\rangle|f_i\rangle \tag{14}$$

defines a purification of ρ on $\mathcal{H}\otimes\mathcal{K}$ (we have set $\eta_i = 0$ for $m < i \leq p$). Reciprocally, let $|\Psi\rangle$ be a purification of ρ on $\mathcal{H}\otimes\mathcal{K}$. Denote as before the eigenvalues and orthonormal eigenvectors of ρ by p_k and $|k\rangle$. As argued above, one can find a unitary U on $\mathcal{K}$ such that

$$|\Psi\rangle = \sum_{k=1}^r \sqrt{p_k}|k\rangle U|f_k\rangle = \sum_{i=1}^p\sum_{k=1}^r \sqrt{p_k}\langle f_i|U|f_k\rangle|k\rangle|f_i\rangle = \sum_{i=1}^p \sqrt{\eta_i}|\psi_i\rangle|f_i\rangle \tag{15}$$

with $\sqrt{\eta_i}|\psi_i\rangle = \sum_k \sqrt{p_k}\langle f_i|U|f_k\rangle|k\rangle$. Hence $|\Psi\rangle$ has the form (14). Taking the partial trace over $\mathcal{K}$, one can associate to it a unique pure state decomposition, which is given by (13).

Since two purifications $|\Psi\rangle$ and $|\Phi\rangle$ of the same state ρ are related by a local unitary U acting on the ancilla space $\mathcal{K}$, this implies that any two pure state decompositions $\rho = \sum_{i=1}^m \eta_i|\psi_i\rangle\langle\psi_i|$ and $\rho = \sum_{j=1}^p \xi_j|\phi_j\rangle\langle\phi_j|$ are related by

$$\sqrt{\xi_j}|\phi_j\rangle = \sum_{i=1}^{\max\{m,p\}} u_{ji}\sqrt{\eta_i}|\psi_i\rangle \,, \tag{16}$$

where (u_{ji}) is a unitary matrix with size $\max\{m, p\}$ (if $m < i \leq p$ we set as before $\eta_i = 0$).

D. Entangled and separable states

Let us consider a bipartite system AB. If this system is in a tensor product state $|\Psi_{\rm sep}\rangle = |\psi_{\rm A}\rangle|\phi_{\rm B}\rangle$ with $|\psi_{\rm A}\rangle \in \mathcal{H}_{\rm A}$ and $|\phi_{\rm B}\rangle \in \mathcal{H}_{\rm B}$, then the expectation value of the product of two local observables $A \otimes 1$ and $1 \otimes B$ coincides with the product of the expectations values, i.e.,

$$G_{AB}(|\Psi_{\rm sep}\rangle) = \langle\Psi_{\rm sep}|A \otimes B|\Psi_{\rm sep}\rangle - \langle\Psi_{\rm sep}|A \otimes 1|\Psi_{\rm sep}\rangle\langle\Psi_{\rm sep}|1 \otimes B|\Psi_{\rm sep}\rangle = 0\ . \tag{17}$$

This means that the random outcomes of measurements of the local observables $A \otimes 1$ and $1 \otimes B$ are uncorrelated. More generally, if one thinks of AB as a pair of particles located far apart (e.g. a photon pair shared by two observers Alice and Bob), this pair is in a product state if and only if there are no correlations between the results of arbitrary local measurements performed independently on each particle (for instance, if Alice sends her photon through a polarizer and then to a photodetector, and Bob does the same with his photon, the clicks of the two detectors will be uncorrelated whatever the polarizer angles). One says that $|\Psi_{\rm sep}\rangle = |\psi_{\rm A}\rangle|\phi_{\rm B}\rangle$ is a *separable state*. If the pure state $|\Psi\rangle \in \mathcal{H}_{\rm AB}$ is not a product state one says that it is *entangled*.

By applying the Schmidt decomposition, one sees that $|\Psi\rangle$ is separable if and only if all its Schmidt coefficients vanish except one, that is, if and only if its reduced states $\rho_{\rm A}$ and $\rho_{\rm B}$ are pure. In the opposite, if either $\rho_{\rm A}$ or $\rho_{\rm B}$ is proportional to the identity matrix (maximally mixed state), we say that $|\Psi\rangle$ is *maximally entangled*. Such states have the form

$$|\Psi_{\rm ent}\rangle = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}|\alpha_i\rangle|\beta_i\rangle\ , \tag{18}$$

where $\{|\alpha_i\rangle\}_{i=1}^n$ and $\{|\beta_i\rangle\}_{i=1}^n$ are orthonormal families in $\mathcal{H}_{\rm A}$ and $\mathcal{H}_{\rm B}$ and $n = \min\{n_{\rm A}, n_{\rm B}\}$. For instance, denoting by $|0\rangle$ and $|1\rangle$ the canonical basis vectors of $\mathbb{C}^2$, the EPR (or Bell) states $|\Phi_\pm\rangle = (|0\rangle|0\rangle \pm |1\rangle|1\rangle)/\sqrt{2}$ and $|\Psi_\pm\rangle = (|0\rangle|1\rangle \pm |1\rangle|0\rangle)/\sqrt{2}$ are maximally entangled states of two qubits, and any maximally entangled two-qubit state is an EPR state, up to a local unitary transformation $U_{\rm A} \otimes U_{\rm B}$.

For mixed states, entanglement is no longer equivalent to being a product state. The "good" definition of mixed state entanglement is due to Werner.[165]

Definition 2.D.1. *A mixed state ρ of a bipartite system* AB *is separable if it admits a pure state decomposition*

$$\rho = \sum_i \eta_i |\psi_i \otimes \phi_i\rangle\langle\psi_i \otimes \phi_i| \tag{19}$$

in terms of pure separable states $|\psi_i \otimes \phi_i\rangle \in \mathcal{H}_{\rm AB}$. If such a decomposition does not exist then ρ is entangled. The set of all separable states of AB *forms a convex subset of $\mathcal{E}(\mathcal{H}_{\rm AB})$, which is denoted by $\mathcal{S}_{\rm AB}$.*

It follows from the Carathéodory theorem that the number of pure product states in the decomposition (19) can always be chosen to be smaller or equal to $(n_{\rm A}n_{\rm B})^2 + 1$.

According to this definition, a state is separable if it *could* have been prepared from pure product states only. This does not mean that it has actually been prepared using such states. For example, if one prepares two qubits in the maximally entangled states $|\Phi_+\rangle$ and $|\Phi_-\rangle$ with equal probabilities, the corresponding state

$$\rho = \frac{1}{2}|\Phi_+\rangle\langle\Phi_+| + \frac{1}{2}|\Phi_-\rangle\langle\Phi_-| = \frac{1}{2}|0\rangle\langle 0| \otimes |0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| \otimes |1\rangle\langle 1| \tag{20}$$

is separable! This unexpected result is inherent to the ambiguity of the state preparation discussed in Subsection II C. This quantum ambiguity unfortunately obliges us to look for all possible state preparations of a given mixed state ρ to decide whether ρ is entangled or not. This makes this problem highly non-trivial.

An explicit complete characterization of $\mathcal{S}_{\rm AB}$ is known for qubit-qubit and qubit-qutrit systems only, that is, for $(n_{\rm A}, n_{\rm B}) = (2, 2)$, $(2, 3)$, and $(3, 2)$. In such a case, the Peres-Horodecki criterion[125,80,81] gives a necessary and sufficient condition for ρ to be entangled. This criterion is

formulated in terms of the partial transpose. Given two orthonormal bases $\{|i\rangle\}$ of $\mathcal{H}_\mathsf{A}$ and $\{|k\rangle\}$ of $\mathcal{H}_\mathsf{B}$, the partial transpose ρ^{T_B} of ρ with respect to B has matrix elements in the product basis $\{|i\rangle|k\rangle\}$ given by

$$\langle i|\langle k|\rho^{T_\mathsf{B}}|j\rangle|l\rangle = \langle i|\langle l|\rho|j\rangle|k\rangle \ . \tag{21}$$

One defines similarly ρ^{T_A} and note that $\rho^{T_\mathsf{A}} = (\rho^{T_\mathsf{B}})^T$. It follows from Definition 2.D.1 that if ρ is separable then $\rho^{T_\mathsf{A}} \geq 0$ and $\rho^{T_\mathsf{B}} \geq 0$, i.e., ρ^{T_A} and ρ^{T_B} are states of AB. Thus, if ρ^{T_A} (or, equivalently, ρ^{T_B}) has negative eigenvalues then ρ is necessarily entangled. Since the transpose is a positive but not CP map, such negative eigenvalues may indeed exist. However, if $n_\mathsf{A} n_\mathsf{B} > 6$, certain entangled states have non-negative partial transposes.[81] It is remarkable that this does not happen when $n_\mathsf{A} n_\mathsf{B} \leq 6$: then $\rho^{T_\mathsf{A}} \geq 0$ if and only if $\rho \in \mathcal{S}_\mathsf{AB}$.[80] Two remarks should be made at this point. First, states with non-negative partial transposes cannot undergo entanglement distillation and therefore form an interesting subset of $\mathcal{E}(\mathcal{H}_\mathsf{AB})$ on their own, which contains $\mathcal{S}_\mathsf{AB}$ (see Ref. 82 for more detail). Second, extending the Peres criterion to all positive but not CP linear maps $\Lambda_\mathsf{B} : \mathcal{B}(\mathcal{H}_\mathsf{B}) \to \mathcal{B}(\mathcal{H}_\mathsf{A})$ (i.e., asking that $1 \otimes \Lambda_\mathsf{B}(\rho) \geq 0$ for any such map) yields a necessary and sufficient condition for entanglement, valid whatever the space dimensions n_A and n_B.[80] Due to the lack of an explicit characterization of such maps, this condition is unfortunately not very helpful in general except for $(n_\mathsf{A}, n_\mathsf{B}) = (2, 2)$ or $(3, 2)$ (When $(n_\mathsf{A}, n_\mathsf{B}) = (2, 2)$ or $(3, 2)$, any positive map $\Lambda : \mathcal{B}(\mathcal{H}_\mathsf{B}) \to \mathcal{B}(\mathcal{H}_\mathsf{A})$ can be written as $\Lambda = \mathcal{M}_1 + \mathcal{M}_2 \circ T$, where $\mathcal{M}_1$ and $\mathcal{M}_2$ are CP and T is the transposition.[170] The fact that the partial transpose criterion is sufficient for entanglement follows from this characterization.[80])

Let us also mention another necessary but not sufficient (even for two qubits) condition for entanglement, which relies on the Schmidt decomposition (10) for mixed states. By using the fact that $\sum_m \sqrt{\mu_m}$ defines a norm on $\mathcal{E}(\mathcal{H}_\mathsf{AB})$, one can show that if $\rho \in \mathcal{S}_\mathsf{AB}$ then $\sum_m \sqrt{\mu_m} \leq 1$.[39] Hence $\sum_m \sqrt{\mu_m} > 1$ implies that ρ is entangled.

Once a state has been recognized as separable, it may be of relevance to determine its decomposition(s) into pure product states. This problem has been addressed in Refs. 169, 137 and 162 for two qubits.

Definition 2.D.1 can be extended straightforwardly to multipartite systems $\mathsf{A}_1 \ldots \mathsf{A}_k$. Then different kinds of entanglement can be defined according to the chosen partition of $\{\mathsf{A}_1, \ldots, \mathsf{A}_k\}$. In this article we will not consider multipartite entanglement, which is a challenging subject in its own.[67,82]

III. QUANTUM MEASUREMENTS

In this section we review the notions of quantum operations and generalized measurements and give the basic theorems, namely, the Stinespring theorem, the Kraus decomposition, and the Neumark extension theorem. We start by a physical description of a von Neumann measurement.

A. Physical realization of a measurement process

A measurement on a quantum system S is realized by coupling S with a measurement apparatus. This apparatus consists of a macroscopic pointer P interacting with an environment E playing the role of an infinite bath. One may think of P as the center of mass of the needle of a meter. The environment E then includes all the other degrees of freedom of the macroscopic apparatus. The coupling of the measured system S with the pointer transforms the initially uncorrelated state $|\psi\rangle \otimes |0\rangle$ of the composite system SP into an entangled state,

$$|\psi\rangle \otimes |0\rangle \longrightarrow |\Psi_\mathsf{SP}^\mathrm{ent}\rangle = \sum_{i,l} c_{il} |\alpha_{il}\rangle \otimes |i\rangle \ . \tag{22}$$

Our assumption that S and P are initially in pure states is made to simplify the foregoing discussion and can be easily relaxed. The states $|\alpha_{il}\rangle$ form an orthonormal basis of the system Hilbert space $\mathcal{H}_\mathsf{S}$ (measurement basis), which is the eigenbasis of the measured observable A, i.e., $A|\alpha_{il}\rangle = a_i|\alpha_{il}\rangle$. The index l labels if necessary the different orthogonal eigenstates of A with the same degenerate

eigenvalue a_i. In ideal measurements $c_{il} = \langle \alpha_{il}|\psi\rangle$. The states $|i\rangle$ are the pointer states of the apparatus. After a sufficiently long coupling time between S and P, these states are macroscopically distinct and thus nearly orthogonal, $\langle i|j\rangle \simeq \delta_{ij}$ (hereafter δ_{ij} stands for the Kronecker symbol, equal to 1 if $i = j$ and zero otherwise). The transformation (22) is a unitary transformation, i.e., $|\Psi_{\mathsf{SP}}^{\mathrm{ent}}\rangle = U_{\mathsf{SP}}|\psi\rangle|0\rangle$ where U_{SP} is a unitary evolution operator on $\mathcal{H}_{\mathsf{SP}}$. One usually calls such a transformation the *pre-measurement*.[64] This unitary evolution induces quantum correlations between S and P, such that each eigenprojector $\Pi_i = \sum_l |\alpha_{il}\rangle\langle\alpha_{il}|$ of A is in one-to-one correspondence with a pointer state $|i\rangle$. The resulting state (22) is a superposition of macroscopically distinct states, sometimes referred to as a "Schrödinger cat state." The pointer states are singled out by their robustness against environment-induced decoherence. More precisely, if the pointer P is initially in the state $|i\rangle$, its interaction with the environment E does not entangle P and E. Letting P and E interact during a time t much larger than the decoherence time, the SP-entangled state $|\Psi_{\mathsf{SP}}^{\mathrm{ent}}\rangle$ is transformed into a statistical mixture in which all the coherences between the pointer states $|i\rangle$ have disappeared. After tracing out the environment degrees of freedom, the reduced state of SP is modified according to

$$|\Psi_{\mathsf{SP}}^{\mathrm{ent}}\rangle\langle\Psi_{\mathsf{SP}}^{\mathrm{ent}}| \longrightarrow \rho_{\mathsf{SP}}^{\mathrm{p.m.}} = \sum_{ikl} c_{ik}\overline{c_{il}}|\alpha_{ik}\rangle\langle\alpha_{il}| \otimes |i\rangle\langle i| = \sum_i \Pi_i\,\rho\,\Pi_i \otimes |i\rangle\langle i|\,, \tag{23}$$

$\rho = |\psi\rangle\langle\psi|$ being the initial system state. The final SP-state has no quantum correlations but is classically correlated: indeed, each pointer state $|i\rangle$ goes hand in hand with the system state

$$\rho_{S|i} = p_i^{-1}\Pi_i\,\rho\,\Pi_i, \quad p_i = \mathrm{tr}(\Pi_i\rho)\,. \tag{24}$$

Concrete models for the pointer and its coupling with the system and the environment have been investigated in Refs. 2, 3, 147, and 148; in these works the aforementioned decoherence time and the time duration of the measurement are estimated in the more realistic situation where the two transformations (22) and (23) occur simultaneously. The readout of the pointer (that is, the observation of the position of the needle) cannot significantly alter the macroscopic state $|i\rangle$. It merely selects one of the measurement outcomes,

$$\textbf{outcome } i\text{:} \qquad \rho_{\mathsf{SP}}^{\mathrm{p.m.}} \longrightarrow \rho_{\mathsf{SP}|i} = \rho_{\mathsf{S}|i} \otimes |i\rangle\langle i| \qquad \text{(wavepacket reduction)}. \tag{25}$$

After the measurement yielding the outcome i the measured system is in the conditional state $\rho_{S|i}$, and this outcome occurs with probability p_i (Born rule). The transformation (25) results from the knowledge of the random outcome, it should not be regarded as a true dynamical process. It is actually analog to a state preparation (see Sec. II C). In mathematical terms, it corresponds to a convex decomposition of $\rho_{\mathsf{SP}}^{\mathrm{p.m.}} = \sum_i p_i \rho_{\mathsf{SP}|i}$.

We point out that recent progresses in the understanding of quantum measurement processes via dynamical models and their interpretation with a statistical physics viewpoint have been made by Allahverdyan, Balian, and Nieuwenhuizen.[4]

B. Quantum operations

In the absence of readout of the measurement result, one does not know which state $\rho_{S|i}$ has been prepared and the system is after the measurement in the average state

$$\mathcal{M}_\Pi(\rho) = \sum_i \Pi_i\,\rho\,\Pi_i, \tag{26}$$

where ρ is the state before the measurement.

Since $\{\Pi_i\}$ is the spectral measure of the self-adjoint operator A, the Π_i form a family of projectors in $\mathcal{B}(\mathcal{H}_{\mathsf{S}})_{\mathrm{s.a.}}$ satisfying $\Pi_i\Pi_j = \delta_{ij}\Pi_i$ and $\sum_i \Pi_i = 1$. We will refer in the sequel to such a family as an *orthonormal family of projectors*. It is easy to show that the map $\mathcal{M}_\Pi$ is CP (as a sum of CP maps) and trace-preserving. In quantum information, such maps are called *quantum operations*.

Definition 3.B.1. *A quantum operation $\mathcal{M} : \mathcal{B}(\mathcal{H}_{\mathsf{S}}) \to \mathcal{B}(\mathcal{H}'_{\mathsf{S}})$ is a trace-preserving CP map from $\mathcal{B}(\mathcal{H}_{\mathsf{S}})$ into $\mathcal{B}(\mathcal{H}'_{\mathsf{S}})$.*

A necessary and sufficient condition for a linear map $\mathcal{M} : \mathcal{B}(\mathcal{H}_\mathsf{S}) \to \mathcal{B}(\mathcal{H}'_\mathsf{S})$ to be CP is that it satisfies $\mathcal{M} \otimes 1(|\Psi_{\rm ent}\rangle\langle\Psi_{\rm ent}|) \geq 0$ for the maximally entangled state $|\Psi_{\rm ent}\rangle = n_\mathsf{S}^{-1/2} \sum_k |k\rangle|k\rangle$ in $\mathcal{H}_\mathsf{S} \otimes \mathcal{H}_\mathsf{S}$, where $\{|k\rangle\}$ is an orthonormal basis of $\mathcal{H}_\mathsf{S}$. In fact, $\mathcal{M} \otimes 1(|\Psi_{\rm ent}\rangle\langle\Psi_{\rm ent}|)$ coincides with the operator $O^{\mathcal{R}}_{\mathcal{M}}$ defined in Sec. II A up to a factor $1/n_\mathsf{S}$, and it has been argued above that $\mathcal{M}$ is CP if and only if $O^{\mathcal{R}}_{\mathcal{M}} \geq 0$.

A quantum operation is the quantum analog of a stochastic matrix $\mathcal{M}^{\rm clas}$ giving the transition probabilities $q(j|i)$ of a classical Markov process,

$$\mathbf{p} = (p_1, \ldots, p_n) \;\mapsto\; \mathcal{M}^{\rm clas}\mathbf{p} \quad \text{with} \quad (\mathcal{M}^{\rm clas}\mathbf{p})_j = \sum_{i=1}^{n} q(j|i)\, p_i, \;\; q(j|i) \geq 0, \quad \sum_{j=1}^{n} q(j|i) = 1. \tag{27}$$

Save for the wavepacket reduction (25), all physical dynamical processes on quantum systems are given by quantum operations. In order to include the transformation (25), many authors define a more general notion of quantum operation by relaxing the trace-preserving condition and replacing it by $\mathrm{tr}(\mathcal{M}(\rho)) \leq 1$ for any $\rho \in \mathcal{E}(\mathcal{H})$. The state transformation is then given by the non-linear map $\rho \mapsto \mathcal{M}(\rho)/\,\mathrm{tr}(\mathcal{M}(\rho))$. Theorems 3.B.2 and 3.B.3 below can easily be adapted to this more general definition (in particular, the Kraus decomposition (31) holds, with Kraus operators A_i satisfying $\sum_i A_i^* A_i \leq 1$). Let a system S interact with another system E at times $t \geq 0$. If S and E are initially in a product state $\rho(0) \otimes \rho_\mathsf{E}(0)$ and SE can be considered as an isolated system, so that its dynamics is governed by the Schrödinger equation, then the reduced state of S at time t reads

$$\rho(t) = \mathrm{tr}_\mathsf{E}\big(e^{-\mathrm{i}tH_\mathsf{SE}}\rho(0) \otimes \rho_\mathsf{E}(0)e^{\mathrm{i}tH_\mathsf{SE}}\big)\,. \tag{28}$$

Here $H_\mathsf{SE} = H_\mathsf{S} + H_\mathsf{E} + \lambda H_{\rm int}$ is the Hamiltonian of SE, where H_S and H_E are the Hamiltonians of S and E, $H_{\rm int}$ their coupling Hamiltonian, and λ the coupling constant. The time-evolved state (28) is related to the initial state $\rho(0)$ by a quantum operation $\mathcal{M}_t$, i.e., $\rho(t) = \mathcal{M}_t\rho(0)$. The Stinespring theorem says that any quantum operation $\mathcal{M}$ can be viewed as a reduced evolution of the system coupled to an auxiliary system (ancilla).

Theorem 3.B.2. (Stinespring[151]) *Let $\mathcal{M}$ be a quantum operation $\mathcal{B}(\mathcal{H}_\mathsf{S}) \to \mathcal{B}(\mathcal{H}_\mathsf{S})$. Then one can find an ancilla Hilbert space $\mathcal{H}_\mathsf{E}$, a state $|\epsilon_0\rangle \in \mathcal{H}_\mathsf{E}$, and a unitary operator U on $\mathcal{H}_\mathsf{SE}$ such that $\mathcal{M}(\rho) = \mathrm{tr}_\mathsf{E}(U\rho \otimes |\epsilon_0\rangle\langle\epsilon_0|\, U^*)$.*

It is appropriate at this point to review a few well-known facts from the theory of CP maps on C^*-algebras. The *adjoint* $\mathcal{M}^*$ with respect to the trace of $\mathcal{M} : \mathcal{B}(\mathcal{H}_\mathsf{S}) \to \mathcal{B}(\mathcal{H}'_\mathsf{S})$ is the map $\mathcal{M}^* : \mathcal{B}(\mathcal{H}'_\mathsf{S}) \to \mathcal{B}(\mathcal{H}_\mathsf{S})$ defined by $\mathrm{tr}[A\mathcal{M}(\rho)] = \mathrm{tr}[\mathcal{M}^*(A)\rho]$ for any $A \in \mathcal{B}(\mathcal{H}'_\mathsf{S})$ and $\rho \in \mathcal{B}(\mathcal{H}_\mathsf{S})$. If $\mathcal{M}$ is a quantum operation then $\mathcal{M}^*$ is also a CP map and is unity-preserving, $\mathcal{M}^*(1) = 1$. According to Stinespring's theorem, one has

$$\mathcal{M}^*(X) = \langle\epsilon_0|U^*X \otimes 1U|\epsilon_0\rangle \tag{29}$$

for any $X \in \mathcal{B}(\mathcal{H})$. It follows that $\mathcal{M}^*$ satisfies the Kadyson-Schwarz inequality

$$|\mathcal{M}^*(X)|^2 \leq \mathcal{M}^*(|X|^2)\,. \tag{30}$$

Theorem 3.B.3. (Kraus[95]) *A linear map $\mathcal{M}$ from $\mathcal{B}(\mathcal{H}_\mathsf{S})$ into itself is a quantum operation if and only if it admits the representation*

$$\mathcal{M}(\rho) = \sum_i A_i \rho A_i^*\,, \tag{31}$$

where $\{A_i\}$ is a countable family of operators on $\mathcal{H}_\mathsf{S}$ satisfying $\sum_i A_i^* A_i = 1$.

For infinite dimensional Hilbert spaces and in the more general C*-algebra setting, the Kraus decomposition holds under the additional assumption that $\mathcal{M}$ is normal, that is, ultra-weakly continuous. One usually deduces it from Stinespring's theorem. In our finite-dimensional setting, however, a simple direct proof of Theorem 3.B.3 exists (see Remark 3.B.4 below). One can then obtain the

Stinespring theorem from the Kraus decomposition as follows. Let $\{|k\rangle\}_{k=1}^{n_S}$ be an orthonormal basis of $\mathcal{H}_S$ and $\mathcal{H}_E$ be a (possibly infinite-dimensional) Hilbert space with orthonormal basis $\{|\epsilon_i\rangle\}$. Define the vectors $|\Psi_{k0}\rangle = \sum_i A_i|k\rangle|\epsilon_i\rangle$. Using $\sum_i A_i^* A_i = 1$, one finds that these vectors form an orthonormal family in $\mathcal{H}_{SE}$, which can be completed so as to get an orthonormal basis $\{|\Psi_{kl}\rangle\}$. Then $\mathcal{M}^*(X) = \langle\epsilon_0|U^* X \otimes 1\, U|\epsilon_0\rangle$ for any $X \in \mathcal{B}(\mathcal{H}_S)$, where the unitary U on $\mathcal{H}_{SE}$ is defined by $U|k\rangle|\epsilon_l\rangle = |\Psi_{kl}\rangle$ for any k and l.

Remark 3.B.4. Any quantum operation $\mathcal{B}(\mathcal{H}_S) \to \mathcal{B}(\mathcal{H}_S)$ with $\dim \mathcal{H}_S = n_S < \infty$ admits a Kraus decomposition (31) with at most n_S^2 operators A_i. Consequently, one can choose the ancilla space $\mathcal{H}_E$ in Theorem 3.B.2 of dimension $\dim \mathcal{H}_E = n_S^2$.

Sketch the proof.[117] To show that $\mathcal{M}$ has the form (31), consider the operator $B = \mathcal{M} \otimes 1(|\Psi_{ent}\rangle\langle\Psi_{ent}|)$ with $|\Psi_{ent}\rangle = n_S^{-1/2} \sum_k |k\rangle|k\rangle \in \mathcal{H}_{SS}$ as above. Since $\mathcal{M}$ is CP, one has $B \geq 0$. Let $|\widetilde{\Phi}_i\rangle$ be orthogonal eigenvectors of B, normalized in such a way that $n_S B = \sum_i |\widetilde{\Phi}_i\rangle\langle\widetilde{\Phi}_i|$. Then define the Kraus operators A_i as the operators associated to $|\widetilde{\Phi}_i\rangle$ by the isomorphism (5) between $\mathcal{B}(\mathcal{H}_S)$ and $\mathcal{H}_{SS}$. □

It is important to realize that the Kraus decomposition is not unique. For indeed, if $\{A_i\}_{i=1}^p$ is a family of Kraus operators for $\mathcal{M}$ and $(u_{ji})_{i,j=1}^q$ is a unitary matrix of size $q \geq p$, then the operators

$$B_j = \sum_{i=1}^{p} \overline{u}_{ji} A_i, \quad j = 1, \ldots, q\ , \tag{32}$$

define another family of Kraus operators for $\mathcal{M}$. Conversely, two families $\{A_i\}_{i=1}^p$ and $\{B_j\}_{j=1}^q$ of Kraus operators for $\mathcal{M}$ with $p \leq q < \infty$ are related to each other by (32). Actually, let B and $|\Psi_{ent}\rangle$ be defined as in the Remark 3.B.4 above. Then $B = \sum_i |\widetilde{\mu}_i\rangle\langle\widetilde{\mu}_i| = \sum_j |\widetilde{\nu}_j\rangle\langle\widetilde{\nu}_j|$ with

$$|\widetilde{\mu}_i\rangle = n_S^{-\frac{1}{2}} \sum_k (A_i|k\rangle)|k\rangle, \quad |\widetilde{\nu}_j\rangle = n_S^{-\frac{1}{2}} \sum_k (B_j|k\rangle)|k\rangle. \tag{33}$$

In view of the link (16) between pure state decompositions of a non-negative operator, one has $|\widetilde{\nu}_j\rangle = \sum_i \overline{u}_{ji}|\widetilde{\mu}_i\rangle$ with $(u_{ji})_{i,j=1}^q$ unitary. This implies (32).

Given a purification $|\Psi\rangle$ of ρ on $\mathcal{H}_S \otimes \mathcal{H}_R$ and a quantum operation $\mathcal{M} : \mathcal{B}(\mathcal{H}_S) \to \mathcal{B}(\mathcal{H}'_S)$, it is natural to ask about purifications of $\mathcal{M}(\rho)$. A slight generalization of Theorem 3.B.2 ensures that there exist a vector $|\epsilon_0\rangle \in \mathcal{H}_E$ and a unitary $U : \mathcal{H}_S \otimes \mathcal{H}_E \to \mathcal{H}'_S \otimes \mathcal{H}'_E$ such that $\mathcal{M}(\rho) = \mathrm{tr}_{E'}(U\rho \otimes |\epsilon_0\rangle\langle\epsilon_0|U^*)$. Therefore,

$$|\Psi_{\mathcal{M}}\rangle = 1_R \otimes U|\Psi\rangle|\epsilon_0\rangle = \sum_{k=1}^{n} \sum_{i=1}^{p} \sqrt{p_k}(A_i|k\rangle)|f_k\rangle|\epsilon'_i\rangle \tag{34}$$

is a purification of $\mathcal{M}(\rho)$ on $\mathcal{H}'_S \otimes \mathcal{H}_R \otimes \mathcal{H}'_E$. In the second equality, $\{|k\rangle\}$ is an orthonormal eigenbasis of ρ, $\{|f_k\rangle\}$ is the orthonormal family of $\mathcal{H}_R$ such that $|\Psi\rangle = \sum_k \sqrt{p_k}|k\rangle|f_k\rangle$, and $\{|\epsilon'_i\rangle\}$ is an orthonormal basis of $\mathcal{H}'_E$ such that $U|k\rangle|\epsilon_0\rangle = \sum_i (A_i|k\rangle)|\epsilon'_i\rangle$ (see the expression of U in terms of the Kraus operators after Theorem 3.B.3).

C. Generalized measurements

For the quantum operation $\mathcal{M}_\Pi$ defined by (26), the orthogonal projectors Π_i form a family of Kraus operators. One may wonder if more general quantum operations, given by Kraus operators A_i which are not necessarily orthogonal projectors, correspond to some kind of measurements. The answer is yes: such operations can always be obtained by coupling the system S to an auxiliary system E (the ancilla) and subsequently performing a von Neumann measurement on E.

Theorem 3.C.1. (Neumark extension theorem) *Let $\{A_i\}_{i=1}^p$ be a family of operators such that $\sum_i A_i^* A_i = 1$. Then there exist a space $\mathcal{H}_E$ with dimension $\dim \mathcal{H}_E = p$, a pure state $|\epsilon_0\rangle \in \mathcal{H}_E$, an orthonormal family $\{\pi_i^E\}$ of projectors in $\mathcal{B}(\mathcal{H}_E)$, and a unitary operator U on $\mathcal{H}_{SE}$ such that for*

any density matrix $\rho \in \mathcal{E}(\mathcal{H}_{\mathsf{S}})$,

$$A_i \rho A_i^* = \mathrm{tr}_{\mathsf{E}}(1 \otimes \pi_i^{\mathsf{E}} U \rho \otimes |\epsilon_0\rangle\langle\epsilon_0| U^* 1 \otimes \pi_i^{\mathsf{E}}) \; . \tag{35}$$

Proof. Use the same arguments as in the above proof of Stinespring's theorem from Theorem 3.B.3, and define $\pi_i^{\mathsf{E}} = |\epsilon_i\rangle\langle\epsilon_i|$. □

Definition 3.C.2. *A generalized measurement is given by a family* $\{M_i\}$ *of non-negative operators* M_i *satisfying* $\sum_i M_i = 1$ *(positive operator valued measure, abbreviated as POVM) together with a family of operators* $\{A_i\}$ *such that* $M_i = A_i^* A_i$. *The conditional state* $\rho_{S|i}$ *given outcome* i *and the probability of this outcome read*

$$\rho_{S|i} = p_i^{-1} A_i \rho A_i^*, \quad p_i = \mathrm{tr}(M_i \rho) \; . \tag{36}$$

According to Theorem 3.C.1, any generalized measurement can be realized by letting the system S interact with an ancilla E in the state $|\epsilon_0\rangle$ and subsequently performing a von Neumann measurement on E, that is, coupling E to a macroscopic apparatus with pointer P. The interaction between S and E first transforms the initial state $\rho_{\mathsf{S}} \otimes |\epsilon_0\rangle\langle\epsilon_0|$ into $\rho_{\mathsf{SE}} = U \rho_{\mathsf{S}} \otimes |\epsilon_0\rangle\langle\epsilon_0| U^*$, U being a unitary evolution operator on $\mathcal{H}_{\mathsf{SE}}$, and the subsequent von Neumann measurement leads to the wavepacket reduction for the system SP (compare with (24) and (25))

$$\textbf{outcome } i\text{:} \qquad \rho_{\mathsf{SP}} \to \rho_{\mathsf{SP}|i} = p_i^{-1} \mathrm{tr}_{\mathsf{E}}(1 \otimes \pi_i^{\mathsf{E}} \rho_{\mathsf{SE}} 1 \otimes \pi_i^{\mathsf{E}}) \otimes |i\rangle\langle i| = p_i^{-1} A_i \rho_{\mathsf{S}} A_i^* \otimes |i\rangle\langle i| \; , \tag{37}$$

where $p_i = \mathrm{tr}(1 \otimes \pi_i^{\mathsf{E}} \rho_{\mathsf{SE}}) = \mathrm{tr}(M_i \rho_{\mathsf{S}})$ is the probability of outcome i, in agreement with (36).

One has $A_i = U_i M_i^{1/2}$ (polar decomposition) for some unitary operator U_i depending on i. The conditional states $\rho_{\mathsf{S}|i}$ are thus characterized by the POVM $\{M_i\}$ up to unitary conjugations, which introduce a freedom in choosing the output state associated to each measurement outcome. For instance, if $M_i = |\widetilde{\mu}_i\rangle\langle\widetilde{\mu}_i|$ are of rank one then $A_i = |i\rangle\langle\widetilde{\mu}_i|$ for some arbitrary normalized vector $|i\rangle$ and the output conditional states are $\rho_{\mathsf{S}|i} = |i\rangle\langle i|$. One usually takes the vectors $|i\rangle$ to form an orthonormal basis (which can be identified to the pointer state basis of Sec. III A), in such a way that the states $\rho_{\mathsf{S}|i}$ be perfectly distinguishable (this happens if the $\rho_{\mathsf{S}|i}$ are orthogonal only, see Sec. V below). One should keep in mind, however, that the probability $p_i = \langle\widetilde{\mu}_i|\rho|\widetilde{\mu}_i\rangle$ of outcome i is independent of the choice of $\{|i\rangle\}$. If one is interested only in functions of the post-measurement states $\rho_{\mathsf{S}|i}$ which are invariant under unitary conjugations (as, for instance, the von Neumann entropy), then the generalized measurement can be fully specified by the measurement operators M_i. Thanks to the Neumark extension theorem, these operators may be written as

$$M_i = A_i^* A_i = \langle\epsilon_0| U^* 1 \otimes \pi_i^{\mathsf{E}} U |\epsilon_0\rangle. \tag{38}$$

As stressed above, in the absence of read-out the state of the system after the measurement is the average of the conditional states,

$$\mathcal{M}(\rho) = \sum_i p_i \rho_{S|i} = \sum_i A_i \rho A_i^* \; , \tag{39}$$

in analogy with (26). This defines a quantum operation $\mathcal{M}$, the Kraus decomposition of which specifies the state preparation associated with the wavepacket reduction.

Writing the spectral decomposition of each operator M_i, one observes that

$$M_i = \sum_{k=1}^{r_i} |\widetilde{\mu}_{ik}\rangle\langle\widetilde{\mu}_{ik}|, \quad \sum_i M_i = \sum_{i,k} |\widetilde{\mu}_{ik}\rangle\langle\widetilde{\mu}_{ik}| = 1, \tag{40}$$

where $r_i = \mathrm{rank}(M_i)$ and $|\widetilde{\mu}_{ik}\rangle$ are unnormalized eigenvectors with norms equal to the square roots of the corresponding eigenvalues. The last condition in (40) implies that either $\{|\widetilde{\mu}_{ik}\rangle\}$ is an orthonormal basis, in which case $\{M_i\}$ is an orthonormal family of projectors (von Neumann measurement), or $\{|\widetilde{\mu}_{ik}\rangle\}$ is a non-orthogonal family containing more than n_{S} vectors, in which case at least two eigenvalues $\|\widetilde{\mu}_{ik}\|$ are strictly smaller than one and $\{M_i\}$ is not a von Neumann measurement.

The set of all POVMs is a convex set. Its boundary and extremal points have been studied in Ref. 45.

Remark 3.C.3. An alternative version of Theorem 3.C.1 states that if $m = \sum_i r_i$ *with* $r_i = \mathrm{rank}(M_i)$, *then there exist a space* $\mathcal{H}_\mathsf{E}$ *with dimension* $m - n_\mathsf{S} + 1$, *a state* $|\epsilon_0\rangle \in \mathcal{H}_\mathsf{E}$, *and a von Neumann measurement* $\{\Pi_i^{\mathsf{SE}}\}$ *on* $\mathcal{H}_{\mathsf{SE}}$ *such that*

$$M_i = \langle \epsilon_0 | \Pi_i^{\mathsf{SE}} | \epsilon_0 \rangle. \tag{41}$$

The interesting point is that the dimension of the ancilla space $\mathcal{H}_\mathsf{E}$ *can be smaller than* p *in Theorem 3.C.1 (for instance,* $\dim \mathcal{H}_\mathsf{E} = p - n_\mathsf{S} + 1$ *for rank-one operators* M_i*).*

Sketch of the proof.[126] Note that $m \geq n_\mathsf{S}$ by the observation above. Define

$$|\zeta_{ik}\rangle = |\widetilde{\mu}_{ik}\rangle |\epsilon_0\rangle + \sum_{l=1}^{m-n_\mathsf{S}} c_{ik,l} |\phi\rangle |\epsilon_l\rangle \ , \tag{42}$$

where $|\widetilde{\mu}_{ik}\rangle$ is as in (40), $|\phi\rangle \in \mathcal{H}_\mathsf{S}$ is an arbitrary state, and $\{|\epsilon_l\rangle\}_{l=0}^{m-n_\mathsf{S}}$ is an orthonormal basis of $\mathcal{H}_\mathsf{E}$. The coefficients $c_{ik,\,l}$ may be chosen such that $\{|\zeta_{ik}\rangle\}$ is an orthonormal family of $\mathcal{H}_{\mathsf{SE}}$. To establish this statement, set $c_{ik,l} = \langle l | \widetilde{\mu}_{ik} \rangle$ for $m - n_\mathsf{S} < l \leq m$, with $\{|l\rangle\}_{l=m-n_\mathsf{S}+1}^{m}$ an orthonormal basis of $\mathcal{H}_\mathsf{S}$, and let $\mathbf{c}_l \in \mathbb{C}^m$ be the vector with components $c_{ik,\,l}$. Then $\mathbf{c}_l \cdot \mathbf{c}_{l'} = \delta_{ll'}$ for any $l, l' > m - n_\mathsf{S}$, as a result of $\sum_i M_i = 1$. One can choose the $(m - n_\mathsf{S})$ other vectors $\mathbf{c}_l$ in such a way that $(\mathbf{c}_1, \ldots, \mathbf{c}_m)$ forms a $m \times m$ unitary matrix. Then $\Pi_i^{\mathsf{SE}} = \sum_k |\zeta_{ik}\rangle\langle\zeta_{ik}|$ has the desired property. □

D. Connections between POVMs, quantum operations, and state ensembles

To each POVM one can associate a quantum operation and vice-versa. Similarly, there is a canonical way to associate to a quantum operation a state ensemble and vice versa. These correspondences depend on an orthonormal basis $\{|i\rangle\}_{i=1}^m$ of a fictitious pointer P with m-dimensional space $\mathcal{H}_\mathsf{P}$. It has been already seen above that one can associate to a POVM $\{M_i\}_{i=1}^m$ on S a quantum operation with Kraus operators A_i such that $M_i = A_i^* A_i$. This operation implements the state changes in the measurement process in the absence of readout. If we imagine that S is coupled to P and that the measurement is performed on both S and P, one may consider the Kraus operators $A_{ik} = |k\rangle\langle i|\widetilde{\mu}_{ik}|$ such that $M_i = \sum_k A_{ik}^* A_{ik}$, where $\{|k\rangle\}_{k=1}^{n_\mathsf{S}}$ is an orthonormal basis of $\mathcal{H}_\mathsf{S}$ and $|\widetilde{\mu}_{ik}\rangle$ are the unnormalized eigenvectors of M_i in (40). Provided that there is no readout of the measurement on S, one may trace the post-measurement states over $\mathcal{H}_\mathsf{S}$. The conditional states of P are given by $\rho_{\mathsf{P}|i} = p_{ik}^{-1} \mathrm{tr}_\mathsf{S}(A_{ik} \rho A_{ik}^*) = |i\rangle\langle i|$ with $p_{ik} = \langle \widetilde{\mu}_{ik} | \rho | \widetilde{\mu}_{ik} \rangle$, and the corresponding probability is $p_i = \sum_k p_{ik} = \mathrm{tr}(M_i \rho)$. The state changes in the absence of readout are implemented by the quantum operation $\mathcal{M} : \mathcal{B}(\mathcal{H}_\mathsf{S}) \to \mathcal{B}(\mathcal{H}_\mathsf{P})$ defined by

$$\mathcal{M}(\rho) = \sum_i \mathrm{tr}(M_i \rho) |i\rangle\langle i|, \quad \rho \in \mathcal{E}(\mathcal{H}_\mathsf{S}) \quad \Leftrightarrow \quad \mathcal{M}^*(|i\rangle\langle j|) = M_i \delta_{ij}, \quad i, j = 1, \ldots, m \ . \tag{43}$$

Conversely, if $\mathcal{M}$ is a quantum operation $\mathcal{B}(\mathcal{H}_\mathsf{S}) \to \mathcal{B}(\mathcal{H}_\mathsf{P})$ then $M_i = \mathcal{M}^*(|i\rangle\langle i|)$ defines a POVM $\{M_i\}_{i=1}^m$ (actually, $M_i \geq 0$ by the positivity of $\mathcal{M}^*$ and $\sum_i M_i = \mathcal{M}^*(1) = 1$). Therefore, for a given orthonormal basis $\{|i\rangle\}_{i=1}^m$ of $\mathcal{H}_\mathsf{P}$, there is a one-to-one correspondence between POVMs $\{M_i\}_{i=1}^m$ on $\mathcal{H}_\mathsf{S}$ and quantum operations $\mathcal{M} : \mathcal{B}(\mathcal{H}_\mathsf{S}) \to \mathcal{B}(\mathcal{H}_\mathsf{P})$ of the form (43).

A similar one-to-one correspondence can be found between state ensembles on $\mathcal{H}_\mathsf{S}$ with fixed probabilities $\{\eta_i\}_{i=1}^m$ and quantum operations $\mathcal{B}(\mathcal{H}_\mathsf{P}) \to \mathcal{B}(\mathcal{H}_\mathsf{S})$ such that $\mathcal{M}(|i\rangle\langle j|) = 0$ for $i \neq j$. This correspondence is given by

$$\rho_i = \mathcal{M}(|i\rangle\langle i|), \quad i = 1, \ldots, m \ . \tag{44}$$

In fact, if $\mathcal{M} : \mathcal{B}(\mathcal{H}_\mathsf{P}) \to \mathcal{B}(\mathcal{H}_\mathsf{S})$ is a quantum operation then $\{\rho_i, \eta_i\}_{i=1}^m$ is clearly an ensemble on $\mathcal{H}_\mathsf{S}$. Conversely, if $\{\rho_i, \eta_i\}_{i=1}^m$ is an ensemble of m states, let us write the spectral decompositions $\rho_i = \sum_k p_{ik} |\psi_{ik}\rangle\langle\psi_{ik}|$. Then the operation with Kraus operators $A_{ik} = \sqrt{p_{ik}} |\psi_{ik}\rangle\langle i|$ has the required property.

IV. TRANSPOSE OPERATION AND LEAST SQUARE MEASUREMENT

A. Recovery operation in quantum error correction

The notion of transpose operation was introduced by Ohya and Petz in their monograph.[119] It plays the role of an approximate reversal of a quantum operation, in a sense that will be made more precise below.

Definition 4.A.1. *Let $\mathcal{M} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ be a quantum operation and $\rho \in \mathcal{E}(\mathcal{H})$ be a state such that $\mathcal{M}(\rho) > 0$. The transpose operation of $\mathcal{M}$ for ρ is the quantum operation $\mathcal{R}_{\mathcal{M},\rho} : \mathcal{B}(\mathcal{H}') \to \mathcal{B}(\mathcal{H})$ with Kraus operators $R_i = \rho^{\frac{1}{2}} A_i^* \mathcal{M}(\rho)^{-\frac{1}{2}}$, where $\{A_i\}$ is a family of Kraus operators for $\mathcal{M}$. It is independent of the Kraus decomposition of $\mathcal{M}$. Actually, for any $\sigma \in \mathcal{E}(\mathcal{H}')$,*

$$\mathcal{R}_{\mathcal{M},\rho}(\sigma) = \rho^{\frac{1}{2}} \mathcal{M}^*\big(\mathcal{M}(\rho)^{-\frac{1}{2}} \sigma \mathcal{M}(\rho)^{-\frac{1}{2}}\big) \rho^{\frac{1}{2}} \ . \tag{45}$$

One easily checks that $\sum_i R_i^* R_i = 1$, so that $\mathcal{R}_{\mathcal{M},\rho}$ is indeed a quantum operation, and that $\mathcal{R}_{\mathcal{M},\rho} \circ \mathcal{M}(\rho) = \rho$. Furthermore, transposing twice amounts to do nothing, that is, the transpose of $\mathcal{R}_{\mathcal{M},\rho}$ for the state $\mathcal{M}(\rho)$ is equal to $\mathcal{M}$.

The operation $\mathcal{R}_{\mathcal{M},\rho}$ appears naturally in the context of quantum error correction. The problem of quantum error correction is to send a state ρ over a noisy quantum communication channel in such a way that ρ is resilient to the effect of the noise in the channel. The state ρ is encoded via a unitary transformation into a subspace $\mathcal{H}_C$ of the Hilbert space $\mathcal{H}$ of the quantum channel. The noise is described by some quantum operation $\mathcal{M}$.

Proposition 4.A.2. *Let $\mathcal{M}$ be a quantum operation on $\mathcal{B}(\mathcal{H})$ with Kraus operators $\{A_i\}$. Let Π_C denote the orthogonal projector onto a subspace $\mathcal{H}_C \subset \mathcal{H}$ and $\mathbb{E}_C : \rho \mapsto \Pi_C \, \rho \, \Pi_C$ be the conditional expectation onto the space of operators supported on $\mathcal{H}_C$. There exists a recovery quantum operation $\mathcal{R}$ on $\mathcal{B}(\mathcal{H})$ satisfying $\mathcal{R} \circ \mathcal{M} \circ \mathbb{E}_C = \mathbb{E}_C$ if and only if the following condition holds:*

$$\mathbb{E}_C(A_i^* A_j) = a_{ij} \Pi_C, \tag{46}$$

where (a_{ij}) is a self-adjoint matrix. If this condition is satisfied then for any ρ with support $\mathrm{ran}(\rho) \subset \mathcal{H}_C$, *the transpose operation $\mathcal{R}_{\mathcal{M},\rho}$ is a recovery quantum operation.*

We refer the reader to the book of Nielsen and Chuang[117] for a proof of the necessary and sufficient condition (46). Some bibliographic information on this topic can also be found there.

Proof of the second statement. By taking advantage of the non-uniqueness of the Kraus decomposition, (46) can be transformed into $\mathbb{E}_C(B_i^* B_j) = p_i \delta_{ij} \Pi_C$, where the Kraus operators B_i are given by (32) with $(u_{ij})(a_{ij})(u_{ij})^*$ the diagonal matrix with entries p_i. Together with the polar decomposition, this implies $B_j \Pi_C = \sqrt{p_j} W_j$ with $W_j = V_j \Pi_C$ satisfying $W_i^* W_j = \delta_{ij} \Pi_C$, the V_j being some unitary operators. Thus the subspaces $V_j \mathcal{H}_C$ are orthogonal for different j's and the restriction of $\sum_j W_j W_j^*$ to the subspace $\mathcal{V} = \oplus_j V_j \mathcal{H}_C$ equals the identity. If $\rho = \mathbb{E}_C(\rho)$ and the restriction of ρ to $\mathcal{H}_C$ is invertible, then $\mathcal{M}(\rho) = \sum_j p_j W_j \rho W_j^*$ and $\mathcal{M}(\rho)^{-1/2} = \sum_j W_j \rho^{-1/2} W_j^* / \sqrt{p_j}$, the last operator being defined on $\mathcal{V}$. A simple calculation then shows that $\mathcal{R}_{\mathcal{M},\rho} \circ \mathcal{M} \circ \mathbb{E}_C = \mathbb{E}_C$, as stated in the proposition. □

B. Transpose operation as an approximate reverse operation

Since the condition (46) is not always fulfilled, it is natural to ask whether one can find an optimal imperfect recovery map, which would enable to recover a given ensemble $\{\rho_i, \eta_i\}$ subject to some noise with a maximal fidelity. A notion of fidelity has been introduced by Schumacher.[141] Its definition is as follows (for more detail and motivations from classical information theory, see Ref. 117). Given a state $\rho \in \mathcal{E}(\mathcal{H}_\mathsf{S})$, consider a purification $|\Psi_\rho\rangle$ of ρ on $\mathcal{H}_\mathsf{S} \otimes \mathcal{H}_\mathsf{R}$, where R is a reference system with Hilbert space $\mathcal{H}_\mathsf{R} \simeq \mathcal{H}_\mathsf{S}$. For instance, $|\Psi_\rho\rangle$ can be given by (12). If ρ is a mixed state then $|\Psi_\rho\rangle$ is SR-entangled (Sec. II D). The *entanglement fidelity* of ρ quantifies how

well this entanglement is preserved when the system S is subject to some noise modelized by a quantum operation $\mathcal{M}$ on $\mathcal{B}(\mathcal{H}_\mathsf{S})$. It is defined by

$$F_\mathrm{e}(\rho, \mathcal{M}) = \langle \Psi_\rho | \mathcal{M} \otimes 1(|\Psi_\rho\rangle\langle\Psi_\rho|)|\Psi_\rho\rangle \ . \tag{47}$$

Since different purifications of ρ on $\mathcal{H}_\mathsf{SR}$ are related by unitaries acting on $\mathcal{H}_\mathsf{R}$, the right-hand side of (47) does not depend on the chosen purification. As a consequence of the positivity and the trace-preserving property of $\mathcal{M}$, one has $0 \le F_\mathrm{e}(\rho, \mathcal{M}) \le \mathrm{tr}_\mathsf{SR}[\mathcal{M} \otimes 1(|\Psi_\rho\rangle\langle\Psi_\rho|)] = \mathrm{tr}[\mathcal{M}(\rho)] = 1$. Plugging (12) and (31) into (47), a simple calculation yields

$$F_\mathrm{e}(\rho, \mathcal{M}) = \sum_j \left|\mathrm{tr}(A_j \rho)\right|^2 \ , \tag{48}$$

where $\{A_j\}$ is a family of Kraus operators for $\mathcal{M}$. Note that the sum in the right-hand side does not depend on the choice of Kraus decomposition (this follows from (32)), as it should be. For a pure state $\rho_\psi = |\psi\rangle\langle\psi|$, the entanglement fidelity reduces to the input-output fidelity $F(\rho_\psi, \mathcal{M}) = \langle\psi|\mathcal{M}(|\psi\rangle\langle\psi|)|\psi\rangle$. One infers from (48) that $F_\mathrm{e}(\rho, \mathcal{M})$ is a convex function of ρ.

Let us now consider an ensemble of states $\{\rho_i, \eta_i\}_{i=1}^m$. The corresponding average entanglement fidelity is defined by

$$\overline{F}_\mathrm{e}(\{\rho_i, \eta_i\}, \mathcal{M}) = \sum_i \eta_i F_\mathrm{e}(\rho_i, \mathcal{M}) \ . \tag{49}$$

This fidelity belongs to the interval [0, 1].

Proposition 4.B.1. (Barnum and Knill[19]) *If the states ρ_i commute with $\rho = \sum_i \eta_i \rho_i$, then*

$$\overline{F}_\mathrm{e}\big(\{\rho_i, \eta_i\}, \mathcal{R}_{\mathcal{M},\rho} \circ \mathcal{M}\big) \ge \overline{F}_\mathrm{e}\big(\{\rho_i, \eta_i\}, \mathcal{R}_\mathrm{opt} \circ \mathcal{M}\big)^2, \tag{50}$$

where $\mathcal{R}_{\mathcal{M},\rho}$ is the transpose operation of $\mathcal{M}$ for ρ and $\mathcal{R}_\mathrm{opt}$ the optimal recovery quantum operation $\mathcal{R}$ maximizing $\overline{F}_e(\{\rho_i, \eta_i\}, \mathcal{R} \circ \mathcal{M})$.

Hence, if the minimal fidelity error is $1 - \overline{F}_e(\{\rho_i, \eta_i\}, \mathcal{R}_\mathrm{opt} \circ \mathcal{M}) = \eta$, then the fidelity error by using $\mathcal{R}_{\rho,\mathcal{M}}$ as the recovery operation is at most twice larger than this minimal error.

Proof. Taking advantage of the non-uniqueness of the Kraus decomposition, one can choose for any fixed i some families $\{R_j^{\mathrm{opt}\,(i)}\}$ and $\{A_k^{(i)}\}$ of Kraus operators for $\mathcal{R}^\mathrm{opt}$ and $\mathcal{M}$ satisfying

$$\mathrm{tr}\big(R_j^{\mathrm{opt}\,(i)} A_k^{(i)} \rho_i\big) = 0, \quad j \ne k. \tag{51}$$

Actually, given any families $\{R_m^\mathrm{opt}\}$ for $\mathcal{R}^\mathrm{opt}$ and $\{A_l\}$ for $\mathcal{M}$, the operators $R_j^{\mathrm{opt}\,(i)} = \sum_m u_{jm}^{(i)} R_m^\mathrm{opt}$ and $A_k^{(i)} = \sum_l \overline{v}_{kl}^{(i)} A_l$ have the required property if $(u_{jm}^{(i)})$ and $(v_{kl}^{(i)})$ are the unitary matrices in the singular decomposition of $(\mathrm{tr}(R_m^\mathrm{opt} A_l \rho_i))$. Since $\{R_j^{\mathrm{opt}\,(i)} A_k^{(i)}\}$ is a Kraus family for $\mathcal{R}^\mathrm{opt} \circ \mathcal{M}$, one obtains from (48), (49), and (51)

$$\overline{F}_\mathrm{e}\big(\{\rho_i, \eta_i\}, \mathcal{R}^\mathrm{opt} \circ \mathcal{M}\big) = \sum_{i,j} \eta_i \left|\mathrm{tr}(R_j^{\mathrm{opt}(i)} A_j^{(i)} \rho_i)\right|^2 \ . \tag{52}$$

We first consider the case $\rho_\mathcal{M} = \mathcal{M}(\rho) > 0$. Without loss of generality, we may assume that $\mathrm{ran}(R_j^{\mathrm{opt}(i)}) \subset \mathrm{ran}\,\rho_i \subset \mathrm{ran}\,\rho$, so that the operators

$$X_{ij} = \eta_i^{\frac14} \rho_\mathcal{M}^{-\frac14} A_j^{(i)} \rho^{\frac14} \rho_i^{\frac12}, \quad Y_{ij} = \eta_i^{\frac14} \rho_\mathcal{M}^{-\frac14} B_j^{(i)} \rho^{\frac14} \rho_i^{\frac12} \quad \text{and} \quad (B_j^{(i)})^* = \rho^{-\frac12} R_j^{\mathrm{opt}\,(i)} \rho_\mathcal{M}^{\frac12} \tag{53}$$

are well-defined. Since $[\rho_i, \rho] = 0$, one finds by using twice the Cauchy-Schwarz inequality

$$\begin{aligned} \overline{F}_\mathrm{e}(\{\rho_i, \eta_i\}, \mathcal{R}^\mathrm{opt} \circ \mathcal{M})^2 &= \left(\sum_{i,j} \left|\mathrm{tr}(Y_{ij}^* X_{ij})\right|^2\right)^2 \le \sum_{i,j} \big(\mathrm{tr}(Y_{ij}^* Y_{ij})\big)^2 \sum_{i,j} \big(\mathrm{tr}(X_{ij}^* X_{ij})\big)^2 \\ &\le \sum_{i,j,k} \left|\mathrm{tr}(Y_{ij}^* Y_{ik})\right|^2 \sum_{i,j,k} \left|\mathrm{tr}(X_{ij}^* X_{ik})\right|^2 \ . \end{aligned} \tag{54}$$

The transpose operation $\mathcal{R}_{\rho,\mathcal{M}}$ has Kraus operators $R_j^{(i)} = \rho^{\frac{1}{2}}(A_j^{(i)})^* \rho_{\mathcal{M}}^{-\frac{1}{2}}$. As a result,

$$\overline{F}_{\mathrm{e}}\big(\{\rho_i, \eta_i\}, \mathcal{R}_{\rho,\mathcal{M}} \circ \mathcal{M}\big) = \sum_{i,j,k} \eta_i \big|\mathrm{tr}(R_j^{(i)} A_k^{(i)} \rho_i)\big|^2 = \sum_{i,j,k} \big|\mathrm{tr}(X_{ij}^* X_{ik})\big|^2 . \tag{55}$$

The first sum in the last member of (54) is equal to $\overline{F}_{\mathrm{e}}(\{\rho_i, \eta_i\}, \mathcal{R}^{\mathrm{opt}} \circ \mathcal{B})$, where $\mathcal{B}$ is the CP map defined by $\mathcal{B}(\sigma) = \sum_k B_k^{(i)} \sigma (B_k^{(i)})^*$ (note that $\mathcal{B}$ does not depend on i). Even if $\mathcal{B}$ is not trace-preserving, with the help of (47) this fidelity can be bounded from above by $\mathrm{tr}[\mathcal{R}^{\mathrm{opt}} \circ \mathcal{B}(\rho)]$, which equals unity thanks to the identity $\mathcal{B}(\rho) = \mathcal{M}(\rho)$. This yields the inequality (50). If $\rho_{\mathcal{M}}$ is not invertible, one approximates $\mathcal{M}$ by some quantum operations $\mathcal{M}_\varepsilon$ satisfying $\mathcal{M}_\varepsilon(\rho) > 0$ for $\varepsilon > 0$ and $\mathcal{M}_\varepsilon \to \mathcal{M}$ as $\varepsilon \to 0$, and obtains the result by continuity. □

C. Least square measurement

Let us consider an ensemble $\{\rho_i, \eta_i\}_{i=1}^m$ of states of the system S forming a convex decomposition of $\rho_{\mathrm{out}} = \sum_i \eta_i \rho_i$. For any i, we denote by $\rho_i = \sum_k p_{ik} |\psi_{ik}\rangle\langle\psi_{ik}|$ the spectral decomposition of ρ_i and set $\rho_i = A_i A_i^*$, where $A_i = \sqrt{\rho_i} U_i$ is defined up to a unitary U_i. Introducing as in Sec. III D an arbitrary orthonormal basis $\{|k\rangle\}_{k=1}^{n_{\mathsf{S}}}$ of $\mathcal{H}_{\mathsf{S}}$ and a fictitious pointer with m-dimensional space $\mathcal{H}_{\mathsf{P}}$ and orthonormal basis $\{|i\rangle\}_{i=1}^m$, one can choose

$$A_i = \sum_{k=1}^{n_{\mathsf{S}}} \sqrt{p_{ik}} |\psi_{ik}\rangle\langle k| \langle i| \ \in \ \mathcal{B}(\mathcal{H}_{\mathsf{SP}}, \mathcal{H}_{\mathsf{S}}). \tag{56}$$

We remark that A_i is associated to a purification of $\rho_i \otimes |i\rangle\langle i|$ on $\mathcal{H}_{\mathsf{SP}} \otimes \mathcal{H}_{\mathsf{S}}$ via the isometry (5) between $\mathcal{B}(\mathcal{H}_{\mathsf{SP}}, \mathcal{H}_{\mathsf{S}})$ and $\mathcal{H}_{\mathsf{SP}} \otimes \mathcal{H}_{\mathsf{S}}$, namely, $|\Psi_i\rangle = \sum_k \sqrt{p_{ik}} |\psi_{ik}\rangle |i\rangle |k\rangle$. Moreover, $|\Psi_{\mathrm{out}}\rangle = \sum_i \sqrt{\eta_i} |\Psi_i\rangle$ is a purification of ρ_{out} on the same space.

The *least square measurement* associated to $\{\rho_i, \eta_i\}_{i=1}^m$ is given by the Kraus and measurement operators

$$R_i^{\mathrm{lsm}} = \sqrt{\eta_i} A_i^* \rho_{\mathrm{out}}^{-\frac{1}{2}} = \sum_k \sqrt{\eta_i p_{ik}} |k\rangle|i\rangle\langle\psi_{ik}| \rho_{\mathrm{out}}^{-\frac{1}{2}}, \quad M_i^{\mathrm{lsm}} = \big|R_i^{\mathrm{lsm}}\big|^2 = \eta_i \rho_{\mathrm{out}}^{-\frac{1}{2}} \rho_i \rho_{\mathrm{out}}^{-\frac{1}{2}} \tag{57}$$

for $i = 1, \ldots, m$. One indeed checks that $\sum_i M_i^{\mathrm{lsm}} = 1$, so that (57) defines a generalized measurement in the sense of Definition 3.C.2. This measurement bears several names: it was referred to as the "pretty good measurement" in Ref. 71 and is also called "square-root measurement" by many authors. While the operators M_i^{lsm} and thus the outcome probabilities $q_i = \mathrm{tr}(M_i^{\mathrm{lsm}} \sigma_{\mathsf{S}})$ (here σ_{S} is the system state) only depend on $\{\rho_i, \eta_i\}$, the post-measurement states also depend on the choice of the basis $\{|i\rangle\}$, as highlighted in Sec. III. The conditional and average post-measurement states of the pointer P are

$$\textbf{outcome } i\text{:} \ \ \sigma_{\mathsf{S}} \ \mapsto \ \sigma_{\mathsf{P}|i} = q_i^{-1} \, \mathrm{tr}_{\mathsf{S}}(R_i^{\mathrm{lsm}} \sigma_{\mathsf{S}} (R_i^{\mathrm{lsm}})^*) = |i\rangle\langle i| \tag{58}$$

$$\textbf{no readout}\text{:} \ \ \sigma_{\mathsf{S}} \ \mapsto \ \sigma_{\mathsf{P}} = \mathcal{M}^{\mathrm{lsm}}(\sigma_{\mathsf{S}}) = \sum_{i=1}^m q_i \sigma_{\mathsf{P}|i} = \sum_{i=1}^m q_i |i\rangle\langle i| \, . \tag{59}$$

For a pure state ensemble $\{|\psi_i\rangle, \eta_i\}_{i=1}^m$, the least square measurement consists of rank-one measurement operators $M_i = |\widetilde{\mu}_i\rangle\langle\widetilde{\mu}_i|$ with $|\widetilde{\mu}_i\rangle = \sqrt{\eta_i} \rho_{\mathrm{out}}^{-\frac{1}{2}} |\psi_i\rangle$. The vectors $|\widetilde{\mu}_i\rangle$ have the following property,[78,54] which elucidates the name given to the measurement: they minimize the sum of the square norms $\| |\widetilde{\mu}_i\rangle - \sqrt{\eta_i} |\psi_i\rangle \|^2$ under the constraint $\sum_i |\widetilde{\mu}_i\rangle\langle\widetilde{\mu}_i| = 1$. If the $|\psi_i\rangle$ are linearly independent and span $\mathcal{H}_{\mathsf{S}}$, so that $m = n$, then $\{|\widetilde{\mu}_i\rangle\}$ is an orthonormal basis of $\mathcal{H}_{\mathsf{S}}$. In that case $\{M_i^{\mathrm{lsm}}\}$ is a von Neumann measurement (see Sec. III C).

Remark 4.C.1. The aforementioned property of a least square measurement can be stated as follows:

$$\min_{\{|\widetilde{\mu}_i\rangle\}}\left\{\sum_{i=1}^{m}\left\||\widetilde{\mu}_i\rangle-\sqrt{\eta_i}|\psi_i\rangle\right\|^2\right\}=n_{\mathsf{S}}+1-2\,\mathrm{tr}(\rho_{\rm out}^{\frac{1}{2}})\quad with\quad \rho_{\rm out}=\sum_i \eta_i|\psi_i\rangle\langle\psi_i|\,, \tag{60}$$

the minimum being over all families $\{|\widetilde{\mu}_i\rangle\}_{i=1}^m$ *in* $\mathcal{H}_{\mathsf{S}}$ *such that* $\sum_i |\widetilde{\mu}_i\rangle\langle\widetilde{\mu}_i|=1$. *This minimum is achieved if and only if* $|\widetilde{\mu}_i\rangle=\sqrt{\eta_i}\rho_{\rm out}^{-1/2}|\psi_i\rangle$ *(up to irrelevant phase factors).*

Sketch of the proof.[54,83] Define $A=\sum_i\sqrt{\eta_i}|\psi_i\rangle\langle i|$ and $B=\sum_i|\widetilde{\mu}_i\rangle\langle i|$ in analogy with (56). Then observe that the sum to be minimized in (60) is equal to $\|A^*-B^*\|_2^2=1+n_{\mathsf{S}}-2\mathrm{Re}\ \mathrm{tr}(AB^*)$, and use (3). □

As suggested by this result, the least square measurement plays an important role in distinguishing quantum states drawn from a given ensemble. This point will be discussed in Sec. V C below.

Let us recall from Sec. III D that the relation $\rho_i=\mathcal{M}(|i\rangle\langle i|)$, where $\{|i\rangle\}_{i=1}^m$ is a fixed orthonormal basis of $\mathcal{H}_{\mathsf{P}}$, can be used to associate to a quantum operation $\mathcal{M}:\mathcal{B}(\mathcal{H}_{\mathsf{P}})\to\mathcal{B}(\mathcal{H}_{\mathsf{S}})$ an ensemble $\{\rho_i,\eta_i\}_{i=1}^m$ on $\mathcal{H}_{\mathsf{S}}$. Conversely, if $\{\rho_i,\eta_i\}$ is an ensemble on $\mathcal{H}_{\mathsf{S}}$, the operation $\mathcal{M}$ with Kraus operators $A_{ik}=A_i|k\rangle=\sqrt{p_{ik}}|\Psi_{ik}\rangle\langle i|$ satisfies this relation (here A_i is the operator (56)). Similarly, the relation (43) establishes a one-to-one correspondence between POVMs $\{M_i\}$ on $\mathcal{H}_{\mathsf{S}}$ and quantum operations $\mathcal{R}:\mathcal{B}(\mathcal{H}_{\mathsf{S}})\to\mathcal{B}(\mathcal{H}_{\mathsf{P}})$. It was recognized by Barnum and Knill[19] that *the least square measurement associated to the ensemble* $\{\rho_i=\mathcal{M}(|i\rangle\langle i|),\eta_i\}$ *is nothing but the measurement corresponding to the transpose operation* $\mathcal{R}_{\mathcal{M},\rho_{\rm in}}$ *of* $\mathcal{M}$ *for the state* $\rho_{in}=\sum_i\eta_i|i\rangle\langle i|$. Actually, since $\mathcal{M}(\rho_{\rm in})=\rho_{\rm out}$, according to the Definition 4.A.1,

$$R_{ik}=\rho_{\rm in}^{\frac{1}{2}}A_{ik}^*\rho_{\rm out}^{-\frac{1}{2}}=\sqrt{\eta_i\,p_{ik}}|i\rangle\langle\psi_{ik}|\rho_{\rm out}^{-\frac{1}{2}} \tag{61}$$

are Kraus operators for $\mathcal{R}_{\mathcal{M},\rho_{\rm in}}$. Thus

$$M_i^{\rm lsm}=\eta_i\rho_{\rm out}^{-\frac{1}{2}}\rho_i\rho_{\rm out}^{-\frac{1}{2}}=\sum_k R_{ik}^*R_{ik}=\mathcal{R}_{\mathcal{M},\rho_{\rm in}}^*(|i\rangle\langle i|)\,. \tag{62}$$

Conversely, it is immediate to verify that $\mathcal{R}_{\mathcal{M},\rho_{\rm in}}(\sigma)=\sum_{ik}R_{ik}\sigma R_{ik}^*=\sum_i\mathrm{tr}(M_i^{\rm lsm}\sigma)|i\rangle\langle i|$, hence $\mathcal{R}_{\mathcal{M},\rho_{\rm in}}$ is the operation associated to $\{M_i^{\rm lsm}\}$ by the relation (43).

V. QUANTUM STATE DISCRIMINATION

The carriers of information in quantum communication and quantum computing are quantum systems, and the information is encoded in the states of those systems. After processing the information, it is necessary to perform measurements in order to read out the result of the computation. In other words, one has to determine the output state of the system. If these possible outputs form a set of orthogonal states, that is, if they are given by m known density matrices ρ_i with orthogonal supports, then it is easy to devise a measurement which discriminates them without any error (a von Neumann measurement with projectors Π_i onto $\mathrm{ran}(\rho_i)$ will do the job). However, when the ρ_i are non-orthogonal a perfect discrimination is impossible. Indeed, if two non-orthogonal states $|\psi_1\rangle$ and $|\psi_2\rangle$ could be discriminated perfectly then one could duplicate those states by producing copies of $|\psi_i\rangle$ if the measurement outcome is $i=1,2$, without prior knowledge on which of the two states one actually possesses. This would contradict the no-cloning theorem. In its simplest form, this theorem tells us that no unitary evolution on a system S initially in state $|\psi\rangle$ and a register R initially in state $|\phi\rangle$ can transform $|\Psi\rangle=|\psi\rangle|\phi\rangle$ into $|\Psi'\rangle=|\psi\rangle|\psi\rangle$ for any $|\psi\rangle$ belonging to a set of distinct non-orthogonal states, e.g., $|\psi\rangle\in\{|\psi_1\rangle,|\psi_2\rangle\}$. Actually, the scalar products $\langle\Psi_1|\Psi_2\rangle=\langle\psi_1|\psi_2\rangle$ and $\langle\Psi_1'|\Psi_2'\rangle=\langle\psi_1|\psi_2\rangle^2$ are different if $\langle\psi_1|\psi_2\rangle\neq0,1$. More generally, one cannot duplicate unknown states by using any (not necessarily unitary) quantum evolution, except when these states pertain to a family of orthogonal states.[18] Consequently, one can extract less information from an ensemble of non-orthogonal states than from an ensemble of orthogonal ones.

It is of interest to find the best measurement to distinguish non-orthogonal states ρ_i with the smallest possible failure probability. We study this state discrimination problem in this section. This is a quite important issue in quantum cryptography and in quantum communication in general. As emphasized in the Introduction of this article, we aim at explaining some typical questions, providing examples, and establishing basic general results that will be used in Secs. VIII and XI, rather than giving a full account on the subject. We refer the reader to the review articles[38,26,25] for more complete presentations. Measurements for distinguishing quantum states can also be optimized using other criteria than the minimal probability of equivocation. For instance, one can try to maximize the mutual information between the initial distribution of the state ensemble and the distribution of the measurement outcomes. This optimization problem, which plays an important role in the transmission of information in quantum channels, is briefly discussed at the end of this section.

Before entering into the detail of the theory, let us make a philosophical remark concerning the quantum-classical differences. Let us inquire about the quantum analog of the celebrated experiment in classical probability which consists of picking up randomly colored balls contained in an urn. In quantum mechanics, the readout of the system's state (the color of the ball in the classical analogy) is performed by a measurement perturbing the system. If the urn contains an ensemble of non-orthogonal states, we have just seen above that there is no way to identify with certainty which state from the ensemble has been picked up. Therefore, the starting assumption that the color of the ball is known once it has been extracted from the urn is not fulfilled in the quantum world and identifying these colors is already a non-trivial task!

A. Discriminating quantum states drawn from a given ensemble

We review in this subsection two strategies for discriminating non-orthogonal states, known as the ambiguous and unambiguous state discriminations. Let us consider an ensemble $\{\rho_i, \eta_i\}_{i=1}^m$ of states ρ_i with prior probabilities η_i. For instance, the ρ_i can be some states of the electromagnetic field encoding m symbols of a given alphabet, the ith symbol occurring with frequency η_i. In order to send a message, a sender prepares random states drawn from the ensemble and gives them to a receiver. To decode the message the latter must identify these states by performing measurements. He wants to find the measurement that minimizes the failure probability.

A first strategy, called *ambiguous (or minimal error) quantum state discrimination*, consists in looking for a generalized measurement with m outcomes yielding the maximal success probability $P_S = \sum_i \eta_i p_{i|i}$, $p_{i|i}$ being the probability of the measurement outcome i given that the state is ρ_i. Here, the number of possible outcomes is chosen to be equal to the number of states in the ensemble. The conditional probability of the outcome j given the state ρ_i is (see Sec. III C)

$$p_{j|i} = \mathrm{tr}(M_j \rho_i) \tag{63}$$

so that the maximal success probability reads

$$P_S^{\mathrm{opt}}(\{\rho_i, \eta_i\}) = \max_{\mathrm{POVM}\ \{M_i\}} \left\{ \sum_{i=1}^m \eta_i \,\mathrm{tr}(M_i \rho_i) \right\}, \tag{64}$$

where the maximum is over all POVMs $\{M_i\}_{i=1}^m$.

A second strategy consists in seeking for a generalized measurement with $(m+1)$ outcomes enabling to identify perfectly each state ρ_i, but such that one of the outcomes leads to an inconclusive result. This strategy, originally proposed by Ivanovic[85] and further investigated by Dieks and Peres,[52,124] is called *unambiguous quantum state discrimination*. In other words, if the measurement outcome is $j \in \{1, \ldots, m\}$ then the receiver is certain that the state is ρ_j, whereas if $j = 0$ he does not know. This means that $p_{j|i} = p_{i|i}\,\delta_{ij}$ with $p_{i|i} > 0$, for any $i, j = 1, \ldots, m$. The probability of occurrence of the inconclusive outcome, $P_0 = \sum_i \eta_i p_{0|i}$, must be minimized. Since $p_{0|i} = 1 - p_{i|i}$, the success probability is obtained from the same formula (64) as for ambiguous discrimination, but with a maximum over all POVMs $\{M_j\}_{j=0}^m$ such that $\mathrm{tr}(M_j \rho_i) = p_{i|i}\delta_{ij}$ for $j \neq 0$. For pure states

$\rho_i = |\psi_i\rangle\langle\psi_i|$, the rank-one measurement operators M_j satisfying this condition are

$$M_j = \frac{p_{j|j}}{|\langle\psi_j^*|\psi_j\rangle|^2}|\psi_j^*\rangle\langle\psi_j^*|, \quad j = 1, \ldots, m\ , \tag{65}$$

with the dual normalized vectors $|\psi_j^*\rangle$ defined by $\langle\psi_j^*|\psi_i\rangle = \delta_{ij}\langle\psi_i^*|\psi_i\rangle$. The remaining problem is to find the values of the probabilities $p_{j|j}$ which maximize the success probability (64) under the constraint that $\{M_j\}_{j=0}^m$ is a POVM, that is,

$$M_0 = 1 - \sum_{j=1}^{m} \frac{p_{j|j}}{|\langle\psi_j^*|\psi_j\rangle|^2}|\psi_j^*\rangle\langle\psi_j^*| \ \geq\ 0. \tag{66}$$

This is a non-trivial problem, which has been solved so far in particular cases only. Upper and lower bounds on the maximal success probability can be found in terms of the scalar products $\langle\psi_i|\psi_j\rangle$ (see, e.g., Ref. 26).

It is worth noting that unambiguous discrimination is not always possible. For instance, a pure state ensemble $\{|\psi_i\rangle,\ \eta_i\}$ with linearly dependent vectors $|\psi_i\rangle$ cannot be discriminated unambiguously.[37] Indeed, assume that $|\psi_{i_0}\rangle$ is a linear combination of the other states $|\psi_i\rangle$. Together with the no-error condition $p_{j|i} = p_{i|i}\ \delta_{ij}$, which is equivalent to $|\psi_i\rangle \in \ker M_j$ for any $j \notin \{0,\ i\}$, this means that $|\psi_{i_0}\rangle \in \ker(M_{i_0})$ and thus $p_{i_0|i_0} = 0$, in contradiction with the requirement $p_{i_0|i_0} > 0$. The same argument shows that one cannot discriminate unambiguously an ensemble of mixed states $\{\rho_i,\ \eta_i\}$ such that one state ρ_{i_0} has its support $\mathrm{ran}(\rho_{i_0})$ contained in the sum of the supports of the other states.

Ambiguous and unambiguous quantum state discriminations have many applications. For instance, the discrimination of two non-orthogonal states plays a central role in the quantum cryptography protocol proposed by Bennett in 1992 to distribute a secrete key between two parties.[21] We will not elaborate further on these applications. Let us also mention that other optimization schemes than those discussed above have been worked out.[26,25] State discriminations have been implemented experimentally by using polarized photons in pure states (see Ref. 42 and references therein) and, more recently, in mixed states.[111]

B. Ambiguous and unambiguous discriminations of two states

1. *Ambiguous discrimination*

The simplest example of ambiguous discrimination is the case of $m = 2$ states ρ_1 and ρ_2. Then the optimal success probability and measurement are easy to determine.[74] One starts by writing the measurement operator M_2 as $1\ -\ M_1$ in the expression of the success probability,

$$P_{\mathrm{S,a}}^{\{M_i\}}(\{\rho_i, \eta_i\}) = \eta_1\,\mathrm{tr}(M_1\rho_1) + \eta_2\,\mathrm{tr}(M_2\rho_2) = \frac{1}{2}\big(1 - \mathrm{tr}\,\Lambda\big) + \mathrm{tr}(M_1\Lambda) \tag{67}$$

with $\Lambda = \eta_1\rho_1\ -\ \eta_2\rho_2$. The maximum of $\mathrm{tr}(M_1\Lambda)$ over all M_1 satisfying $0 \leq M_1 \leq 1$ is achieved when M_1 is the spectral projector Π_1 associated to the positive eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p > 0$ of the Hermitian matrix Λ. Consequently, the maximal success probability is given by the Helstrom formula

$$P_{\mathrm{S,a}}^{\mathrm{opt}}(\{\rho_i, \eta_i\}) = \frac{1}{2}\big(1 + \mathrm{tr}\,|\Lambda|\big), \quad \Lambda = \eta_1\rho_1 - \eta_2\rho_2. \tag{68}$$

The optimal measurement is a von Neumann measurement $\{\Pi_1^{\mathrm{opt}},\ 1 - \Pi_1^{\mathrm{opt}}\}$ with Π_1^{opt} the projector onto the support of $\Lambda_+ = (\Lambda\ +\ |\Lambda|)/2$. If $\Lambda \geq 0$ the optimal measurement is $\{\Pi_1^{\mathrm{opt}} = 1,\ \Pi_2^{\mathrm{opt}} = 0\}$, meaning that no measurement can outperform the simple guess that the state is ρ_1 (a similar statement holds for ρ_2 if $\Lambda \leq 0$). For pure states $\rho_i = |\psi_i\rangle\langle\psi_i|$, (68) reduces to

$$P_{\mathrm{S,a}}^{\mathrm{opt}}(\{|\psi_i\rangle, \eta_i\}) = \frac{1}{2}\Big(1 + \sqrt{1 - 4\eta_1\eta_2|\langle\psi_1|\psi_2\rangle|^2}\Big) \tag{69}$$

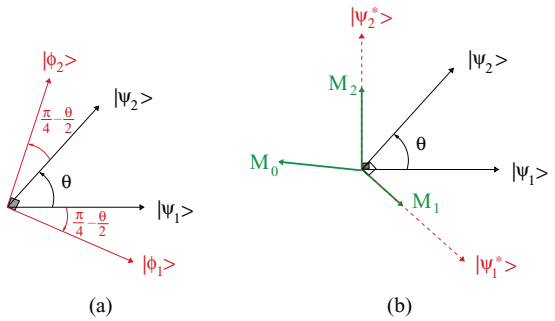

FIG. 1. Optimal measurement $\{M_i^{\rm opt}\}$ in the discrimination of two non-orthogonal pure states $|\psi_1\rangle$ and $|\psi_2\rangle$ with equal prior probabilities $\eta_i = 1/2$. (a) For ambiguous discrimination, $\{M_i^{\rm opt}\}$ is the von Neumann measurement in the two orthogonal states $|\phi_1\rangle$ and $|\phi_2\rangle$ with $|\langle\phi_i|\psi_i\rangle| = \cos(\frac{\pi}{4} - \frac{\theta}{2})$, that is, it is the least square measurement associated to $\{|\psi_i\rangle, \eta_i\}$ (b) For unambiguous discrimination, if the maximal prior probability $\eta_{\max}$ is larger than $q_1 = 1/(1 + \cos^2\theta)$, then the von Neumann measurement in the orthonormal basis $\{|\psi_1\rangle, |\psi_2^*\rangle\}$ (if $\eta_{\max} = \eta_2 > \eta_1$) or $\{|\psi_2\rangle, |\psi_1^*\rangle\}$ (if $\eta_{\max} = \eta_1 > \eta_2$) indicated by the red dashed vectors yields the smallest failure probability. Failure occurs when the outcome corresponds to the first vector in these two bases (inconclusive result). If $1 - q_1 < \eta_1 < q_1$, a smaller failure probability is obtained by using the generalized measurement with rank-one operators M_i indicated schematically by the green vectors.

and the optimal measurement consists of the rank-one eigenprojectors of Λ for the positive and negative eigenvalues. When $\eta_1 = \eta_2$, these are the projections onto the two orthogonal subspaces placed symmetrically with respect to $|\psi_1\rangle$ and $|\psi_2\rangle$, as represented in Fig. 1.

2. Unambiguous discrimination of two pure states

The power of generalized measurements is illustrated in the unambiguous discrimination of two pure states $|\psi_1\rangle$ and $|\psi_2\rangle$. Indeed, we will show that such measurements enable to distinguish quantum states better than von Neumann measurements (this can be considered as the main physical motivation to introduce generalized measurements[127]). Clearly, the Hilbert space $\mathcal{H}$ can be restricted to its two-dimensional subspace spanned by $|\psi_1\rangle$ and $|\psi_2\rangle$. The unambiguity condition implies $|\psi_1\rangle \in \ker M_2$ and $|\psi_2\rangle \in \ker M_1$, so that the measurement operators M_1 and M_2 are of rank one and given by (65). We can already observe at this point that the number of outcomes is larger than the space dimension, so that the unambiguous discrimination strategy cannot be realized with a von Neumann measurement.

The optimal success probability is given by[86]

$$P_{\rm S,u}^{\rm opt}(\{|\psi_i\rangle, \eta_i\}) = \begin{cases} 1 - 2\sqrt{\eta_1\eta_2}|\langle\psi_1|\psi_2\rangle| & \text{if } 1 - q_1 \le \eta_1 \le q_1 \\ \eta_{\max}(1 - |\langle\psi_1|\psi_2\rangle|^2) & \text{if } \eta_{\max} \ge q_1 \end{cases} \tag{70}$$

with $\eta_{\max} = \max\{\eta_1, \eta_2\}$ and $q_1 = 1/(1 + |\langle\psi_1|\psi_2\rangle|^2)$. It is instructive to establish this formula by using the Neumark extension theorem.[26] Thanks to Theorem 3.C.1, one can represent $\{M_j\}$ as a von Neumann measurement on the larger space $\mathcal{H} \otimes \mathcal{H}_{\mathsf{E}}$, with $\mathcal{H}_{\mathsf{E}} \simeq \mathbb{C}^3$. Let $\{A_j\}_{j=0}^2$ be the Kraus operators for the measurement and $|\epsilon_0\rangle$, U, and π_j^{E} be as in this theorem. We may assume that $\pi_j^{\mathsf{E}} = |j\rangle\langle j|$ are of rank one, where $\{|j\rangle\}_{j=0}^2$ is an orthonormal basis of $\mathcal{H}_{\mathsf{E}}$ (see the proof of Theorem 3.C.1). One writes

$$|\Psi_i'\rangle = U|\psi_i\rangle|\epsilon_0\rangle = \sum_{j=0}^{2} \sqrt{p_{j|i}}|\varphi_{j|i}\rangle|j\rangle \tag{71}$$

for $i = 1, 2$, where $\sqrt{p_{j|i}}|\varphi_{j|i}\rangle = \langle j|\Psi_i'\rangle \in \mathcal{H}$ are in general non-orthogonal for distinct j's and $\|\varphi_{j|i}\| = 1$. By (35) and (36) the unnormalized post-measurement states are $\widetilde{\rho}_{j|i} = \langle j|\Psi_i'\rangle\langle\Psi_i'|j\rangle = p_{j|i}|\varphi_{j|i}\rangle\langle\varphi_{j|i}|$, hence $p_{j|i}$ and $|\varphi_{j|i}\rangle$ can be interpreted as the probability of outcome j and the corresponding conditional state for the input state $|\psi_i\rangle$. Since we require $p_{2|1} = p_{1|2} = 0$, the unitarity of U imposes the conditions $p_{0|i} = 1 - p_{i|i}$ and $\langle\Psi_1'|\Psi_2'\rangle = \sqrt{p_{0|1}p_{0|2}}\langle\varphi_{0|1}|\varphi_{0|2}\rangle = \langle\psi_1|\psi_2\rangle$. The

last relation implies that the probabilities $p_{0|i}$ satisfy

$$p_{0|1}p_{0|2} \geq p_{0|1}p_{0|2}|\langle\varphi_{0|1}|\varphi_{0|2}\rangle|^2 = \cos^2\theta, \tag{72}$$

where we have set $\cos\theta = |\langle\psi_1|\psi_2\rangle|$. Note that this bound could have been obtained directly from (66), which is easy to solve since we are dealing here with 2×2 matrices.[26]

In order to maximize the success probability $P_S = \sum_i \eta_i p_{i|i} = 1 - \sum_i \eta_i p_{0|i}$, we are looking for the smallest possible $p_{0|1}$ and $p_{0|2}$. For such $p_{0|i}$'s the inequality (72) is an equality. Assuming $\cos\theta > 0$, this holds whenever $|\varphi_{0|2}\rangle = e^{i\delta}|\varphi_{0|1}\rangle$ with $\delta = \arg\langle\psi_1|\psi_2\rangle$. Accordingly, the conditional post-measurement state for the inconclusive outcome is the same irrespective of the input state $|\psi_i\rangle$. This is physically meaningful since if this post-measurement state was depending on $|\psi_i\rangle$ then one could perform a new measurement on it to increase further the success probability. In summary, for the optimal measurement one has

$$|\Psi_i'\rangle = \sqrt{p_{i|i}}|\varphi_{i|i}\rangle|i\rangle + \sqrt{p_{0|i}}\, e^{i\delta_i}|\Phi_0\rangle \tag{73}$$

with $|\Phi_0\rangle = |\varphi_{0|1}\rangle|0\rangle$ and $\delta_1 = 0$, $\delta_2 = \delta$.

The failure probability

$$P_0 = \eta_1 p_{0|1} + \eta_2 \frac{\cos^2\theta}{p_{0|1}} \tag{74}$$

is easy to minimize as a function of $p_{0|1}$. The minimum is achieved for $p_{0|1}^{\rm opt} = \sqrt{\eta_2/\eta_1}\cos\theta$ and is equal to $P_0^{\rm opt} = 2\sqrt{\eta_1\eta_2}\cos\theta$. This yields the upper expression in (70). The restrictions on the values of η_1 come from the conditions $p_{0|1}^{\rm opt} \leq 1$ and $p_{0|2}^{\rm opt} \leq 1$. When $\eta_1 \leq 1 - q_1$, the minimum is achieved for $p_{0|1}^{\rm opt} = 1$ and $p_{0|2}^{\rm opt} = \cos^2\theta$, i.e., $p_{1|1}^{\rm opt} = 0$ and $p_{2|2}^{\rm opt} = \sin^2\theta$. In such a case only the state $|\psi_2\rangle$ can be identified with certainty, as $|\psi_1\rangle$ always produces an inconclusive outcome. Strictly speaking this does not correspond to an unambiguous discrimination. One can nevertheless determine the optimal measurement, characterized by $M_1^{\rm opt} = 0$ and by two orthogonal projectors $M_2^{\rm opt} = |\psi_2^*\rangle\langle\psi_2^*|$ and $M_0^{\rm opt} = |\psi_1\rangle\langle\psi_1|$, see (65). A similar statement holds when $\eta_1 \geq q_1$ by exchanging the indices 1 and 2. The corresponding success probability is given by the lower expression in (70).

These results are summarized in Fig. 1. As claimed above, when $1 - q_1 < \eta_1 < q_1$ generalized measurements, obtained via a coupling of the system with an ancilla and a measurement on the latter, do better in decoding the message than a von Neumann measurement performed directly on the system.

3. Unambiguous discrimination of two mixed states

Let us now turn to the case of two mixed states ρ_1 and ρ_2. Such states cannot be unambiguously discriminated when $\mathrm{ran}\,\rho_1$ is contained in $\mathrm{ran}\,\rho_2$ or vice versa. By the unambiguity condition, $\mathrm{ran}\,M_1 \subset \ker\rho_2$ and $\mathrm{ran}\,M_2 \subset \ker\rho_1$. A trivial situation is when $\ker\rho_1 \perp \ker\rho_2$, in which case the optimal POVM is the von Neumann measurement with M_1 and M_2 equal to the projectors on $\ker\rho_2$ and $\ker\rho_1$, respectively. Then the minimal failure probability is $P_0^{\rm opt} = \mathrm{tr}[(\eta_1\rho_1 + \eta_2\rho_2)\Pi_0]$, Π_0 being the projector onto $\mathrm{ran}\,\rho_1 \cap \mathrm{ran}\,\rho_2$. One can as before restrict the Hilbert space so that $\mathrm{ran}\,\rho_1 + \mathrm{ran}\,\rho_2 = \mathcal{H}$. If $\mathrm{ran}\,\rho_1$ and $\mathrm{ran}\,\rho_2$ have co-dimension one in $\mathcal{H}$, then M_1 and M_2 are of rank one and take the form (65) with $|\psi_1^*\rangle \in \ker\rho_2$, $|\psi_2^*\rangle \in \ker\rho_1$, and $|\langle\psi_i^*|\psi_i\rangle|^2$ replaced by $R_i = \langle\psi_i^*|\rho_i|\psi_i^*\rangle$. A simple generalization of (70) then yields[135]

$$P_{\rm S,u}^{\rm opt}(\{\rho_i,\eta_i\}) = P_{\rm S}^{\rm opt}(R_i,\eta_i) \equiv \begin{cases} \dfrac{\eta_1R_1+\eta_2R_2-2\sqrt{\eta_1\eta_2R_1R_2}\cos\theta}{\sin^2\theta} & \text{if } \cos^2\theta < \min\left\{\dfrac{\eta_1R_1}{\eta_2R_2}, \dfrac{\eta_2R_2}{\eta_1R_1}\right\} \\ \max\{\eta_1R_1, \eta_2R_2\} & \text{otherwise} \end{cases} \tag{75}$$

with $\cos\theta = |\langle\psi_1^*|\psi_2^*\rangle|$. For kernels of dimensions $d_2 \geq d_1 > 1$, by a standard linear algebra argument one can construct two orthonormal bases $\{|\psi_{2k}^*\rangle\}_{k=1}^{d_1}$ of $\ker\rho_1$ and $\{|\psi_{1k}^*\rangle\}_{k=1}^{d_2}$ of $\ker\rho_2$ such that $\langle\psi_{1k}^*|\psi_{2l}^*\rangle = \delta_{kl}\cos\theta_k$, with $\theta_k \in [0, \pi/2]$. Let us take $M_i = \sum_k M_{ik}$ for $i = 1, 2$, with $M_{ik} = m_{ik}|\psi_{ik}^*\rangle\langle\psi_{ik}^*|$. Optimizing $P_{\rm S,u}^{\{M_i\}}$ over the non-negative numbers m_{ik} under the constraint

$1 - M_1 - M_2 \geq 0$ reduces to the optimization problem for rank-one measurement operators studied before (in fact, this constraint is equivalent to $1 - M_{1k} - M_{2k} \geq 0$ for $k = 1, \ldots, d_1$ and $1 - M_{1k} \geq 0$ for $d_1 < k \leq d_2$). This gives the lower bound[135]

$$P_{\mathrm{S,u}}^{\mathrm{opt}}(\{\rho_i, \eta_i\}) \geq \sum_{k=1}^{d_1} P_{\mathrm{S}}^{\mathrm{opt}}(R_{ik}, \eta_i) + \eta_1 \sum_{d_1 < k \leq d_2} R_{1k} \quad \text{with } R_{ik} = \langle \psi_{ik}^* | \rho_i | \psi_{ik}^* \rangle. \tag{76}$$

An upper bound can be obtained in terms of the fidelity between the states ρ_1 and ρ_2 defined by $F(\rho_1, \rho_2) = (\mathrm{tr}|\sqrt{\rho_1}\sqrt{\rho_2}|)^2$ (see Proposition 5.E.2 and Remark 7.D.4 below),[135]

$$P_{\mathrm{S,u}}^{\mathrm{opt}}(\{\rho_i, \eta_i\}) \leq \begin{cases} 1 - 2\sqrt{\eta_1 \eta_2 F(\rho_1, \rho_2)} & \text{if } F(\rho_1, \rho_2) < \dfrac{\eta_{\min}}{\eta_{\max}} \\ \eta_{\max}(1 - F(\rho_1, \rho_2)) & \text{otherwise.} \end{cases} \tag{77}$$

A nice application of two mixed state discrimination is the state comparison problem.[17] Consider two independent copies of a given system, the state of which is drawn from the pure state ensemble $\{|\psi_i\rangle, 1/2\}_{i=1,2}$. One would like to decide with the help of an appropriate measurement if the two copies are in the same state or not, without further information on the actual state of each copies. If $|\psi_1\rangle$ and $|\psi_2\rangle$ are not orthogonal, this can only be done with a probability of success $P_{\mathrm{S,\,comp}} < 1$. This amounts to discriminate the two mixed states

$$\begin{aligned} \rho_{\mathrm{eq}} &= \frac{1}{2}|\psi_1 \otimes \psi_1\rangle\langle\psi_1 \otimes \psi_1| + \frac{1}{2}|\psi_2 \otimes \psi_2\rangle\langle\psi_2 \otimes \psi_2| \\ \rho_{\mathrm{diff}} &= \frac{1}{2}|\psi_1 \otimes \psi_2\rangle\langle\psi_1 \otimes \psi_2| + \frac{1}{2}|\psi_2 \otimes \psi_1\rangle\langle\psi_2 \otimes \psi_1| \,. \end{aligned} \tag{78}$$

It is shown in Ref. 135 that for such mixed states of rank two, the lower and upper bounds in (76) and (77) coincide. A simple calculation (see Remark 7.D.4 below) then gives the optimal success probability[17]

$$P_{\mathrm{S,comp}}^{\mathrm{opt}} = 1 - |\langle\psi_1|\psi_2\rangle| \,. \tag{79}$$

C. Discrimination with least square measurements

How well does the least square measurement (Sec. IV C) in discriminating ambiguously quantum states? More precisely, let

$$P_{\mathrm{S,a}}^{\mathrm{lsm}}(\{\rho_i, \eta_i\}) = \sum_i \eta_i \,\mathrm{tr}(\rho_i M_i^{\mathrm{lsm}}) \tag{80}$$

be the success probability in discriminating the states ρ_i by performing the least square measurement $\{M_i^{\mathrm{lsm}}\}$ associated to $\{\rho_i, \eta_i\}$. We would like to compare $P_{\mathrm{S,a}}^{\mathrm{lsm}}$ with the optimal success probability.

Let us first observe that if $\rho_i = \mathcal{M}(|i\rangle\langle i|)$, $\mathcal{M}$ being a quantum operation on $\mathcal{B}(\mathcal{H})$ and $\{|i\rangle\}_{i=1}^n$ a fixed orthonormal basis of $\mathcal{H}$, then $P_{\mathrm{S,\,a}}(\{\rho_i, \eta_i\})$ is related to the entanglement fidelity defined in Sec. IV B. Recall that any ensemble $\{\rho_i, \eta_i\}_{i=1}^m$ with $m \leq n$ states can be obtained in this way from an operation $\mathcal{M} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ (since $m \leq n$ we can identify here the pointer space $\mathcal{H}_{\mathsf{P}}$ with a subspace of $\mathcal{H}$, see Sec. III D). To establish the relation with the average fidelity (49), consider a POVM $\{M_i\}_{i=1}^m$ with m measurement operators and let us associate to it the quantum operation $\mathcal{R}$ on $\mathcal{B}(\mathcal{H})$ defined by $\mathcal{R}^*(|i\rangle\langle j|) = M_i \delta_{ij}$. Then

$$P_{\mathrm{S,a}}^{\{M_i\}}\big(\{\rho_i, \eta_i\}\big) = \sum_{i=1}^m \eta_i \,\mathrm{tr}[\mathcal{R}^*(|i\rangle\langle i|)\rho_i] = \sum_{i=1}^m \eta_i \langle i|\mathcal{R} \circ \mathcal{M}(|i\rangle\langle i|)|i\rangle = \overline{F}_{\mathrm{e}}\big(\{|i\rangle, \eta_i\}, \mathcal{R} \circ \mathcal{M}\big) \tag{81}$$

thanks to the equality of the entanglement fidelity with the input-output fidelity for pure states. In view of the one-to-one correspondence between POVMs with $m \leq n$ operators and quantum operations on $\mathcal{B}(\mathcal{H})$ we obtain the following relation between $P_{\mathrm{S,\,a}}^{\mathrm{opt}}$ and the maximal fidelity over all recovery operations $\mathcal{R}$ on $\mathcal{B}(\mathcal{H})$:

$$P_{\mathrm{S,\,a}}^{\mathrm{opt}}(\{\rho_i, \eta_i\}_{i=1}^m) = \max_{\mathcal{R}} \big\{\overline{F}_{\mathrm{e}}(\{|i\rangle, \eta_i\}_{i=1}^m, \mathcal{R} \circ \mathcal{M})\big\}, \quad m \leq n. \tag{82}$$

Furthermore, the optimal measurement operators are given in terms of the optimal recovery operation $\mathcal{R}^{\rm opt}$ by $M_i^{\rm opt} = (\mathcal{R}^{\rm opt})^*(|i\rangle\langle i|)$. According to Proposition 4.B.1, taking $\mathcal{R}$ to be the transpose operation $\mathcal{R}_{\mathcal{M},\rho_{\rm in}}$ of $\mathcal{M}$ for the state $\rho_{\rm in} = \sum_i \eta_i |i\rangle\langle i|$ gives an entanglement fidelity larger than the square of the right-hand side of (82). But the measurement associated to $\mathcal{R}_{\mathcal{M},\rho_{\rm in}}$ is the least square measurement, i.e., $M_i^{\rm lsm} = \mathcal{R}^*_{\mathcal{M},\rho_{\rm in}}(|i\rangle\langle i|)$ (see Sec. IV C). As a result, Proposition 4.B.1 yields the following inequality.

Corollary 5.C.1. *If $m \le n = \dim \mathcal{H}$, then*

$$P_{\rm S,a}^{\rm opt}(\{\rho_i, \eta_i\}_{i=1}^m) \le \left(P_{\rm S,a}^{\rm lsm}(\{\rho_i, \eta_i\}_{i=1}^m)\right)^{\frac{1}{2}} . \tag{83}$$

Thus, if the error probability for discriminating $\{\rho_i, \eta_i\}$ using the least square measurement is small, then it is at most twice the minimal error probability $P_{\rm err,a}^{\rm opt} = 1 - P_{\rm S,a}^{\rm opt}$, up to a small correction of the order of $(P_{\rm S,a}^{\rm lsm})^2$. Small error probabilities occur for almost orthogonal states. Therefore, for such states least square measurements are nearly optimal.[71,19]

It is worth mentioning that least square measurements are also asymptotically optimal for discriminating ambiguously equiprobable linearly independent pure states.[78] In addition, they optimally discriminate equiprobable states drawn from a symmetric ensemble, like, for instance, the states $\rho_i = U^{i-1}\rho_1(U^{i-1})^*$ related between themselves through conjugations by powers of a single unitary operator U satisfying $U^m = \pm 1$ (see Refs. 15, 16, 41, and 54 and references therein). Necessary and sufficient conditions for the optimality of least square measurements in state discrimination have been investigated in Refs. 57 and 138.

D. General results on ambiguous discrimination

Let $\{\rho_i, \eta_i\}_{i=1}^m$ be an ensemble of m states of a system with a n-dimensional Hilbert space $\mathcal{H}$. Hereafter, we assume that $\eta_i > 0$ for all $i = 1, \ldots, m$, so that m is the actual number of states to discriminate. We denote by $\widetilde{\rho}_i = \eta_i \rho_i$ the unnormalized states with trace equal to the prior probability η_i. To shorten notation, the dependence of the success probability $P_{\rm S}$ on the ensemble is not written explicitly. The following proposition contains one of the few results in ambiguous discrimination applying to arbitrary ensembles.

Proposition 5.D.1.[77,171,55] *The optimal success probability in ambiguous state discrimination is given by*

$$P_{\rm S,a}^{\rm opt} = \inf_{\Upsilon \ge \widetilde{\rho}_i} \left\{ {\rm tr}(\Upsilon) \right\} , \tag{84}$$

where the infimum is over all self-adjoint operators Υ satisfying $\Upsilon \ge \widetilde{\rho}_i$ for any $i = 1, \ldots, m$. Moreover, the POVM $\{M_i^{\rm opt}\}_{i=1}^m$ is optimal if and only if the operator $\Upsilon^{\rm opt} = \sum_i \widetilde{\rho}_i M_i^{\rm opt}$ satisfies the two conditions

(i) $\Upsilon^{\rm opt}$ *is self-adjoint*;
(ii) $\Upsilon^{\rm opt} \ge \widetilde{\rho}_i$ *for any* $i = 1, \ldots, m$.

In such a case, the infimum in the right-hand side of (84) is attained for $\Upsilon = \Upsilon^{\rm opt}$.

The fact that (ii) is sufficient to ensure the optimality of $\{M_i^{\rm opt}\}$ is obvious from the relation

$$P_{\rm S,a}^{\rm opt} - P_{\rm S,a}^{\{M_i\}} = \sum_{i=1}^m {\rm tr}[(\Upsilon^{\rm opt} - \widetilde{\rho}_i) M_i]. \tag{85}$$

The necessary and sufficient conditions (i) and (ii) are due to Holevo,[77] who derived them by considering a specific one-parameter family $\{M_i(\varepsilon)\}$ of POVMs such that $M_i(0) = M_i^{\rm opt}$ and by exploiting the fact that $\partial P_{\rm S,a}^{\{M_i(\varepsilon)\}}/\partial\varepsilon = 0$ for $\varepsilon = 0$ (see Ref. 74, Chap. 4). Yuen, Kennedy, and Lax[171] proposed another derivation based on a duality argument in vector space optimization. We shall present below the related proof of Eldar, Megretski, and Verghese.[55]

Let us note that (i) and (ii) imply

$$(\Upsilon^{\rm opt} - \widetilde{\rho}_i)M_i^{\rm opt} = M_i^{\rm opt}(\Upsilon^{\rm opt} - \widetilde{\rho}_i) = 0, \quad i = 1, \ldots, m \ . \tag{86}$$

In fact, since $\sum_i {\rm tr}[(\Upsilon^{\rm opt} - \widetilde{\rho}_i)M_i^{\rm opt}] = 0$ and $\Upsilon^{\rm opt} - \widetilde{\rho}_i \geq 0$ by (ii), one deduces that $|(\Upsilon^{\rm opt} - \widetilde{\rho}_i)^{1/2}(M_i^{\rm opt})^{1/2}|^2 = 0$ (recall that $A \geq 0$ and ${\rm tr}(A) = 0$ imply $A = 0$). One concludes from this equality that $(\Upsilon^{\rm opt} - \widetilde{\rho}_i)M_i^{\rm opt} = 0$. It is easy to see by eliminating $\Upsilon^{\rm opt}$ that (86) is equivalent to

$$M_i^{\rm opt}(\widetilde{\rho}_i - \widetilde{\rho}_j)M_j^{\rm opt} = 0, \quad i, j = 1, \ldots, m \ . \tag{87}$$

The condition (87) automatically implies that $\Upsilon^{\rm opt}$ is self-adjoint. Hence a necessary and sufficient condition for $\{M_i^{\rm opt}\}$ to be optimal is given by conditions (ii) and (87).

Except in special cases such as ensembles of equiprobable states related by a symmetry,[15,16,54,41] it is difficult in practice to obtain the optimal measurement and success probability from the above necessary and sufficient conditions. Nevertheless, the formulas (84) and (86) are helpful for computing these quantities numerically. For indeed, the minimization task in (84) is simpler than the maximization in (64) and can be solved efficiently with the help of convex semidefinite programs.[55]

Proof. The main idea is to show that the minimization problem in (84) is dual to the maximization problem in (64). More precisely, there exists a convex set $\Gamma \subset \mathcal{B}(\mathcal{H})_{\rm s.a.}$ such that

$$P_{\rm S,a}^{\{M_i\}} \leq {\rm tr}(\Upsilon), \qquad \forall\, \{M_i\} \text{ POVM}, \ \forall\, \Upsilon \in \Gamma, \tag{88}$$

and the maximum of the left-hand member is equal to the minimum of the right-hand member, i.e., $P_{\rm S,a}^{\rm opt} = \min_{\Upsilon\in\Gamma} {\rm tr}(\Upsilon)$. The set Γ is defined by

$$\Gamma = \left\{\Upsilon \in \mathcal{B}(\mathcal{H})_{\rm s.a.}\,;\ \Upsilon \geq \widetilde{\rho}_i,\ i = 1, \ldots, m\right\}. \tag{89}$$

Then ${\rm tr}(\Upsilon) - P_{\rm S,a}^{\{M_i\}} = \sum_i {\rm tr}[(\Upsilon - \widetilde{\rho}_i)M_i] \geq 0$ for any $\Upsilon \in \Gamma$, so that (88) holds true. Let us now define the following convex subset Ω of the real vector space $\mathcal{B}(\mathcal{H})_{\rm s.a.} \times \mathbb{R}$:

$$(B, x) \in \Omega \quad \Leftrightarrow \quad B = \sum_{i=1}^m B_i - 1\ ,\ x = r - \sum_{i=1}^m {\rm tr}(B_i\widetilde{\rho}_i) \ \text{ with } \ B_i \geq 0 \ \text{ and } \ r > P_{\rm S,a}^{\rm opt}\ . \tag{90}$$

This space is endowed with the scalar product $\langle (B, x)\,,\ (C, y)\rangle = {\rm tr}(BC) + xy$. Since Ω is convex and does not contain (0, 0), by the separating hyperplane theorem one can find a non-vanishing vector $(\Upsilon_a, a) \in \mathcal{B}(\mathcal{H})_{\rm s.a.} \times \mathbb{R}$ such that $\langle (\Upsilon_a, a)\,, (B, x)\rangle \geq 0$ for any $(B, x) \in \Omega$, that is

$${\rm tr}\Big[\Upsilon_a\Big(\sum_{i=1}^m B_i - 1\Big)\Big] + a\Big(r - \sum_{i=1}^m {\rm tr}(B_i\widetilde{\rho}_i)\Big) \geq 0\ . \tag{91}$$

Taking $B_i = t|\varphi\rangle\langle\varphi|$ if $i = k$ and zero otherwise, with $|\varphi\rangle \in \mathcal{H}$ and $t > 0$, and letting $t \to \infty$, we obtain $\langle\varphi|\Upsilon_a|\varphi\rangle - a\langle\varphi|\widetilde{\rho}_k|\varphi\rangle \geq 0$. But $|\varphi\rangle$ and k are arbitrary, hence

$$\Upsilon_a \geq a\widetilde{\rho}_i, \quad i = 1, \ldots, m\ . \tag{92}$$

Similarly, taking $B_i = 0$ for all i and $r \to P_{\rm S,a}^{\rm opt}$, (91) yields

$$a\,P_{\rm S,a}^{\rm opt} \geq {\rm tr}(\Upsilon_a)\ . \tag{93}$$

From the same choice of B_i and $r \to \infty$ one gets $a \geq 0$. If $a = 0$ then $\Upsilon_a \geq 0$ and ${\rm tr}(\Upsilon_a) = 0$ by (92) and (93). This would imply $\Upsilon_a = 0$, in contradiction with $(\Upsilon_a, a) \neq (0, 0)$. Thus $a > 0$. The self-adjoint operator $\Upsilon^{\rm opt} = \Upsilon_a/a$ satisfies $\Upsilon^{\rm opt} \geq \widetilde{\rho}_i$ for all i (i.e., $\Upsilon^{\rm opt} \in \Gamma$) and ${\rm tr}(\Upsilon^{\rm opt}) \leq P_{\rm S,a}^{\rm opt}$, see (92) and (93). The converse of the last inequality follows from (88). Whence $P_{\rm S,a}^{\rm opt} = {\rm tr}(\Upsilon^{\rm opt}) = \min_{\Upsilon\in\Gamma} {\rm tr}(\Upsilon)$, as claimed in the proposition. This identity implies $\sum_i {\rm tr}[(\Upsilon^{\rm opt} - \widetilde{\rho}_i)M_i^{\rm opt}] = 0$ if $\{M_i^{\rm opt}\}$ is an optimal POVM. But all traces in the sum are non-negative, thus they vanish and (86) is satisfied by the arguments given above to derive this equation. It results from (86) that $\Upsilon^{\rm opt} = \sum_i \widetilde{\rho}_i M_i^{\rm opt} = \sum_i M_i^{\rm opt}\widetilde{\rho}_i$. This concludes the proof. □

Let us consider the success probability

$$P_{\mathrm{S,a}}^{\mathrm{opt\ v.N.}}(\{\rho_i,\eta_i\}) = \max_{\{\Pi_i\}}\left\{\sum_{i=1}^m \eta_i \operatorname{tr}(\Pi_i\rho_i)\right\}, \tag{94}$$

where the maximum is over all von Neumann measurements $\{\Pi_i\}_{i=1}^m$. A natural question is whether this probability may be equal to $P_{\mathrm{S,a}}^{\mathrm{opt}}$, i.e., whether the states ρ_i may be discriminated optimally with a von Neumann measurement. We have already argued above that this is not always the case, even for pure states. A simple consequence of Proposition 5.D.1 is that the equality holds for *linearly independent states*. The states ρ_i are called linearly independent if their eigenvectors $|\zeta_{ij}\rangle$ with non-zero eigenvalues form a linearly independent family $\{|\zeta_{ij}\rangle\}_{i=1,\ldots,m}^{j=1,\ldots,r_i}$ in $\mathcal{H}$ (here r_i is the rank of ρ_i). We say that they span the Hilbert space $\mathcal{H}$ if $\mathcal{H} = \operatorname{span}\{|\zeta_{ij}\rangle\}_{i=1,\ldots,m}^{j=1,\ldots,r_i}$. Without loss of generality one can restrict $\mathcal{H}$ to a subspace $\mathcal{H}'$ spanned by the ρ_i.

Corollary 5.D.2.[55] *Let $\{|\psi_i\rangle,\eta_i\}_{i=1}^m$ be an ensemble of pure states spanning $\mathcal{H}$. Then the optimal measurement operators M_i^{opt} in ambiguous state discrimination are of rank one. More generally, for any ensemble $\{\rho_i,\eta_i\}_{i=1}^m$ spanning $\mathcal{H}$, the optimal measurement operators have ranks $\operatorname{rank}(M_i^{\mathrm{opt}}) \le \operatorname{rank}(\rho_i)$ for all $i = 1,\ldots,m$.*

Corollary 5.D.3.[56] *Let $\{\rho_i,\eta_i\}_{i=1}^m$ be an ensemble of linearly independent states spanning $\mathcal{H}$. Then an optimal measurement in ambiguous state discrimination is a von Neumann measurement with orthogonal projectors $M_i^{\mathrm{opt}} = \Pi_i^{\mathrm{opt}}$ of rank $r_i = \operatorname{rank}(\rho_i)$. In particular, the probabilities (64) and (94) are equal.*

Proof. Let us set $N_i^{\mathrm{opt}} = \Upsilon^{\mathrm{opt}} - \widetilde{\rho}_i$. The relation (86) implies $\operatorname{ran} M_i^{\mathrm{opt}} \subset \ker N_i^{\mathrm{opt}}$, hence $\operatorname{rank}(M_i^{\mathrm{opt}}) \le \dim(\ker N_i^{\mathrm{opt}})$. Since the rank of the sum of two matrices is smaller or equal to the sum of their ranks, $\operatorname{rank}(\Upsilon^{\mathrm{opt}}) \le \operatorname{rank}(N_i^{\mathrm{opt}}) + r_i$ and thus $\dim(\ker N_i^{\mathrm{opt}}) \le \dim(\ker \Upsilon^{\mathrm{opt}}) + r_i$. But $\ker \Upsilon^{\mathrm{opt}} \subset [\operatorname{ran}(\rho_i)]^{\perp}$ for all i according to the condition (ii) of Proposition 5.D.1. Consequently, if the states ρ_i span $\mathcal{H}$ then $\ker \Upsilon^{\mathrm{opt}} = \{0\}$. This shows that $\operatorname{rank}(M_i^{\mathrm{opt}}) \le r_i$. If furthermore the ρ_i are linearly independent, then $\sum_i r_i = n = \dim \mathcal{H}$. Introducing the spectral decomposition $M_i^{\mathrm{opt}} = \sum_k |\widetilde{\mu}_{ik}\rangle\langle\widetilde{\mu}_{ik}|$ with unnormalized vectors $|\widetilde{\mu}_{ik}\rangle$, $k = 1,\ldots,r_i$, and noting that the sum $\sum_{i,k} |\widetilde{\mu}_{ik}\rangle\langle\widetilde{\mu}_{ik}| = 1$ contains at most n terms, it follows that $\{|\widetilde{\mu}_{ik}\rangle\}$ is an orthonormal basis of $\mathcal{H}$. Thus M_i^{opt} are orthogonal projectors of rank r_i. □

E. Bounds on the maximal success probability

We now establish some inequalities satisfied by $P_{\mathrm{S}}^{\mathrm{opt}}$ for any number m of states to discriminate. A review of various upper bounds for ambiguous discrimination can be found in Ref. 132. We only discuss here the bounds involving the fidelity

$$F(\rho,\sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1^2 = \left(\operatorname{tr}[(\sqrt{\sigma}\rho\sqrt{\sigma})^{\frac{1}{2}}]\right)^2. \tag{95}$$

The properties of this fidelity will be analyzed in Sec. VII. Let us only mention here that $F(\rho,\sigma)$ is symmetric under the exchange of ρ and σ (actually, $\sqrt{\sigma}\rho\sqrt{\sigma}$ and $\sqrt{\rho}\sigma\sqrt{\rho}$ have the same non-zero eigenvalues) and reduces for pure states $\rho_\psi = |\psi\rangle\langle\psi|$ and $\sigma_\phi = |\phi\rangle\langle\phi|$ to the square modulus of the scalar product $\langle\psi|\phi\rangle$, i.e., $F(\rho_\psi,\sigma_\phi) = |\langle\psi|\phi\rangle|^2$. More generally, $F(\rho,\sigma)$ can be seen as a measure of non-orthogonality of ρ and σ.

The following lower and upper bounds on the maximum success probability $P_{\mathrm{S,a}}^{\mathrm{opt}}$ for ambiguous state discrimination are taken from Refs. 19 and 112, respectively. The upper bound is established in Ref. 19 (and is often reported in subsequent works) with an unnecessary extra factor of two in front of the sum (after correcting the obvious misprints in this reference).

Proposition 5.E.1. (Barnum and Knill,[19] Montanaro[112]). *For any ensemble* $\{\rho_i, \eta_i\}_{i=1}^m$, *one has*

$$1 - \sum_{i>j} \sqrt{\eta_i \eta_j F(\rho_i, \rho_j)} \leq P_{\mathrm{S,a}}^{\mathrm{opt}}(\{\rho_i, \eta_i\}) \leq 1 - \sum_{i>j} \eta_i \eta_j F(\rho_i, \rho_j)\,. \tag{96}$$

The inequalities (96) make quantitative the intuitive fact that the more pairwise orthogonal are the states ρ_i, the larger is the success probability to discriminate them, and conversely.

Proof. Let $\rho_i = A_i A_i^*$, the operators A_i being, for instance, given by (56). Given a POVM $\{M_i\}$ with Kraus operators R_i (i.e., $M_i = R_i^* R_i$), we set

$$S_{ij} = \sqrt{\eta_j} R_i A_j, \quad B_{ij} = \sqrt{\eta_i \eta_j} A_i^* A_j. \tag{97}$$

We view $S = (S_{ij})_{i,j=1}^m$ and $B = (B_{ij})_{i,j=1}^m$ as $m \times m$ matrices with values in $\mathcal{B}(\mathcal{H})$, which are related by $S^*S = B \geq 0$ (this follows from $\sum_i R_i^* R_i = 1$). Observe that

$$P_{\mathrm{S,a}}^{\{M_i\}} = \sum_j \eta_j \operatorname{tr}(M_j \rho_j) = 1 - \sum_{i \neq j} \eta_j \operatorname{tr}(M_i \rho_j) = 1 - \sum_{i \neq j} \|S_{ij}\|_2^2 \tag{98}$$

and

$$\eta_i \eta_j F(\rho_i, \rho_j) = \eta_i \eta_j \|\sqrt{\rho_i}\sqrt{\rho_j}\|_1^2 = \eta_i \eta_j \|U_i^* \sqrt{\rho_i}\sqrt{\rho_j} U_j\|_1^2 = \|B_{ij}\|_1^2, \tag{99}$$

where $\|\cdot\|_{1,2}$ are the trace and Hilbert Schmidt norms. We have used in (99) the polar decomposition $A_i = \sqrt{\rho_i} U_i$ and the unitary invariance of these norms. The main idea to prove the first inequality in (96) is to bound from below the optimal success probability $P_{\mathrm{S,a}}^{\mathrm{opt}}$ by the success probability $P_{\mathrm{S,a}}^{\mathrm{lsm}}$ for discriminating the states with the least square measurement.[19] For the latter, the matrix S in (97) is the square root of B (in fact, according to (57), $S_{ij}^{\mathrm{lsm}} = \sqrt{\eta_i \eta_j} A_i^* \rho_{\mathrm{out}}^{-1/2} A_j$ so that $S^{\mathrm{lsm}} \geq 0$, and it has been argued above that $|S|^2 = B$). For instance, if the ρ_i are pure states $|\psi_i\rangle$, B and S^{lsm} can be identified with the scalar product matrices $(\langle\widetilde{\psi}_i|\widetilde{\psi}_j\rangle)_{i,j=1}^m$ and $(\langle\widetilde{\mu}_i|\widetilde{\psi}_j\rangle)_{i,j=1}^m$, respectively, with $|\widetilde{\psi}_i\rangle = \sqrt{\eta_i}|\psi_i\rangle$ and $|\widetilde{\mu}_i\rangle = \sqrt{\eta_i}\rho_{\mathrm{out}}^{-1/2}|\psi_i\rangle$, the latter being the vectors describing the least square measurement (Sec. IV C). The identity $S^{\mathrm{lsm}} = \sqrt{B}$ then becomes evident from the definition of a POVM. (This remarkable identity has been singled out for pure states in Ref. 72. The authors of this reference suggest to use it as a definition of the least square measurement.) Therefore, in view of (98), $P_{\mathrm{S,a}}^{\mathrm{opt}} \geq P_{\mathrm{S,a}}^{\mathrm{lsm}} = 1 - \sum_{i \neq j} \|(\sqrt{B})_{ij}\|_2^2$. The lower bound in (96) comes from the following norm inequality proven in Appendix B: for any fixed $j = 1, \ldots, m$,

$$\sum_{i, i \neq j} \|(\sqrt{B})_{ij}\|_2^2 \leq \frac{1}{2} \sum_{i, i \neq j} \|B_{ij}\|_1\,, \tag{100}$$

where the last sum is related to the fidelities by (99).

It remains to establish the upper bound. With the notation above, this bound takes the form

$$\frac{1}{2} \sum_{i \neq j} \|B_{ij}\|_1^2 \leq \sum_{i \neq j} \|S_{ij}\|_2^2\,. \tag{101}$$

Fixing j again and introducing the notation $\|\cdot\|_{1/2}$ as in (2) (note that this is not a norm), if one can show that

$$\Big\| \sum_{i, i \neq j} |B_{ij}|^2 \Big\|_{\frac{1}{2}} \leq \sum_{i, i \neq j} \Big(\|S_{ij}\|_2^2 + \|S_{ji}\|_2^2 \Big) \tag{102}$$

then the required inequality (101) will be proven. Actually, $\sum_i \|B_{ij}\|_1^2 = \sum_i \||B_{ij}|^2\|_{1/2} \leq \|\sum_i |B_{ij}|^2\|_{1/2}$ by the inverse Minkowski inequality (B1) in Appendix B. In order to show (102), let us introduce the following $(m-1) \times (m-1)$ matrices with values in $\mathcal{B}(\mathcal{H})$:

$$\begin{aligned} C^{(j)} &= \sum_{i, i \neq j} (S_{ji})^* \otimes |i\rangle\langle 1|, \quad D^{(j)} = S_{jj} \otimes |1\rangle\langle 1| \\ E^{(j)} &= \sum_{i \neq j} \sum_{k \neq j} (S_{ki})^* \otimes |i\rangle\langle k|, \; F^{(j)} = \sum_{k, k \neq j} S_{kj} \otimes |k\rangle\langle 1| \end{aligned} \tag{103}$$

(here $|i\rangle\langle k|$ stands for the matrix with vanishing entries except in the ith raw and kth column, which has a unit entry). An explicit calculation leads to

$$\left\|C^{(j)}D^{(j)}+E^{(j)}F^{(j)}\right\|_1^2=\left\|\sum_{i,i\neq j}|B_{ij}|^2\right\|_{\frac{1}{2}},\quad \left\|C^{(j)}\right\|_2^2=\sum_{i,i\neq j}\|S_{ji}\|_2^2,\quad \left\|F^{(j)}\right\|_2^2=\sum_{k,k\neq j}\|S_{kj}\|_2^2. \tag{104}$$

Furthermore,

$$\left\|C^{(j)}\right\|_2^2+\left\|D^{(j)}\right\|_2^2+\left\|E^{(j)}\right\|_2^2+\left\|F^{(j)}\right\|_2^2=\sum_{i,k}\|S_{ik}\|_2^2=\sum_k\eta_k\,\mathrm{tr}(\rho_k)=1. \tag{105}$$

We can now take advantage of the norm inequality (B4) of Appendix B. Because of (105), this gives

$$\left\|C^{(j)}D^{(j)}+E^{(j)}F^{(j)}\right\|_1^2\leq\left\|C^{(j)}\right\|_2^2+\left\|F^{(j)}\right\|_2^2. \tag{106}$$

We plug the equalities (104) into this result to obtain (102). This concludes the proof. □

Let us now turn to unambiguous discrimination. The following easy-to-derive bound generalizes the upper line in (77).

Proposition 5.E.2.[59] *The maximum success probability for unambiguous state discrimination is bounded by*

$$P_{\mathrm{S,u}}^{\mathrm{opt}}(\{\rho_i,\eta_i\})\leq 1-\left(\frac{2m}{m-1}\sum_{i>j}\eta_i\eta_j F(\rho_i,\rho_j)\right)^{\frac{1}{2}}. \tag{107}$$

Proof. The failure probability $P_0=1-P_{\mathrm{S,\,u}}$ satisfies

$$\begin{aligned}P_0^2&=\left(\sum_{i=1}^m\eta_i\,\mathrm{tr}(M_0\rho_i)\right)^2\geq\frac{m}{m-1}\sum_{i\neq j}\eta_i\eta_j\,\mathrm{tr}(M_0\rho_i)\,\mathrm{tr}(M_0\rho_j)\\&\geq\frac{m}{m-1}\sum_{i\neq j}\eta_i\eta_j\left|\mathrm{tr}(U_{ij}\sqrt{\rho_i}M_0\sqrt{\rho_j})\right|^2,\end{aligned} \tag{108}$$

where U_{ij} are arbitrary unitary operators and the first and second bounds follow from the Cauchy-Schwarz inequality. Expressing M_0 as $1-\sum_i M_i$ and using $\mathrm{ran}\,M_i\subset\ker\rho_j$ for $i\neq j$, one gets $\mathrm{tr}(U_{ij}\sqrt{\rho_i}M_0\sqrt{\rho_j})=\mathrm{tr}(U_{ij}\sqrt{\rho_i}\sqrt{\rho_j})$. Maximizing over all unitaries U_{ij} and using the formula $F(\rho_i,\rho_j)=\max_U|\,\mathrm{tr}(U\sqrt{\rho_i}\sqrt{\rho_j})|^2$, one obtains (107). □

One infers from the last two propositions and the Cauchy-Schwarz inequality that

Corollary 5.E.3. *The minimal failure probabilities $P_{\mathrm{err,a}}^{\mathrm{opt}}=1-P_{\mathrm{S,a}}^{\mathrm{opt}}$ and P_0^{opt} for discriminating m states ambiguously and unambiguously satisfy $P_0^{\mathrm{opt}}\geq 2P_{\mathrm{err,a}}^{\mathrm{opt}}/(m-1)$.*

In particular, as noted in Ref. 26, for two states P_0^{opt} is at least twice larger than $P_{\mathrm{err,a}}^{\mathrm{opt}}$.

F. The Holevo bound

Let us come back to the issue of encoding an input message A in an ensemble $\{\rho_i,\eta_i\}$ of quantum states and transmitting it to a receiver. From an information point of view, it makes sense to optimize the measurement in such a way as to maximize the mutual information between the input message A and the output message B reconstructed by the receiver (that is, B is the set of measurement outcomes). This mutual information is defined as[143]

$$I_{A:B}=H(A)+H(B)-H(A,B), \tag{109}$$

where $H(A)=-\sum_i\eta_i\ln\eta_i$ is the Shannon entropy of the input message, $H(B)=-\sum_j p_j\ln p_j$ is the Shannon entropy of the measurement outcomes B with probabilities $p_j=\sum_i\eta_i\,\mathrm{tr}(M_j\rho_i)$,

and $H(A, B) = -\sum_{i,j} p_{ij} \ln p_{ij}$ is the Shannon entropy of the joint process (A, B) with probabilities $p_{ij} = \eta_i p_{j|i} = \eta_i \operatorname{tr}(M_j \rho_i)$, see (63). One can show from the concavity of the logarithm that $I_{A:B} \geq 0$ and $I_{A:B} = 0$ if and only if A and B are independent.

The conditional Shannon entropies are defined by

$$H(B|A) = -\sum_i \eta_i \sum_j p_{j|i} \ln p_{j|i}, \quad H(A|B) = -\sum_j p_j \sum_i \eta_{i|j} \ln \eta_{i|j}, \tag{110}$$

where $p_{j|i} = \operatorname{tr}(M_j \rho_i)$ is the conditional probability of the measurement outcome j given the state ρ_i and $\eta_{i|j}$ the conditional (*a posteriori*) probability that the state is ρ_i given the outcome j. The latter is given by the Bayes rule $\eta_{i|j} = \eta_i p_{j|i}/p_j$. The conditional entropy $H(A|B)$ represents the lack of knowledge of the receiver on the state of the ensemble that was sent to him, after he has performed the measurement. In general, the measurement producing the lowest value of $H(A|B)$ is not a von Neumann measurement.[50] Thanks to the well-known relation $H(A, B) = H(A) + H(B|A) = H(B) + H(A|B)$, the mutual information can be expressed in terms of these conditional entropies as[143]

$$I_{A:B} = H(A) - H(A|B) = H(B) - H(B|A). \tag{111}$$

As $H(A|B) \geq 0$ one has $I_{A:B} \leq H(A)$, with equality if and only if B is a function of A. This means that if $I_{A:B}$ is maximal, i.e., $I_{A:B} = H(A)$, the receiver can reconstruct without any error the message A from his measurement outcomes. As stressed at the beginning of this section, this is never the case if A is encoded using non-orthogonal states ρ_i. Hence $I_{A:B} < H(A)$ for non-orthogonal states. The maximum

$$\max_{\text{POVM } \{M_i\}} \{I_{A:B}\} \tag{112}$$

measures the maximal amount of information accessible to the receiver, that is, how well can he reconstruct the message. The determination of the optimal measurement maximizing $I_{A:B}$ appears to be a more difficult task than the minimization of the probability of error in state discrimination. However, one can place an upper bound on the maximal information (112) by means of the Holevo inequality

$$I_{A:B} \leq \chi_{\text{Holevo}} = S(\rho) - \sum_i \eta_i S(\rho_i), \quad \rho = \sum_i \eta_i \rho_i, \tag{113}$$

where $S(\rho) = -\operatorname{tr}(\rho \ln \rho)$ is the von Neumann entropy of ρ. The proof of this important result relies on the monotonicity of the quantum mutual information under certain quantum operations (see Remark 10.C.3 below). The positive number χ_{Holevo} is called the Holevo quantity. We will show below that $\chi_{\text{Holevo}} \leq H(\{\eta_i\})$ with equality if and only if the ρ_i have orthogonal supports (see (121)). We thus recover the aforementioned fact that for non-orthogonal states ρ_i the maximum (112) is smaller than the entropy $H(A)$ of the input message.

VI. QUANTUM ENTROPIES

In this section we give the definitions and main properties of the von Neumann entropy, the corresponding relative entropy, and the quantum Rényi relative entropies. For classical systems these entropies reduce to the Shannon entropy, the Kullback-Leibler divergence, and the Rényi divergences, respectively, which are central objects in classical information theory. To begin with we recall in Sec. VI A the standard properties of the von Neumann entropy. The most important result for our purpose is the monotonicity of the corresponding relative entropy with respect to quantum operations and the characterization of pairs of states which have the same relative entropy than their transformed states under a given operation. The proof of this result, which will be used later in Sec. X, is given in Sec. VI B. We finally present in Sec. VI C the quantum version of the Rényi divergences introduced recently in Refs. 114, 166, and 60. This quantum version contains as special cases the von Neumann relative entropy and the logarithm of the fidelity (95). The fidelity and the closely related Bures distance will be the subject of Sec. VII. Together with the von Neumann relative entropy, it plays a major role in our geometrical approach of quantum correlations

(Sec. XI). The generalization of this approach to the whole family formed by the relative Rényi entropies constitutes an interesting open problem that will not be deeply explored in this article. The reader may thus skip Sec. VI C in a first reading.

A. The von Neumann entropy

The entropy $H(\{p_k\}) = -\sum_k p_k \ln p_k$ introduced by Shannon in his two celebrated 1948 papers[143] quantifies the amount of information at our disposal on the state of a classical system. It vanishes when the state is perfectly known and takes its maximum value (equal to $\ln n$ if the system has n distinct possible states) when one has no information on this state at all, that is, if all possible states are equiprobable. The quantum analog of the Shannon entropy is the von Neumann entropy

$$S(\rho) = -\operatorname{tr}(\rho \ln \rho). \tag{114}$$

This is a unitary invariant quantity, i.e., $S(U\rho U^*) = S(\rho)$ for U unitary. Moreover, S is additive for composite systems, i.e., $S(\rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}}) = S(\rho_{\mathsf{A}}) + S(\rho_{\mathsf{B}})$ for any states ρ_{A} and ρ_{B} of the systems A and B. Another important property of S is its strictly concavity, i.e., for any states ρ_0, ρ_1, and $0 \le \eta \le 1$ it holds $S((1-\eta)\rho_0 + \eta\rho_1) \ge (1-\eta)S(\rho_0) + \eta S(\rho_1)$, with equality if and only if $\rho_0 = \rho_1$ or $\eta \in \{0, 1\}$. (This comes from the strict convexity of $f(x) = x \ln x$. Actually, it is not hard to prove that if f is strictly convex then the map $\rho \in \mathcal{E}(\mathcal{H}) \mapsto \operatorname{tr}[f(\rho)]$ is strictly convex.[33])

A much less trivial property of importance in quantum information theory is the so-called strong subadditivity

$$S(\rho_{\mathsf{AB}}) + S(\rho_{\mathsf{BC}}) - S(\rho_{\mathsf{ABC}}) - S(\rho_{\mathsf{B}}) \ge 0, \tag{115}$$

where ρ_{ABC} is a state of ABC with marginals $\rho_{\mathsf{AB}} = \operatorname{tr}_{\mathsf{C}}(\rho_{\mathsf{ABC}})$, $\rho_{\mathsf{BC}} = \operatorname{tr}_A(\rho_{\mathsf{ABC}})$, and $\rho_{\mathsf{B}} = \operatorname{tr}_{\mathsf{AC}}(\rho_{\mathsf{ABC}})$. The inequality (115) was first proven by Lieb and Ruskai[99] by using a former work of Lieb[98] on the concavity of the map $\rho \mapsto \operatorname{tr}(K^*\rho^{1+\beta}K\rho^{-\beta})$ for $-1 \le \beta \le 0$ (see Lemma 6.C.2 below). Alternatively, (115) is a direct consequence of the monotonicity of the relative entropy (Theorem 6.B.1 below), which can be established by other means than Lieb's concavity theorem. Choosing $\mathcal{H}_{\mathsf{B}} = \mathbb{C}$, the strong subadditivity (115) implies that S is subadditive, i.e., $S(\rho_{\mathsf{AC}}) \le S(\rho_{\mathsf{A}}) + S(\rho_{\mathsf{C}})$.

As is well know in statistical physics, the von Neumann entropy $S(\rho)$ is the Legendre transform of the free energy $\Phi(\beta, H) = -\beta^{-1} \ln \operatorname{tr}(e^{-\beta H})$. More precisely, one has (see Ref. 33, Theorem 2.13)

$$S(\rho) = \inf_{H \in \mathcal{B}(\mathcal{H})_{\text{s.a.}}} \left\{\beta \operatorname{tr}(H\rho) - \beta\Phi(\beta, H)\right\}, \quad \Phi(\beta, H) = \inf_{\rho \in \mathcal{E}(\mathcal{H})} \left\{\operatorname{tr}(H\rho) - \beta^{-1}S(\rho)\right\}, \tag{116}$$

and the last infimum is attained if and only if ρ is the Gibbs state $\rho_\beta = e^{-\beta H}/\operatorname{tr}(e^{-\beta H})$. The free energy is a concave function of the energy observable H.

The following identity will be used repeatedly in Secs. IX and X:

$$S(\rho_{\mathsf{A}}) = S(\rho_{\mathsf{B}}) \quad \text{if } \rho_{\mathsf{A}} \text{ and } \rho_{\mathsf{B}} \text{ are the reduced states of the pure state } |\Psi_{\mathsf{AB}}\rangle \text{ of } \mathsf{AB}. \tag{117}$$

It is a consequence of Theorem 2.B.1, since if $|\Psi_{\mathsf{AB}}\rangle$ has Schmidt coefficients μ_i then $S(\rho_{\mathsf{A}}) = S(\rho_{\mathsf{B}}) = -\sum_i \mu_i \ln \mu_i$.

A last identity worthwhile mentioning here is

$$S(\rho) = \min_{\{|\psi_i\rangle, \eta_i\}} H(\{\eta_i\}) = \min_{\{|\psi_i\rangle, \eta_i\}} \left\{-\sum_{i=1}^m \eta_i \ln \eta_i\right\}, \tag{118}$$

where the minimum is over all pure state decompositions of ρ. Furthermore, a decomposition minimizes $H(\{\eta_i\})$ if and only if it is a spectral decomposition of ρ. These statements can be justified as follows (an alternative proof can be found in Ref. 117). Let $\{|k\rangle, p_k\}_{i=1}^r$ be a spectral decomposition of ρ, with $r = \operatorname{ran}(\rho)$. An arbitrary pure state decomposition $\{|\psi_i\rangle, \eta_i\}_{i=1}^m$ of ρ has the form $\sqrt{\eta_i}|\psi_i\rangle = \sum_k u_{ik}\sqrt{p_k}|k\rangle$, where (u_{ik}) is a $m \times m$ unitary matrix and $m \ge r$ (see (16)). Setting $p_k = 0$ for $r < k \le m$ one gets $\eta_i = \sum_k |u_{ik}|^2 p_k$. Since $f(x) = x \ln x$ is strictly convex, one

finds

$$-H(\{\eta_i\}) = \sum_{i=1}^{m} \eta_i \ln \eta_i \leq \sum_{i,k=1}^{m} |u_{ik}|^2 p_k \ln p_k = \sum_{k=1}^{r} p_k \ln p_k = -S(\rho), \tag{119}$$

so that $S(\rho) \leq H(\{\eta_i\})$. By strict convexity, the inequality in (119) is an equality if and only if for any i, there exists some $k_i \in \{1, \ldots, r+1\}$ such that $u_{ik} = 0$ when $k \notin I_i = \{k = 1, \ldots, m; p_k = p_{k_i}\}$. Thus $S(\rho) = H(\{\eta_i\})$ if and only if

$$\sqrt{\eta_i}|\psi_i\rangle = \sqrt{p_{k_i}} \sum_{k \in I_i} u_{ik}|k\rangle \tag{120}$$

are eigenvectors of ρ with eigenvalue $\eta_i = p_{k_i}$ (if $p_{k_i} \neq 0$). It remains to check that $\langle\psi_i|\psi_j\rangle = 0$ when $p_{k_i} = p_{k_j} \neq 0$. This comes from the unitarity of (u_{ik}). This yields the desired result. The inequality (119) can be easily generalized to get

$$S(\rho) \leq H(\{\eta_i\}) + \sum_i \eta_i S(\rho_i) \tag{121}$$

for any ensemble $\{\rho_i, \eta_i\}$ forming a convex decomposition of ρ. (This follows from (118) by writing the spectral decompositions of the ρ_i (see Ref. 117, Sec. 11.3).) Moreover, one has equality if and only if the ρ_i have orthogonal supports.

B. Relative entropy

A related quantity to the von Neumann entropy is the relative entropy introduced by Umegaki[158] and later extended by Araki[10] in the von Neumann algebra setting,

$$S(\rho||\sigma) = \begin{cases} \mathrm{tr}\big(\rho(\ln\rho - \ln\sigma)\big) & \text{if } \ker(\sigma) \subset \ker(\rho) \\ +\infty & \text{otherwise.} \end{cases} \tag{122}$$

Note that by taking $\sigma = 1/n$ proportional to the identity operator, $S(\rho||1/n) = \ln n - S(\rho)$ is the difference between the maximal and the von Neumann entropy of ρ. The relative entropy has the following properties:

(i) $S(\rho||\sigma) \geq 0$ with equality if and only if $\rho = \sigma$;
(ii) unitary invariance $S(U\rho U^*||U\sigma U^*) = S(\rho||\sigma)$ for any unitary U;
(iii) additivity for composite systems: $S(\rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}}||\sigma_{\mathsf{A}} \otimes \sigma_{\mathsf{B}}) = S(\rho_{\mathsf{A}}||\sigma_{\mathsf{A}}) + S(\rho_{\mathsf{B}}||\sigma_{\mathsf{B}})$;
(iv) joint convexity: if $0 \leq \eta \leq 1$ then $S((1-\eta)\rho_0 + \eta\rho_1||(1-\eta)\sigma_0 + \eta\sigma_1) \leq (1-\eta)S(\rho_0||\sigma_0) + \eta S(\rho_1||\sigma_1)$.

The first property (i) follows from Klein's inequality, which states that if f is continuous and strictly convex, then $\mathrm{tr}[f(A) - f(B) - (A-B)f'(B)] \geq 0$, with equality if and only if $A = B$. Its proof can be found, for instance, in the excellent lecture notes of Carlen.[33] The properties (ii) and (iii) are immediate consequences of the cyclicity of the trace and the relation $\ln(\rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}}) = \ln\rho_{\mathsf{A}} \otimes 1 + 1 \otimes \ln\rho_{\mathsf{B}}$, as in the case of the von Neumann entropy. The last property (iv) can be deduced from the strong subadditivity (115).[101,102] It will be proven in Sec. VI C. Let us point out that (i) implies the aforementioned subadditivity $S(\rho_{\mathsf{AC}}) \leq S(\rho_{\mathsf{A}}) + S(\rho_{\mathsf{C}})$ of the von Neumann entropy, with equality if and only if $\rho_{\mathsf{AC}} = \rho_{\mathsf{A}} \otimes \rho_{\mathsf{C}}$ is a product state (in fact, $S(\rho_{\mathsf{AC}}||\rho_{\mathsf{A}} \otimes \rho_{\mathsf{C}}) = S(\rho_{\mathsf{A}}) + S(\rho_{\mathsf{C}}) - S(\rho_{\mathsf{AC}})$).

Another fundamental property of $S(\rho||\sigma)$ is its monotonicity with respect to CP trace-preserving mappings. This monotonicity means that if one performs the same measurement on two states without readout of the outcomes, the pair of post-measurement states has a lower relative entropy than the pair of states before the measurement. This fact was first proven by Lindblad[102] (see also Refs. 10 and 155). Notice that unlike the relative entropy, the von Neumann entropy is not monotonous with respect to non-projective measurements (see Ref. 117, Exercise 11.15). The following theorem provides a necessary and sufficient condition on the two states such that the monotonicity of the relative entropy is satisfied with equality. It is due to Petz.[129]

Theorem 6.B.1. (Monotonicity of the relative entropy[129,73]) *For any quantum operation* $\mathcal{M}: \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ *one has* $S(\rho||\sigma) \geq S(\mathcal{M}(\rho)||\mathcal{M}(\sigma))$ *for all states* $\rho, \sigma \in \mathcal{E}(\mathcal{H})$. *The inequality is an equality if and only if there exists a quantum operation* $\mathcal{R}: \mathcal{B}(\mathcal{H}') \to \mathcal{B}(\mathcal{H})$ *such that* $\mathcal{R} \circ \mathcal{M}(\sigma) = \sigma$ *and* $\mathcal{R} \circ \mathcal{M}(\rho) = \rho$. *This quantum operation is the transpose operation* $\mathcal{R} = \mathcal{R}_{\mathcal{M},\sigma}$ *defined in (45).*

Let us recall from Sec. IV A that the transpose operation $\mathcal{R}_{\mathcal{M},\sigma}$ is the quantum operation with Kraus operators

$$R_i = \sqrt{\sigma} A_i^* \mathcal{M}(\sigma)^{-1/2}\,, \tag{123}$$

where $\{A_i\}$ are some Kraus operators for $\mathcal{M}$. The conditions $\mathcal{R} \circ \mathcal{M}(\sigma) = \sigma$ and $\mathcal{R} \circ \mathcal{M}(\rho) = \rho$, which mean that ρ and σ can be recovered, respectively, from $\mathcal{M}(\rho)$ and $\mathcal{M}(\sigma)$ by means of the same quantum operation $\mathcal{R}$, is clearly sufficient to ensure the equality $S(\rho||\sigma) = S(\mathcal{M}(\rho)||\mathcal{M}(\sigma))$ if monotonicity holds true. It is remarkable that this is also a necessary condition, with $\mathcal{R} = \mathcal{R}_{\mathcal{M},\sigma}$ the approximate reversal of $\mathcal{M}$ introduced in the context of quantum error correction (Sec. IV).

We present below the derivation of this result given by Petz in Ref. 129, which also provides a nice and simple proof of the monotonicity. A completely different proof of the monotonicity, based on Lieb's concavity theorem as in Ref. 102, 33 and 60, will be given in Sec. VI C in the more general setting of the Rényi entropies. It is noteworthy that Petz's derivation does neither rely on the Stinespring theorem nor on the Kraus decomposition (albeit it takes advantage of one of its consequence, namely the Kadison-Schwarz inequality). It makes use of the theory of operator convex functions and of Araki's relative modular operators.[11] Let $\mathcal{M}$ be a quantum operation $\mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ and ρ and σ be two states of $\mathcal{E}(\mathcal{H})$ such that ρ and $\mathcal{M}(\rho)$ are invertible. One can define two relative modular operators by (see Sec. II)

$$\Delta_{\sigma|\rho}(B) = \sigma B \rho^{-1}, \quad \Delta_{\mathcal{M}(\sigma)|\mathcal{M}(\rho)}(B') = \mathcal{M}(\sigma) B' \mathcal{M}(\rho)^{-1}, \quad B \in \mathcal{B}(\mathcal{H})\,,\ B' \in \mathcal{B}(\mathcal{H}'). \tag{124}$$

Proof. Let us set $\rho_{\mathcal{M}} = \mathcal{M}(\rho)$ and $\sigma_{\mathcal{M}} = \mathcal{M}(\sigma)$ and assume that $\rho, \sigma, \rho_{\mathcal{M}}$, and $\sigma_{\mathcal{M}}$ are invertible. In the whole proof these states are fixed, so to simplify notation we write Δ instead of $\Delta_{\sigma|\rho}$ and $\Delta_{\mathcal{M}}$ instead of $\Delta_{\sigma_{\mathcal{M}}|\rho_{\mathcal{M}}}$. We set $\xi = \rho^{\frac{1}{2}}$ and $\xi_{\mathcal{M}} = \rho_{\mathcal{M}}^{\frac{1}{2}}$. One can view these two operators as unit vectors in $\mathcal{B}(\mathcal{H})$ and $\mathcal{B}(\mathcal{H}')$, respectively, for the Hilbert-Schmidt scalar product $\langle \cdot\,,\, \cdot \rangle$. The first observation is that

$$S(\rho||\sigma) = \langle \xi,\, (\ln\rho - \ln\sigma)\xi\rangle = -\langle \xi,\, \ln(\Delta)\xi\rangle = \int_0^\infty \mathrm{d}t \left(\langle \xi,\, (\Delta + t)^{-1}\xi\rangle - (1+t)^{-1} \right). \tag{125}$$

The third equality can be established, for instance, with the help of the first identity in (A2) (see Appendix A). Therefore, in order to prove that $S(\rho||\sigma) \geq S(\rho_{\mathcal{M}}||\sigma_{\mathcal{M}})$, it suffices to show that for any $t > 0$,

$$\langle \xi_{\mathcal{M}}\,,\, (\Delta_{\mathcal{M}} + t)^{-1}\xi_{\mathcal{M}}\rangle \leq \langle \xi,\, (\Delta + t)^{-1}\xi\rangle\,. \tag{126}$$

To this end, let us consider the operator $\mathcal{C}_{\mathcal{M}}$ defined by

$$\mathcal{C}_{\mathcal{M}}(B'\xi_{\mathcal{M}}) = \mathcal{M}^*(B')\xi, \quad B \in \mathcal{B}(\mathcal{H}')\,. \tag{127}$$

Note that $\{B'\xi_{\mathcal{M}}; B' \in \mathcal{B}(\mathcal{H}')\}$ is equal to $\mathcal{B}(\mathcal{H}')$ by the invertibility of $\rho_{\mathcal{M}}$. In the theory of C*-algebras, if this equality is true upon completion of $\{B'\xi_{\mathcal{M}}; B' \in \mathcal{B}'\}$ for the Hilbert-Schmidt norm one says that $(B' \in \mathcal{B}' \mapsto \mathcal{L}_{B'}, \xi_{\mathcal{M}})$ defines a cyclic representation of the algebra $\mathcal{B}'$ on the Hilbert space $\mathcal{B}(\mathcal{H}')$.[29] Hence, (128) defines an operator $\mathcal{C}_{\mathcal{M}}$ from $\mathcal{B}(\mathcal{H}')$ to $\mathcal{B}(\mathcal{H})$. Then

$$\mathcal{C}_{\mathcal{M}}^* \Delta \mathcal{C}_{\mathcal{M}} \leq \Delta_{\mathcal{M}}. \tag{128}$$

Actually, thanks to the Kadison-Schwarz inequality (30) and the relation $(\mathcal{M}^*(B'^*))^* = \mathcal{M}^*(B')$, one has

$$\begin{aligned} \langle \mathcal{C}_{\mathcal{M}}(B'\xi_{\mathcal{M}}),\ \Delta\, \mathcal{C}_{\mathcal{M}}(B'\xi_{\mathcal{M}})\rangle &= \mathrm{tr}\Big(|\mathcal{M}^*(B'^*)|^2 \sigma \Big) \\ &\leq \mathrm{tr}\Big(\mathcal{M}^*(B'B'^*)\sigma \Big) = \langle B'\xi_{\mathcal{M}}\,,\ \Delta_{\mathcal{M}} B'\xi_{\mathcal{M}}\rangle. \end{aligned} \tag{129}$$

One shows similarly that $\|\mathcal{C}_{\mathcal{M}}(B'\xi_{\mathcal{M}})\|_2 \leq \|B'\xi_{\mathcal{M}}\|_2$ for any $B' \in \mathcal{B}(\mathcal{H}')$, hence $\|\mathcal{C}_{\mathcal{M}}\| \leq 1$.

We now use the fact that the function $f(x) = (x + t)^{-1}$ is operator monotone-decreasing and operator convex. The definitions of operator monotone and operator convex functions are given in Appendix A. Together with the bound (128), this implies

$$(\Delta_{\mathcal{M}} + t)^{-1} \leq (\mathcal{C}_{\mathcal{M}}^* \Delta \mathcal{C}_{\mathcal{M}} + t)^{-1} \leq \mathcal{C}_{\mathcal{M}}^* (\Delta + t)^{-1} \mathcal{C}_{\mathcal{M}} + t^{-1}(1 - \mathcal{C}_{\mathcal{M}}^* \mathcal{C}_{\mathcal{M}}). \tag{130}$$

The last inequality follows by applying the Jensen-type inequality (A4) for the operator convex function $g(x) = (x + t)^{-1} - t^{-1}$ satisfying $g(0) = 0$ and the contraction $\mathcal{C}_{\mathcal{M}}$. Since $\mathcal{C}_{\mathcal{M}}(\xi_{\mathcal{M}}) = \xi$ by (127) and $\mathcal{M}^*(1) = 1$, the inequality (130) entails

$$\left\langle \xi_{\mathcal{M}} \,,\, (\Delta_{\mathcal{M}} + t)^{-1} \xi_{\mathcal{M}} \right\rangle \leq \left\langle \xi,\, (\Delta + t)^{-1} \xi \right\rangle + t^{-1} \big(\mathrm{tr}(\rho_{\mathcal{M}}) - \mathrm{tr}(\rho) \big). \tag{131}$$

The term proportional to t^{-1} vanishes because $\mathcal{M}$ is trace preserving, hence one obtains the desired bound (126). We have thus proven the monotonicity of the relative entropy.

In addition to its simplicity, the above proof offers the advantage that it easily yields a necessary and sufficient condition for having $S(\rho||\sigma) = S(\rho_{\mathcal{M}}||\sigma_{\mathcal{M}})$. Actually, this equality holds if and only if (126) is an equality, i.e.,

$$\left\langle \xi_{\mathcal{M}} \,,\, (\Delta_{\mathcal{M}} + t)^{-1} \xi_{\mathcal{M}} \right\rangle = \left\langle \xi_{\mathcal{M}} \,,\, \big(\mathcal{C}_{\mathcal{M}}^* (\Delta + t)^{-1} \mathcal{C}_{\mathcal{M}} + t^{-1}(1 - \mathcal{C}_{\mathcal{M}}^* \mathcal{C}_{\mathcal{M}}) \big) \xi_{\mathcal{M}} \right\rangle \tag{132}$$

for all $t > 0$. But for any operators X, Y, and Z with Z invertible and $X \leq Y$, $\langle Z, XZ\rangle = \langle Z, YZ\rangle$ implies $XZ = YZ$. Hence, we can infer from (130) and (132) that

$$(\Delta_{\mathcal{M}} + t)^{-1} \xi_{\mathcal{M}} = \mathcal{C}_{\mathcal{M}}^* (\Delta + t)^{-1} \xi \,, \quad t > 0, \tag{133}$$

where we have used the identity $\mathcal{C}_{\mathcal{M}}^* \mathcal{C}_{\mathcal{M}}(\xi_{\mathcal{M}}) = \xi_{\mathcal{M}}$ (in fact, one easily finds that $\langle \mathcal{C}_{\mathcal{M}}(B' \xi_{\mathcal{M}}) \,,\, \mathcal{C}_{\mathcal{M}} \xi_{\mathcal{M}} \rangle = \langle B' \xi_{\mathcal{M}} \,,\, \xi_{\mathcal{M}} \rangle$ for any $B' \in \mathcal{B}(\mathcal{H}')$). Therefore,

$$\left\| \mathcal{C}_{\mathcal{M}}^* (\Delta + t)^{-1} \xi \right\|_2^2 = \left\langle (\Delta_{\mathcal{M}} + t)^{-2} \xi_{\mathcal{M}}, \xi_{\mathcal{M}} \right\rangle = \left\langle \mathcal{C}_{\mathcal{M}}^* (\Delta + t)^{-2} \xi, \xi_{\mathcal{M}} \right\rangle = \left\| (\Delta + t)^{-1} \xi \right\|_2^2 , \tag{134}$$

where the second equality is obtained by differentiating (133) with respect to t. Now, the identity $\|\mathcal{C}^*(X)\|_2 = \|X\|_2$ for $\mathcal{C}$ a contraction implies that $\mathcal{C}\mathcal{C}^*(X) = X$ (in fact, then the Cauchy-Schwarz inequality $\langle X, \mathcal{C}\mathcal{C}^*(X)\rangle \leq \|X\|_2 \|\mathcal{C}\mathcal{C}^*(X)\|_2 \leq \|X\|_2^2$ is an equality, so that $\mathcal{C}\mathcal{C}^*(X)$ must be proportional to X). We conclude that

$$\mathcal{C}_{\mathcal{M}} (\Delta_{\mathcal{M}} + t)^{-1} \xi_{\mathcal{M}} = \mathcal{C}_{\mathcal{M}} \mathcal{C}_{\mathcal{M}}^* (\Delta + t)^{-1} \xi = (\Delta + t)^{-1} \xi \tag{135}$$

for any $t > 0$. By means of the functional calculus, one deduces from this identity that

$$\mathcal{C}_{\mathcal{M}} \Delta_{\mathcal{M}}^{-\frac{1}{2}} \xi_{\mathcal{M}} = \Delta^{-\frac{1}{2}} \xi \,. \tag{136}$$

In view of the definitions (124) and (127) and the invertibility of ρ, the last formula gives $\mathcal{M}^*(\sigma_{\mathcal{M}}^{-\frac{1}{2}} \xi_{\mathcal{M}}) = \sigma^{-\frac{1}{2}} \xi$. By multiplying by the adjoint and using the Kadison-Schwarz inequality, we arrive at

$$\sigma^{-\frac{1}{2}} \rho\, \sigma^{-\frac{1}{2}} \leq \mathcal{M}^*(\sigma_{\mathcal{M}}^{-\frac{1}{2}} \rho_{\mathcal{M}} \sigma_{\mathcal{M}}^{-\frac{1}{2}}) \,, \tag{137}$$

that is, $\rho \leq \mathcal{R}_{\mathcal{M},\sigma}(\rho_{\mathcal{M}})$ with $\mathcal{R}_{\mathcal{M},\sigma}$ defined in (45). But $\mathrm{tr}[\rho] = \mathrm{tr}[\rho_{\mathcal{M}}] = \mathrm{tr}[\mathcal{R}_{\mathcal{M},\sigma}(\rho_{\mathcal{M}})]$, whence $\rho = \mathcal{R}_{\mathcal{M},\sigma}(\rho_{\mathcal{M}})$. The other equality $\sigma = \mathcal{R}_{\mathcal{M},\sigma}(\sigma_{\mathcal{M}})$ is obvious. Reciprocally, as stressed above, these two identities imply $S(\rho||\sigma) = S(\rho_{\mathcal{M}}||\sigma_{\mathcal{M}})$ thanks to the monotonicity of the relative entropy and the fact that $\mathcal{R}_{\mathcal{M},\sigma}$ is a quantum operation. □

Let us end this subsection by pointing out that the strong subadditivity of the von Neumann entropy, the joint convexity of the relative entropy, and its monotonicity can be deduced from each other. For instance, the strong subadditivity (115) is a simple consequence of the monotonicity. Actually, one checks that

$$S(\rho_{\mathsf{AB}}) + S(\rho_{\mathsf{BC}}) - S(\rho_{\mathsf{ABC}}) - S(\rho_{\mathsf{B}}) = S(\rho_{\mathsf{ABC}}||\rho_{\mathsf{A}} \otimes \rho_{\mathsf{BC}}) - S(\mathcal{M}_C(\rho_{\mathsf{ABC}})||\mathcal{M}_C(\rho_{\mathsf{A}} \otimes \rho_{\mathsf{BC}})) \tag{138}$$

with $\mathcal{M}_{\mathsf{C}} : \rho \mapsto \mathrm{tr}_{\mathsf{C}}(\rho)$. It is easy to show that $\mathcal{M}_{\mathsf{C}}$ is a CP and trace-preserving map $\mathcal{B}(\mathcal{H}_{\mathsf{ABC}}) \to \mathcal{B}(\mathcal{H}_{\mathsf{AB}})$, therefore (115) follows from Theorem 6.B.1. With the help of this theorem it is also possible to characterize all states ρ_{ABC} such that (115) becomes an equality.[73]

Conversely, Lindblad[101,102] proves the monotonicity inequality from the strong subadditivity. The basic idea is to show that the strong subadditivity of the von Neumann entropy or the closely related Lieb concavity theorem imply the joint convexity (iv) of the relative entropy. The corresponding arguments are given in Sec. VI C 2 below. One can then deduce the monotonicity of the relative entropy from its joint convexity (iv) with the help of Stinespring's theorem as follows.[153,167,60] Recall that if μ_H is the normalized Haar measure on the group $U(n)$ of $n \times n$ unitary matrices, then $\int \mathrm{d}\mu_H(U)\, UBU^* = n^{-1}\,\mathrm{tr}(B)$ for any $B \in \mathcal{B}(\mathcal{H})$ (in fact, all diagonal matrix elements of the left-hand side in an arbitrary basis are equal, as follows from the left-invariance $\mathrm{d}\mu_H(VU) = \mathrm{d}\mu_H(U)$ for $V \in U(n)$; as a result, this left-hand side is proportional to the identity matrix). We infer from Stinespring theorem 3.B.2 that

$$\mathcal{M}(\rho)\otimes(1/n_{\mathsf{E}}) = \int_{U(n_{\mathsf{E}})} \mathrm{d}\mu_H(U_{\mathsf{E}})\,(1\otimes U_{\mathsf{E}})U\rho\otimes|\epsilon_0\rangle\langle\epsilon_0|U^*(1\otimes U_{\mathsf{E}}^*) \tag{139}$$

with U unitary on $\mathcal{H}_{\mathsf{SE}}$. Thanks to the additivity (iii), the joint convexity convexity (iv), and the unitary invariance (ii), we get

$$\begin{aligned} S(\mathcal{M}(\rho)||\mathcal{M}(\sigma)) &= S\big(\mathcal{M}(\rho)\otimes(1/n_{\mathsf{E}})\big|\big|\mathcal{M}(\sigma)\otimes(1/n_{\mathsf{E}})\big) \\ &\le \int_{U(n_{\mathsf{E}})} \mathrm{d}\mu_H(U_{\mathsf{E}})S\big((1\otimes U_{\mathsf{E}})U\rho\otimes|\epsilon_0\rangle\langle\epsilon_0|U^*(1\otimes U_{\mathsf{E}}^*)\big|\big|(1\otimes U_{\mathsf{E}})U\sigma\otimes|\epsilon_0\rangle\langle\epsilon_0|U^*(1\otimes U_{\mathsf{E}}^*)\big) \\ &= \int_{U(n_{\mathsf{E}})} \mathrm{d}\mu_H(U_{\mathsf{E}})S(\rho||\sigma) = S(\rho||\sigma)\,. \end{aligned} \tag{140}$$

By the same argument, one can show a slightly more general result.

Proposition 6.B.2. *Let $f : \mathcal{E}(\mathcal{H})\times\mathcal{E}(\mathcal{H})\to\mathbb{R}$ be a unitary-invariant jointly convex function for any finite Hilbert space $\mathcal{H}$, which satisfies $f(\rho\otimes\tau,\sigma\otimes\tau) = f(\rho,\sigma)$ for all $\rho,\sigma\in\mathcal{E}(\mathcal{H})$, and $\tau\in\mathcal{E}(\mathcal{H}')$. Then f is monotonous with respect to quantum operations.*

C. Quantum relative Rényi entropies

1. Definitions

In the classical theory of information, other entropies than the Shannon entropy play a role when ergodicity breaks down or outside the asymptotic regime. The Rényi entropy depending on a parameter $\alpha > 0$ unifies these different entropies. In the quantum setting, it is defined as

$$S_\alpha(\rho) = (1-\alpha)^{-1}\ln\mathrm{tr}(\rho^\alpha). \tag{141}$$

It is easy to show that $S_\alpha(\rho)$ converges to the von Neumann entropy $S(\rho)$ when $\alpha\to 1$ and that $S_\alpha(\rho)$ is a non-increasing function of α.

A first definition of the quantum relative Réyni entropy is

$$S_\alpha^{(\mathrm{n})}(\rho||\sigma) = (\alpha-1)^{-1}\ln(\mathrm{tr}[\rho^\alpha\sigma^{1-\alpha}]),\quad \alpha>0\,,\ \alpha\neq 1. \tag{142}$$

This entropy appears naturally in the context of the quantum hypothesis testing (Sec. VIII A below). We shall discuss here a symmetrized version proposed recently by Müller-Lennert *et al.*[114] and by Wilde, Winter, and Yang.[166] It is given by

$$S_\alpha(\rho||\sigma) = (\alpha-1)^{-1}\ln\mathrm{tr}\Big[\big(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\,\sigma^{\frac{1-\alpha}{2\alpha}}\big)^\alpha\Big] \tag{143}$$

if $\alpha\in(0,1)$ and $\mathrm{tr}(\sigma\rho)>0$ or if $\alpha>1$ and $\ker\sigma\subset\ker\rho$ (if none of these conditions are satisfied, one sets $S_\alpha(\rho||\sigma) = +\infty$). This relative entropy has been used in Ref. 166 to solve an important open problem related to the transmission of information in noisy quantum channels. It seems likely that much more applications in quantum information theory will be encountered in the future. The entropies S_α appeared recently as central objects in a very different context, namely, the quantum fluctuation relations in out-of-equilibrium statistical physics.[88,89] A nice feature of the family $\{S_\alpha\}_{\alpha>0}$ is that it contains the von Neumann relative entropy, the fidelity entropy, and the

max-entropy as special cases. Furthermore, S_α depends continuously and monotonously on α. The fidelity-entropy is obtained for $\alpha = 1/2$. It is given by $S_{1/2}(\rho||\sigma) = -\ln F(\rho,\sigma)$, where $F(\rho,\sigma)$ is the fidelity (95). The max-entropy is defined by

$$S_\infty(\rho||\sigma) = \lim_{\alpha\to\infty} S_\alpha(\rho||\sigma) = \ln \|\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\| \,, \tag{144}$$

where $\|\cdot\|$ is the operator norm. The second equality follows from $\|A\|_\alpha \to \|A\|$ as $\alpha\to\infty$ (see Sec. II A). Finally, one recovers the von Neumann relative entropy (122) by letting $\alpha \to 1$,

$$S(\rho||\sigma) = \lim_{\alpha\to 1} S_\alpha(\rho||\sigma). \tag{145}$$

To justify this statement, let us set $A(\alpha) = \sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}$. Explicit calculations show that

$$\begin{aligned} \frac{\mathrm{d}\,\mathrm{tr}[A(\alpha)^\alpha]}{\mathrm{d}\alpha} &= \mathrm{tr}[A(\alpha)^\alpha \ln A(\alpha)] + \alpha\,\mathrm{tr}\Big[A(\alpha)^{\alpha-1}\frac{\mathrm{d}A}{\mathrm{d}\alpha}\Big] \\ \frac{\mathrm{d}A}{\mathrm{d}\alpha} &= -\frac{1}{2\alpha^2}\Big(\ln(\sigma)A(\alpha) + A(\alpha)\ln(\sigma)\Big). \end{aligned} \tag{146}$$

Consequently, $S_\alpha(\rho||\sigma) \to (\mathrm{d}\ln\mathrm{tr}[A(\alpha)^\alpha]/\mathrm{d}\alpha)_{\alpha=1} = \mathrm{tr}(\rho\ln\rho - \rho\ln\sigma)$ as $\alpha\to 1$. Note that a similar result holds for the unsymmetrized Rényi entropy (142), i.e., $S(\rho||\sigma) = \lim_{\alpha\to 1} S_\alpha^{(\mathrm{n})}(\rho||\sigma)$. Let us also emphasize that

$$S_\alpha(\rho||\sigma) \le S_\alpha^{(\mathrm{n})}(\rho||\sigma) \tag{147}$$

by the Lieb-Thirring trace inequality (B3).

For commuting matrices $\rho = \sum p_k|k\rangle\langle k|$ and $\sigma = \sum_k q_k|k\rangle\langle k|$, both $S_\alpha(\rho||\sigma)$ and $S_\alpha^{(\mathrm{n})}(\rho||\sigma)$ reduce to the classical Réyni divergence

$$S_\alpha^{\mathrm{clas}}(\mathbf{p}||\mathbf{q}) = (\alpha-1)^{-1}\ln\Big(\sum_{k=1}^n p_k^\alpha q_k^{1-\alpha}\Big) \,, \tag{148}$$

which is non-negative for $\alpha > 0$ by the Hölder inequality.

2. Main properties

It is shown in this subsection that the Rényi relative entropy $S_\alpha(\rho||\sigma)$ satisfies the same properties (i-iv) as the von Neumann relative entropy in Sec. VI B for any $\alpha \in [1/2, 1]$. For $0 < \alpha < \infty$ we define the α-fidelity by

$$F_\alpha(\rho||\sigma) = \|\rho^{\frac{1}{2}}\sigma^{\frac{\beta}{2}}\|_{2\alpha}^2 = \|\sigma^{\frac{\beta}{2}}\rho\sigma^{\frac{\beta}{2}}\|_\alpha = e^{-\beta S_\alpha(\rho||\sigma)} \quad \text{with} \quad \beta = \frac{1-\alpha}{\alpha} \,. \tag{149}$$

Here, we have used the notation $\|A\|_{2\alpha} = (\mathrm{tr}[(A^*A)^\alpha])^{\frac{1}{2\alpha}}$ even if this does not correspond to a norm when $0 < \alpha < 1/2$.

Theorem 6.C.1. *For any $\alpha > 0$, one has*

(i) $S_\alpha(\rho||\sigma) \ge 0$ *with equality if and only if* $\rho = \sigma$;
(ii) $S_\alpha(\rho||\sigma)$ *is unitary invariant*;
(iii) $S_\alpha(\rho||\sigma)$ *is additive for composite systems*;
(iv) $F_\alpha(\rho||\sigma)^\alpha$ *is jointly concave for* $\alpha \in [1/2, 1)$ *and jointly convex for* $\alpha > 1$. *In particular,* $S_\alpha(\rho||\sigma)$ *is jointly convex for* $\alpha \in [1/2, 1]$;
(v) *if* $\alpha \ge 1/2$ *then* $S_\alpha(\rho||\sigma) \ge S_\alpha(\mathcal{M}(\rho)||\mathcal{M}(\sigma))$ *for any quantum operation* $\mathcal{M}$ on $\mathcal{B}(\mathcal{H})$.

The statements (i-iii), as well as (iv-v) for a restricted range of α, namely $\alpha \in (1, 2]$, have been established in Refs. 114 and 166. The justification of (iv-v) in full generality is due to Frank and Lieb.[60]

Proof. The unitary invariance (ii) and additivity (iii) are evident and also hold for the α-fidelity. We now argue that the non-negativity (i) and the monotonicity (iv) can be deduced from the convexity/concavity property (iv). Thanks to Proposition 6.B.2, (iv) implies that if $\alpha \in [1/2, 1)$ then $F_\alpha(\mathcal{M}(\rho)||\mathcal{M}(\sigma)) \geq F_\alpha(\rho||\sigma)$ for any quantum operation $\mathcal{M}$, and the reverse inequality holds true if $\alpha > 1$. The monotonicity of S_α for $\alpha \geq 1/2$ then follows immediately (the case $\alpha = 1$ is obtained by continuity, see (145)). Let $\{|k\rangle\}$ be an orthonormal basis of $\mathcal{H}$ and $\mathcal{M}_\Pi$ be the quantum operation (26) associated to the von Neumann measurement $\{\Pi_k = |k\rangle\langle k|\}$. The monotonicity entails

$$S_\alpha(\rho||\sigma) \geq S_\alpha(\mathcal{M}_\Pi(\rho)||\mathcal{M}_\Pi(\sigma)) = S_\alpha^{\mathrm{clas}}(\mathbf{p}||\mathbf{q}), \tag{150}$$

where $\mathbf{p}$ and $\mathbf{q}$ are the vectors with components $p_k = \langle k|\rho|k\rangle$ and $q_k = \langle k|\sigma|k\rangle$. Since the classical Rényi divergence (148) is non-negative and vanishes if and only if $\mathbf{p} = \mathbf{q}$, we deduce from (150) that $S_\alpha(\rho||\sigma) \geq 0$, with equality if and only if $\langle k|\rho|k\rangle = \langle k|\sigma|k\rangle$ for all k. The orthonormal basis $\{|k\rangle\}$ being arbitrary, this justifies the assertion (i) for $\alpha \geq 1/2$. To show this assertion for $\alpha \in (0, 1/2)$, we argue as in Ref. 114 that

$$S_\alpha(\rho||\sigma) \geq S_\alpha(\mathcal{M}_\Pi(\rho)||\sigma) = S_\alpha^{\mathrm{clas}}(\mathbf{p}||\mathbf{q}) \tag{151}$$

with $0 < \alpha < 1$, $\mathcal{M}_\Pi$ being as before associated with the von Neumann $\{\Pi_k = |k\rangle\langle k|\}$ but with $\{|k\rangle\}$ an orthonormal eigenbasis of σ. Actually, let $\alpha \in (0, 1)$ and let us set $A(\beta) = \sigma^{\frac{\beta}{2}}\rho\sigma^{\frac{\beta}{2}}$ with $\beta = \alpha^{-1} - 1$. By virtue of the Jensen type inequality (A8) of Appendix A, one has

$$\big(\mathcal{M}_\Pi(A(\beta))\big)^\alpha \geq \mathcal{M}_\Pi\big(A(\beta)^\alpha\big) \tag{152}$$

due to the operator concavity of $f(x) = x^\alpha$. Hence, by the trace-preserving property of $\mathcal{M}_\Pi$ and the identity $\sigma^{\frac{\beta}{2}}\mathcal{M}_\Pi(\rho)\sigma^{\frac{\beta}{2}} = \mathcal{M}_\Pi(A(\beta))$,

$$\begin{aligned} S_\alpha(\rho||\sigma) &= (\alpha-1)^{-1}\ln \mathrm{tr}\big[\mathcal{M}_\Pi\big(A(\beta)^\alpha\big)\big] \\ &\geq (\alpha-1)^{-1}\ln \mathrm{tr}\big[\big(\mathcal{M}_\Pi\big(A(\beta)\big)\big)^\alpha\big] = S_\alpha(\mathcal{M}_\Pi(\rho)||\sigma)\,. \end{aligned} \tag{153}$$

This proves (151) and thus the non-negativity of S_α for $\alpha \in (0, 1)$. Observe that $S_\alpha(\rho||\sigma) = S_\alpha(\mathcal{M}_\Pi(\rho)||\sigma)$ if and only if (152) holds with equality, that is, $\langle k|A(\beta)|k\rangle^\alpha = \langle k|A(\beta)^\alpha|k\rangle$ for all k. By the strict concavity of $f(x) = x^\alpha$, $\{|k\rangle\}$ must then be an eigenbasis of $A(\beta)$, and thereby also of ρ. Thus ρ and σ commute and $S_\alpha(\rho||\sigma)$ coincides with the classical Rényi divergence $S_\alpha^{\mathrm{clas}}(\mathbf{p}||\mathbf{q})$. By the aforementioned properties of $S_\alpha^{\mathrm{clas}}(\mathbf{p}||\mathbf{q})$, it follows from (151) that $S_\alpha(\rho||\sigma) = 0$ implies $\mathbf{p} = \mathbf{q}$ and thus $\rho = \sigma$.

It remains to show the statement (iv) of the theorem. Following Ref. 60, we obtain (iv) with the help of a duality formula for $F_\alpha(\rho, \sigma)$ and of Lieb's concavity and Ando's convexity theorems. We omit here the proof of these two important theorems, which can be found in Ref. 33 (see also Ref. 117 for the Lieb theorem). The duality formula will be shown at the end this subsection.

Lemma 6.C.2. (Lieb's concavity and Ando's convexity theorem[8,98]) *For any $K \in \mathcal{B}(\mathcal{H})$ and any $\beta \in [-1, 1]$, the function $(R, S) \mapsto \mathrm{tr}(K^* R^q K S^{-\beta})$ on $\mathcal{B}(\mathcal{H})_+ \times \mathcal{B}(\mathcal{H})_+$ is jointly concave in (R, S) if $-1 \leq \beta \leq 0$ and $0 \leq q \leq 1 + \beta$ and is jointly convex in (R, S) if $0 \leq \beta \leq 1$ and $1 + \beta \leq q \leq 2$.*

Lemma 6.C.3. (Duality formula for the α-fidelity[60]) *If $\alpha \in (0, 1)$ (that is, $\beta = \alpha^{-1} - 1 > 0$) then*

$$F_\alpha(\rho, \sigma)^\alpha = \inf_{H \geq 0}\Big\{\alpha\,\mathrm{tr}(H\rho) + (1-\alpha)\,\mathrm{tr}\big[(\sqrt{H}\sigma^{-\beta}\sqrt{H})^{-\frac{1}{\beta}}\big]\Big\}. \tag{154}$$

If $\alpha > 1$ (that is, $-1 < \beta < 0$), the same identity holds but with the infimum replaced by a supremum.

Given Lemma 6.C.3, if one can show that, for a fixed operator $B \in \mathcal{B}(\mathcal{H})$, the function

$$g_{B,\beta}(\sigma) = \mathrm{tr}\big[(B^*\sigma^{-\beta}B)^{-\frac{1}{\beta}}\big] \tag{155}$$

is concave in σ when $-1 \leq \beta \leq 1$, $\beta \neq 0$, it will follow that $F_\alpha(\rho||\sigma)^\alpha$ is jointly concave for $\alpha \in [1/2, 1)$ (i.e., $0 < \beta \leq 1$) and jointly convex for $\alpha > 1$ (i.e., $-1 < \beta < 0$), thereby proving

Theorem 6.C.1. We first assume $-1 \leq \beta < 0$. For any operator $Y \geq 0$, let us set

$$h_Y(X) = \mathrm{tr}(Y X^{1+\beta}) - (1+\beta)\,\mathrm{tr}(X) \tag{156}$$

with $X \in \mathcal{B}(\mathcal{H})_+$. Given two self-adjoint matrices Y and Z, it is known that (see Ref. 27, Problem III.6.14)

$$\sum_{i=1}^n y_{n-i} z_i \leq \mathrm{tr}(YZ) \leq \sum_{i=1}^n y_i z_i, \tag{157}$$

where $y_1 \geq y_2 \geq \cdots \geq y_n$ and $z_1 \geq z_2 \geq \cdots \geq z_n$ are the eigenvalues of Y and Z in non-increasing order. Therefore,

$$\sup_{X\geq 0}\{h_Y(X)\} = \max_{\mathbf{x}}\Big\{\sum_{i=1}^n \big(y_i x_i^{1+\beta} - (1+\beta)x_i\big)\Big\} = -\beta \sum_{i=1}^n y_i^{-\frac{1}{\beta}} = -\beta\,\mathrm{tr}\big(Y^{-\frac{1}{\beta}}\big), \tag{158}$$

the maximum in the second member being over all vectors $\mathbf{x} \in \mathbb{R}_+^n$. Similarly, it follows from (157) that if $0 < \beta \leq 1$ then $\inf_{X\geq 0}\{h_Y(X)\} = -\beta\,\mathrm{tr}(Y^{-\frac{1}{\beta}})$. Plugging $Y = B^*\sigma^{-\beta}B$ into these identities, one finds

$$g_{B,\beta}(\sigma) = \sup_{X\geq 0}\Big\{-\beta^{-1}\big(\mathrm{tr}(B^*\sigma^{-\beta}BX^{1+\beta}) - (1+\beta)\,\mathrm{tr}(X)\big)\Big\}, \quad -1 \leq \beta < 0 \ \text{ or } \ 0 < \beta \leq 1. \tag{159}$$

Let us introduce the 2×2 block matrices

$$K = \begin{pmatrix} 0 & 0 \\ B^* & 0 \end{pmatrix}, \quad S = \begin{pmatrix} \sigma & 0 \\ 0 & X \end{pmatrix}. \tag{160}$$

A simple calculation gives

$$\mathrm{tr}(B^*\sigma^{-\beta}BX^{1+\beta}) = \mathrm{tr}_{\mathcal{H}\otimes\mathbb{C}^2}(K^* S^{1+\beta} K S^{-\beta}). \tag{161}$$

By Lemma 6.C.2, the right-hand side of (161) is concave (respectively, convex) in S when $-1 \leq \beta < 0$ (respectively, $0 < \beta \leq 1$). As a result, the left-hand side is jointly concave (convex) in (σ, X). But the maximum over X of a jointly concave function $f(\sigma, X)$ is concave in σ. Thanks to (159), we may conclude that $g_{B,\beta}(\sigma)$ is concave in σ for all $\beta \in [-1, 1]$, $\beta \neq 0$. The proof of Theorem 6.C.1 is now complete. □

Let us come back to the duality formula (154). We observe in passing that this formula bears some similarity with the variational formula (116) for the von Neumann entropy.

Proof of lemma 6.C.3. Since $\sigma^{-\frac{\beta}{2}} H \sigma^{-\frac{\beta}{2}}$ has the same non-zero eigenvalues as $\sqrt{H}\sigma^{-\beta}\sqrt{H}$, the quantity inside the infimum in (154) is equal to

$$g(H) = \alpha\,\mathrm{tr}(H\rho) + (1-\alpha)\,\mathrm{tr}\big[(\sigma^{-\frac{\beta}{2}} H \sigma^{-\frac{\beta}{2}})^{-\frac{1}{\beta}}\big]. \tag{162}$$

Differentiating the right-hand side with respect to the matrix elements of H in the some orthonormal basis $\{|i\rangle\}$ and using the relation $\partial\,\mathrm{tr}[f(B)]/\partial B_{ij} = f'(B)_{ji}$ with $f(x)$ a C^1-function, we get

$$\frac{\partial g(H)}{\partial H_{ij}} = \alpha\Big(\rho - \sigma^{-\frac{\beta}{2}}(\sigma^{-\frac{\beta}{2}} H \sigma^{-\frac{\beta}{2}})^{-\frac{1}{\beta}-1}\sigma^{-\frac{\beta}{2}}\Big)_{ji}. \tag{163}$$

Hence $g(H)$ has an extremum if and only if $H = \widehat{H} = \sigma^{\frac{\beta}{2}}(\sigma^{\frac{\beta}{2}}\rho\sigma^{\frac{\beta}{2}})^{\alpha-1}\sigma^{\frac{\beta}{2}} \geq 0$. But

$$g(\widehat{H}) = \mathrm{tr}[(\sigma^{\frac{\beta}{2}}\rho\sigma^{\frac{\beta}{2}})^{\alpha}] = F_\alpha(\rho||\sigma)^\alpha. \tag{164}$$

As $B \in \mathcal{B}(\mathcal{H})_+ \mapsto \mathrm{tr}(B^p)$ is convex for $p \geq 1$ or $p \leq 0$, $g(H)$ is convex if $\alpha \in (0, 1)$ (i.e., $-\beta^{-1} < 0$) and concave if $\alpha > 1$ (i.e., $-\beta^{-1} > 1$). It follows that $g(\widehat{H})$ is a minimum for $\alpha \in (0, 1)$ and a maximum for $\alpha > 1$. □

Let us point out that it follows from Lemma 6.C.2 that the normal-ordered Rényi entropy (142) is also jointly convex for $\alpha \in (0, 1)$. Taking $\alpha \to 1$ and recalling that $S_\alpha^{(\mathrm{n})}(\rho||\sigma) \to S(\rho||\sigma)$, this gives a

direct proof the joint convexity of the relative von Neumann entropy $S(\rho||\sigma)$ from the Lieb concavity theorem, as noted by Lindblad.[101,102] Combined with Proposition 6.B.2, this leads to a completely different justification of the monotonicity of $S(\rho||\sigma)$ in Theorem 6.B.1 than that presented in Sec. VI B. It would be interesting to look for a generalization of the arguments of Petz in Sec. VI B to the case of the α-entropies.

3. Monotonicity in α

As stated above, a very nice feature of the α-entropy (143) is that, like the classical Rényi divergence, it is monotonous in α. This leads in particular to some bound between the relative von Neumann entropy and the fidelity (see (199) below).

Proposition 6.C.4.[114] *For any $\rho, \sigma \in \mathcal{E}(\mathcal{H})$, $S_\alpha(\rho||\sigma)$ is a non-decreasing function of α on $(0, \infty)$.*

Proof. One first derive the following identity similar to (159):

$$\big(g_{B,-\alpha^{-1}}(\sigma)\big)^{\frac{1}{\alpha}} = \big\|B^*\sigma^{1/\alpha}B\big\|_\alpha = \sup_{\tau\geq 0,\mathrm{tr}(\tau)=1} \mathrm{tr}\big(B^*\sigma^{1/\alpha}B\,\tau^{1-1/\alpha}\big), \quad \alpha \geq 1. \tag{165}$$

If $0 < \alpha \leq 1$ the supremum has to be replaced by an infimum. When $\alpha \geq 1$ this identity is nothing but a rewriting of the Hölder's inequality (3). The derivation for $\alpha \in (0, 1)$ relies on (157) and follows the same lines as for the derivation of (159) (apart from the fact that we substituted β by $-1/\alpha$), but one must introduce a Lagrange multiplier to account for the constraint $\mathrm{tr}(\tau) = 1$. Applying the relation (165) for $B = \sigma^{-\frac{1}{2}}\rho^{\frac{1}{2}}$ and plugging the identity $\|\sigma^{\frac{\beta}{2}}\rho\sigma^{\frac{\beta}{2}}\|_\alpha = \|\rho^{\frac{1}{2}}\sigma^\beta\rho^{\frac{1}{2}}\|_\alpha$ into (149), we are led to

$$S_\alpha(\rho||\sigma) = \sup_{\tau\in\mathcal{E}(\mathcal{H})}\big\{-\beta^{-1}\ln F_\alpha(\rho||\sigma;\tau)\big\}, \quad F_\alpha(\rho||\sigma;\tau) = \mathrm{tr}\big(\rho^{\frac{1}{2}}\sigma^\beta\rho^{\frac{1}{2}}\tau^{-\beta}\big) = \big\langle\xi,\ \Delta^\beta_{\sigma|\tau}\xi\big\rangle, \tag{166}$$

for any $\alpha > 0$, $\alpha \neq 1$. In the last identity $\xi = \rho^{\frac{1}{2}}$ and we have introduced the relative modular operator, see (8). For any fixed $\tau \in \mathcal{E}(\mathcal{H})$, one finds

$$\frac{\mathrm{d}}{\mathrm{d}\beta}\Big(-\beta^{-1}\ln F_\alpha(\rho||\sigma;\tau)\Big) = -\frac{1}{\beta^2 F_\alpha(\rho||\sigma;\tau)}\Big(\langle\xi,\ \Delta^\beta_{\sigma|\tau}\ln(\Delta^\beta_{\sigma|\tau})\xi\rangle - \langle\xi,\ \Delta^\beta_{\sigma|\tau}\xi\rangle\ln\langle\xi,\ \Delta^\beta_{\sigma|\tau}\xi\rangle\Big). \tag{167}$$

The Jensen inequality applied to the convex function $f(x) = x\ln x$ implies that the quantity inside the parenthesis in the right-hand side is non-negative. Thus $-\beta^{-1}F_\alpha(\rho||\sigma;\tau)$ is a non-increasing function of β. This is true for any density matrix τ, thus one infers from (166) that $\alpha \mapsto S_\alpha(\rho||\sigma)$ is non-decreasing. □

VII. THE BURES DISTANCE AND UHLMANN FIDELITY

In this section we study the Bures distance on the set of quantum states $\mathcal{E}(\mathcal{H})$. This distance is Riemannian and monotonous with respect to quantum operations. It is a simple function of the fidelity (95). Its metric coincides with the quantum Fisher information quantifying the best achievable precision in the parameter estimation problem discussed in Sec. VIII B. The material of this section (as well as of Sec. VIII) is completely independent from that of Secs. IX and X, so it is possible at this point to proceed directly to Sec. IX. The reading of Secs. VII A–VII C is, however, recommended before going through Sec. XI devoted to the geometrical measures of quantum correlations, where the Bures distance plays the key role. The section is organized as follows. Section VII A contains a short discussion on contractive (i.e., monotonous) distances. It is argued there that the distances induced by the $\|\cdot\|_p$-norm are not contractive save for $p = 1$. The definition and main properties of the Bures distance are given in Secs. VII B–VII D. The Bures metric is determined in Sec. VII E. Finally, Sec. VII F contains the proof of an important result of Petz on the characterization of all Riemannian contractive metrics on $\mathcal{E}(\mathcal{H})$ for finite-dimensional Hilbert spaces $\mathcal{H}$.

A. Contractive and convex distances

In order to quantify how far are two states ρ and σ it is necessary to define a distance on the set $\mathcal{E}(\mathcal{H})$ of quantum states. One has a priori the choice between many distances. The most common ones are the L^p-distances defined by (2). In quantum information theory it seems, however, natural to impose the following requirement.

Definition 7.A.1. *A distance d on the sets of quantum states is contractive if for any finite Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$, any quantum operation $\mathcal{M} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$, and any $\rho, \sigma \in \mathcal{E}(\mathcal{H})$, it holds*

$$d(\mathcal{M}(\rho), \mathcal{M}(\sigma)) \leq d(\rho, \sigma). \tag{168}$$

A contractive distance is in particular invariant under unitary conjugations, i.e.,

$$d\big(U\rho\, U^*, U\sigma\, U^*\big) = d(\rho, \sigma) \quad \text{if } U \text{ is unitary} \tag{169}$$

(in fact, $\rho \mapsto U\rho\, U^*$ is an invertible quantum operation on $\mathcal{B}(\mathcal{H})$). For such a distance, if a generalized measurement is performed on a system, two states are closer from each other after the measurement than before it, and if the system is subject to a unitary evolution the distance between the time-evolved states remains unchanged.

For $p > 1$, the distances d_p (in particular, the Hilbert-Schmidt distance d_2) are not contractive. A counter-example for two qubits is obtained[122] by taking $\mathcal{M}(\rho) = A_1\rho A_1^* + A_2\rho A_2^*$ with

$$A_1 = \sigma_+ \otimes 1, \quad A_2 = \sigma_+\sigma_- \otimes 1, \quad \rho = \frac{1}{2} \otimes \sigma_+\sigma_-, \quad \sigma = \frac{1}{2} \otimes \sigma_-\sigma_+ \tag{170}$$

(here $\sigma_+ = |1\rangle\langle 0|$ is the raising operator and $\sigma_- = \sigma_+^*$). Then $\|\mathcal{M}(\rho) - \mathcal{M}(\sigma)\|_p = 2^{1/p}$ is larger than $\|\rho - \sigma\|_p = 2^{2p-1}$.

Proposition 7.A.2.[136] *The trace distance d_1 is contractive.*

Proof. Let $R = \rho - \sigma = R_+ - R_-$ with $R_\pm = (|R| \pm R)/2 = \pm RP_\pm \geq 0$ the positive and negative parts of R (here P_+ and P_- are the spectral projectors of R on $[0, \infty)$ and $(-\infty, 0)$). Then $\|R\|_1 = \mathrm{tr}(R_+ + R_-) = 2\,\mathrm{tr}(R_+)$ because $\mathrm{tr}(R) = \mathrm{tr}(R_+) - \mathrm{tr}(R_-) = 0$. Since $\mathcal{M}$ is trace preserving and CP, one has $\|\mathcal{M}(R)\|_1 = 2\,\mathrm{tr}[\mathcal{M}(R)_+]$ and $\mathcal{M}(R)_+ = (\mathcal{M}(R_+) - \mathcal{M}(R_-))_+ \leq \mathcal{M}(R_+)$. Thus $\|\mathcal{M}(R)\|_1 \leq 2\,\mathrm{tr}[\mathcal{M}(R_+)] = 2\,\mathrm{tr}[R_+] = \|R\|_1$. □

A distance d on $\mathcal{E}(\mathcal{H})$ is *jointly convex* if for any state ensembles $\{\rho_i, p_i\}$ and $\{\sigma_i, p_i\}$ with the same probabilities p_i,

$$d\Big(\sum_i p_i\rho_i, \sum_i p_i\sigma_i\Big) \leq \sum_i p_i d(\rho_i, \sigma_i). \tag{171}$$

Since they are associated to a norm, the distances d_p are jointly convex for any $p \geq 1$.

B. The Bures distance

We now introduce the Bures distance d_B. This distance is contractive like d_1. It was first considered by Bures in the context of infinite products of von Neumann algebras[32] (see also Ref. 9) and was later studied in a series of papers by Uhlmann.[154,156,157] Uhlmann used it to define parallel transport and related it to the fidelity generalizing the usual fidelity $|\langle\psi|\phi\rangle|^2$ between pure states. Indeed, d_B is a extension to mixed states of the Fubini-Study distance on the projective space $P\mathcal{H}$ of pure states,

$$d_{FS}\big(\rho_\psi, \sigma_\phi\big) = \inf_{|\psi\rangle, |\phi\rangle} \big\||\psi\rangle - |\phi\rangle\big\| = \big(2 - 2|\langle\psi|\phi\rangle|\big)^{\frac{1}{2}}, \tag{172}$$

where the infimum in the second member is over all representatives $|\psi\rangle$ of $\rho_\psi \in P\mathcal{H}$ and $|\phi\rangle$ of $\sigma_\phi \in P\mathcal{H}$ (i.e., $\rho_\psi = |\psi\rangle\langle\psi|$ and $\sigma_\phi = |\phi\rangle\langle\phi|$). Observe that the third member is independent of

these representatives. For two mixed states ρ and σ in $\mathcal{E}(\mathcal{H})$, one can define analogously[156,84]

$$d_{\mathrm{B}}(\rho,\sigma)=\inf_{A,B} d_2(A-B), \tag{173}$$

where the infimum is over all Hilbert-Schmidt matrices A and B satisfying $AA^*=\rho$ and $BB^*=\sigma$. Such matrices are given by $A=\sqrt{\rho}V$ and $B=\sqrt{\sigma}W$ for some unitaries V and W (polar decompositions). If $\rho=\rho_\psi$ and $\sigma=\sigma_\phi$ are pure states, then $A=|\psi\rangle\langle\mu|$ and $B=|\phi\rangle\langle\nu|$ with $\|\mu\|=\|\nu\|=1$, so that (173) reduces to the Fubini-Study distance (172).

For mixed states ρ and σ, the right-hand side of (173) is given by

$$\left(2-2\sup_U \mathrm{Re}\ \mathrm{tr}(U\sqrt{\rho}\sqrt{\sigma})\right)^{\frac{1}{2}} \tag{174}$$

with a supremum over all unitaries $U=WV^*$. This supremum is equal to $\|\sqrt{\rho}\sqrt{\sigma}\|_1$ and is attained if and only if $UU_0|\sqrt{\rho}\sqrt{\sigma}|^{\frac{1}{2}}=|\sqrt{\rho}\sqrt{\sigma}|^{\frac{1}{2}}$, where U_0 is such that $\sqrt{\rho}\sqrt{\sigma}=U_0|\sqrt{\rho}\sqrt{\sigma}|$ (see Sec. II A). Equivalently, the infimum in (173) is attained if and only if the parallel transport condition $A^*B\geq 0$ holds. We obtain the following equivalent definition of d_{B}.

Definition 7.B.1. *For any states $\rho,\sigma\in\mathcal{E}(\mathcal{H})$,*

$$d_{\mathrm{B}}(\rho,\sigma)=\left(2-2\sqrt{F(\rho,\sigma)}\right)^{\frac{1}{2}}, \tag{175}$$

where the Uhlmann fidelity is defined by

$$F(\rho,\sigma)=\|\sqrt{\rho}\sqrt{\sigma}\|_1^2=\left(\mathrm{tr}\left[(\sqrt{\sigma}\rho\sqrt{\sigma})^{\frac{1}{2}}\right]\right)^2. \tag{176}$$

The fidelity $F(\rho,\sigma)$ is symmetric in (ρ,σ) and belongs to the interval $[0,1]$. It is clearly a generalization of the usual pure state fidelity $F(|\psi\rangle,|\phi\rangle)=|\langle\psi|\phi\rangle|^2$. If σ_ϕ is pure, then

$$F(\rho,\sigma_\phi)=\langle\phi|\rho|\phi\rangle \tag{177}$$

for any $\rho\in\mathcal{E}(\mathcal{H})$.

It is immediate on (173) that d_{B} is positive and symmetric, and $d_{\mathrm{B}}(\rho,\sigma)=0$ if and only if $\rho=\sigma$. The triangle inequality is more difficult to show. It can be established with the help of the following astonishing theorem.

Theorem 7.B.2. (Uhlmann[154]) *Let $\rho,\sigma\in\mathcal{E}(\mathcal{H})$, and $|\Psi\rangle$ be a purification of ρ on the space $\mathcal{H}\otimes\mathcal{K}$, with $\dim\mathcal{K}\geq\dim\mathcal{H}$. Then*

$$F(\rho,\sigma)=\max_{|\Phi\rangle}|\langle\Psi|\Phi\rangle|^2 \tag{178}$$

where the maximum is over all purifications $|\Phi\rangle$ of σ on $\mathcal{H}\otimes\mathcal{K}$.

Proof. We give here a simple proof due to Josza.[91] Let us first assume $\mathcal{K}\simeq\mathcal{H}$. Let $|\Psi\rangle$ and $|\Phi\rangle$ be purifications of ρ and σ on $\mathcal{H}\otimes\mathcal{H}$, respectively. As it has been noticed in Sec. II C, by the Schmidt decomposition these purifications can always be written as

$$|\Psi\rangle=\sum_{k=1}^n\sqrt{p_k}|k\rangle|f_k\rangle,\quad |\Phi\rangle=\sum_{k=1}^n\sqrt{q_k}(U|k\rangle)|g_k\rangle, \tag{179}$$

where $\rho=\sum_k p_k|k\rangle\langle k|$ and $\sigma=\sum_k q_kU|k\rangle\langle k|U^*$ are spectral decompositions of ρ and σ, U is a unitary operator on $\mathcal{H}$, and $\{|f_k\rangle\}_{i=1}^n$ and $\{|g_k\rangle\}_{i=1}^n$ are two orthonormal bases of $\mathcal{H}$. Defining the unitaries V and W on $\mathcal{H}$ by $|f_k\rangle=V|k\rangle$ and $|g_k\rangle=W|k\rangle$ for any $k=1,\ldots,n$, we have

$$|\Psi\rangle=\sqrt{\rho}\otimes V|\Sigma\rangle,\quad |\Phi\rangle=\sqrt{\sigma}\,U\otimes W|\Sigma\rangle\quad\text{with}\quad |\Sigma\rangle=\sum_{k=1}^n|k\rangle|k\rangle. \tag{180}$$

The vector $|\Sigma\rangle$ is the vector associated to the identity operator on $\mathcal{B}(\mathcal{H})$ by the isomorphism (5). For any $X, Y \in \mathcal{B}(\mathcal{H})$, one obtains by setting $O = X^T \otimes Y$ in (6) and noting that $\mathrm{tr}(O^{\mathcal{R}}) = \mathrm{tr}(XY)$ that

$$\mathrm{tr}(XY) = \langle \Sigma | X^T \otimes Y | \Sigma \rangle \tag{181}$$

(here X^T is the transpose of X in the basis $\{|k\rangle\}$). Introducing the unitary $U_0 = V^* W U^T$, this gives

$$\sup_{|\Phi\rangle} |\langle \Phi | \Psi \rangle| = \sup_{W} |\langle \Sigma | U^* \sqrt{\sigma}\sqrt{\rho} \otimes W^* V | \Sigma \rangle| = \sup_{U_0} |\,\mathrm{tr}(\sqrt{\rho}\sqrt{\sigma}\, U_0^*)| = \|\sqrt{\rho}\sqrt{\sigma}\|_1 \,. \tag{182}$$

The last equality comes from (3). This proves the desired result. The supremum is achieved by choosing $|\Phi\rangle$ as in (179) with $U = U_0^T (W^*)^T V^T$, U_0 being a unitary in the polar decomposition of $\sqrt{\rho}\sqrt{\sigma}$.

If $\mathcal{K}$ has a dimension m larger than n, we extend ρ and σ to a space $\mathcal{H}' \simeq \mathcal{K}$ by adding to them new orthonormal eigenvectors $|k\rangle$ and $U|k\rangle$ with zero eigenvalues $p_k = q_k = 0$, $k = n+1, \ldots, m$. This does not change the fidelity $F(\rho, \sigma)$, thus $F(\rho, \sigma) = \max_{|\Phi'\rangle} |\langle \Psi' | \Phi' \rangle|^2$, where $|\Psi'\rangle$ is a purification of $\rho' = \sum_{k=1}^m p_k |k\rangle\langle k| = \rho$ on $\mathcal{H}' \otimes \mathcal{H}'$, and similarly for $|\Phi'\rangle$. But $|\Psi'\rangle$ and $|\Phi'\rangle$ have the form (179), hence they belong to $\mathcal{H} \otimes \mathcal{K}$. □

Let ρ, σ, and τ be three states of $\mathcal{E}(\mathcal{H})$ and $|\Psi\rangle$ be a purification of ρ on $\mathcal{H} \otimes \mathcal{H}$. According to Theorem 7.B.2, there exists a purification $|\Phi\rangle$ of σ on $\mathcal{H} \otimes \mathcal{H}$ such that $F(\rho, \sigma) = |\langle \Psi | \Phi \rangle|^2$. One can choose the arbitrary phase factor of $|\Phi\rangle$ in such a way that $\langle \Psi | \Phi \rangle \geq 0$, whence $\sqrt{F(\rho, \sigma)} = \langle \Psi | \Phi \rangle$. Similarly, there exists a purification $|\chi\rangle$ of τ such that $\sqrt{F(\sigma, \tau)} = \langle \Phi | \chi \rangle \geq 0$. In view of (175) and (178),

$$\begin{aligned} d_{\mathrm{B}}(\rho, \tau) &\leq \big(2 - 2|\langle \Psi | \chi \rangle|\big)^{\frac{1}{2}} \\ &\leq \big(2 - 2\mathrm{Re}\,\langle \Psi | \chi \rangle|\big)^{\frac{1}{2}} = \big\| |\Psi\rangle - |\chi\rangle \big\| \\ &\leq \big\| |\Psi\rangle - |\Phi\rangle \big\| + \big\| |\Phi\rangle - |\chi\rangle \big\| = \big(2 - 2\langle \Psi | \Phi \rangle|\big)^{\frac{1}{2}} + \big(2 - 2\langle \Phi | \chi \rangle|\big)^{\frac{1}{2}} \,, \end{aligned} \tag{183}$$

showing that d_{B} satisfies the triangle inequality $d_{\mathrm{B}}(\rho, \tau) \leq d_{\mathrm{B}}(\rho, \sigma) + d_{\mathrm{B}}(\sigma, \tau)$.

Corollary 7.B.3. *The map $(\rho, \sigma) \mapsto d_{\mathrm{B}}(\rho, \sigma)$ defines a distance d_{B} on quantum states, with values in* [0, 1]. *This distance is contractive. Moreover, d_{B}^2 is jointly convex.*

Note that d_{B} is not jointly convex. One gets a counter-example by choosing $\rho_0 = \sigma_0 = |0\rangle\langle 0|$, $\rho_1 = |1\rangle\langle 1|$, $\sigma_1 = |2\rangle\langle 2|$, and $p_0 = p_1 = 1/2$, $\{|0\rangle, |1\rangle, |2\rangle\}$ being an orthonormal family in $\mathcal{H}$.

It is clear on (176) that $F(\rho, \sigma) = 0$ if and only if ρ and σ have orthogonal supports, $\mathrm{ran}\,\rho \perp \mathrm{ran}\,\sigma$. Therefore, two states ρ and σ have a maximal distance $d_{\mathrm{B}}(\rho, \sigma) = 1$ if they are orthogonal and thus perfectly distinguishable.

Proof. We have already established above that d_{B} satisfies all the axioms of a distance. To show the contractivity, it is enough to check that for any quantum operation $\mathcal{M} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ and any states $\rho, \sigma \in \mathcal{E}(\mathcal{H})$,

$$F(\mathcal{M}(\rho), \mathcal{M}(\sigma)) \geq F(\rho, \sigma). \tag{184}$$

This property of the fidelity is a consequence of the contractivity of the relative Rényi entropy for $\alpha = 1/2$ (Theorem 6.C.1*(v)*). It is, however, instructive to re-derive this result from Theorem 7.B.2. According to this theorem, there exist some purifications $|\Psi\rangle$ and $|\Phi\rangle$ of ρ and σ on $\mathcal{H} \otimes \mathcal{K}$ such that $F(\rho, \sigma) = |\langle \Psi | \Phi \rangle|^2$. Now, thanks to (34) one obtains some purifications $|\Psi_{\mathcal{M}}\rangle = 1_{\mathcal{K}} \otimes U |\Psi\rangle |\epsilon_0\rangle$ of $\mathcal{M}(\rho)$ and $|\Phi_{\mathcal{M}}\rangle = 1_{\mathcal{K}} \otimes U |\Phi\rangle |\epsilon_0\rangle$ of $\mathcal{M}(\sigma)$ on $\mathcal{K} \otimes \mathcal{H}' \otimes \mathcal{H}'_{\mathsf{E}}$, with $|\epsilon_0\rangle \in \mathcal{H}_{\mathsf{E}}$ and $U : \mathcal{H} \otimes \mathcal{H}_{\mathsf{E}} \to \mathcal{H}' \otimes \mathcal{H}'_{\mathsf{E}}$ unitary. Thus

$$F(\mathcal{M}(\rho), \mathcal{M}(\sigma)) \geq |\langle \Psi_{\mathcal{M}} | \Phi_{\mathcal{M}} \rangle|^2 = |\langle \Psi | \Phi \rangle|^2 = F(\rho, \sigma). \tag{185}$$

The joint convexity of d_{B}^2 is a consequence of the bound

$$\sqrt{F\Big(\sum_i p_i\rho_i, \sum_i q_i\sigma_i\Big)} \geq \sum_i \sqrt{p_i q_i}\sqrt{F(\rho_i,\sigma_i)}\,, \tag{186}$$

where $\{\rho_i, p_i\}$ and $\{\sigma_i, q_i\}$ are arbitrary ensembles in $\mathcal{E}(\mathcal{H})$. Note that the statement (186) is slightly more general than the joint concavity of the square root of $F(\rho,\sigma)$ proven in Sec. VI C (Theorem 6.C.1*(iv)*). To show that (186) is true, we introduce as before some purifications $|\Psi_i\rangle$ of ρ_i and $|\Phi_i\rangle$ of σ_i on $\mathcal{H}\otimes\mathcal{H}$ such that $\sqrt{F(\rho_i,\sigma_i)} = \langle\Psi_i|\Phi_i\rangle$. Let us define the vectors

$$|\Psi\rangle = \sum_i \sqrt{p_i}|\Psi_i\rangle|\epsilon_i\rangle, \quad |\Phi\rangle = \sum_i \sqrt{p_i}|\Phi_i\rangle|\epsilon_i\rangle \tag{187}$$

in $\mathcal{H}\otimes\mathcal{H}\otimes\mathcal{H}_{\mathsf{E}}$, where $\mathcal{H}_{\mathsf{E}}$ is an auxiliary Hilbert space and $\{|\epsilon_i\rangle\}$ is an orthonormal basis of $\mathcal{H}_{\mathsf{E}}$. Then $|\Psi\rangle$ and $|\Phi\rangle$ are purifications of $\rho = \sum_i p_i\,\rho_i$ and $\sigma = \sum_i q_i\sigma_i$, respectively. One infers from Theorem 7.B.2 that

$$\sqrt{F(\rho,\sigma)} \geq |\langle\Psi|\Phi\rangle| = \sum_i \sqrt{p_i q_i}\langle\Psi_i|\Phi_i\rangle = \sum_i \sqrt{p_i q_i}\sqrt{F(\rho_i,\sigma_i)}\,. \tag{188}$$

This complete the proof of the corollary. □

Note that one cannot replace $\sqrt{F}$ by F in (186), that is, $F(\rho,\sigma)$ is not jointly concave (one can take the same counter-example as that given above for d_{B}.) However, by a slight modification of the proof of Corollary 7.B.3 one can show that $\rho \mapsto F(\rho,\sigma)$ and $\sigma \mapsto F(\rho,\sigma)$ are concave. (In their book,[117] Nielsen and Chuang define the fidelity as the square root of (176). This must be kept in mind when comparing the results in this monograph with those of this article.)

Remark 7.B.4. A consequence of (47) and (177) and of the monotonicity of the fidelity F with respect to partial trace operations (see (184)) is that the entanglement fidelity $F_{\mathrm{e}}(\rho,\mathcal{M})$ of a state ρ with respect to a quantum operation $\mathcal{M}$ satisfies

$$F_{\mathrm{e}}(\rho,\mathcal{M}) \leq F(\rho,\mathcal{M}(\rho)). \tag{189}$$

Remark 7.B.5. As the fidelity satisfies $F(\rho\otimes\rho',\,\sigma\otimes\sigma') = F(\rho,\,\sigma)F(\rho',\,\sigma')$, the Bures distance increases by taking tensor products, $d_{\mathrm{B}}(\rho\otimes\rho',\sigma\otimes\sigma') \geq d_{\mathrm{B}}(\rho,\sigma)$ for any $\rho,\sigma\in\mathcal{E}(\mathcal{H})$, $\rho',\sigma'\in\mathcal{E}(\mathcal{H}')$, with equality if and only if $\rho'=\sigma'$. This has to be contrasted with the trace distance, which does not enjoy this property.

In Subsections VII C and VII D we collect some important properties of the Bures distance. We refer the reader to the monographs Ref. 20 and 117 for a list of names to which these properties should be attached.

C. Bures distance and statistical distance in classical probability

The restriction of a distance d on $\mathcal{E}(\mathcal{H})$ to all density matrices commuting with a given state ρ_0 defines a distance on the simplex $\mathcal{E}_{\mathrm{clas}} = \{\mathbf{p}\in\mathbb{R}_+^n; \sum_i p_i = 1\}$ of classical probabilities on the finite space $\{1,2,\ldots,n\}$. In particular, if ρ and σ are two commuting states with spectral decompositions $\rho = \sum_k p_k|k\rangle\langle k|$ and $\sigma = \sum_k q_k|k\rangle\langle k|$, then

$$d_1(\rho,\sigma) = d_1^{\mathrm{clas}}(\mathbf{p},\mathbf{q}) = \sum_{k=1}^n |p_k - q_k|$$

is the ℓ^1-distance, and

$$d_{\mathrm{B}}(\rho,\sigma) = d_{\mathrm{H}}^{\mathrm{clas}}(\mathbf{p},\mathbf{q}) = \Big(\sum_{k=1}^n(\sqrt{p_k}-\sqrt{q_k})^2\Big)^{\frac12} = \Big(2 - 2\sum_{k=1}^n\sqrt{p_k q_k}\Big)^{\frac12} \tag{190}$$

is the Hellinger distance. A distance closely related to $d_{\mathrm{H}}^{\mathrm{clas}}$ is the so-called statistical distance $\Theta^{\mathrm{clas}}(\mathbf{p},\mathbf{q}) = \arccos(1 - d_{\mathrm{H}}^{\mathrm{clas}}(\mathbf{p},\mathbf{q})^2/2)$, i.e., the angle between the vectors $\mathbf{x} = (\sqrt{p_k})_{k=1}^n$ and $\mathbf{y} = (\sqrt{q_k})_{k=1}^n$ on the unit sphere. Given two non-commuting states ρ and σ, one can consider the distance $d^{\mathrm{clas}}(\mathbf{p},\mathbf{q})$ between the outcome probabilities $\mathbf{p}$ and $\mathbf{q}$ of a measurement performed on the system in states ρ and σ, respectively. It is natural to ask whether there is a relation between $d(\rho,\sigma)$ and the supremum of $d^{\mathrm{clas}}(\mathbf{p},\mathbf{q})$ over all measurements.

Proposition 7.C.1. *For any $\rho, \sigma \in \mathcal{E}(\mathcal{H})$,*

$$d_1(\rho,\sigma) = \sup_{\{M_i\}} d_1^{\mathrm{clas}}(\mathbf{p},\mathbf{q}), \quad d_{\mathrm{B}}(\rho,\sigma) = \sup_{\{M_i\}} d_{\mathrm{H}}^{\mathrm{clas}}(\mathbf{p},\mathbf{q}), \tag{191}$$

where the suprema are over all POVMs $\{M_i\}$ and $p_i = \mathrm{tr}(M_i\rho)$ (respectively $q_i = \mathrm{tr}(M_i\sigma)$) is the probability of the measurement outcome i in the state ρ (respectively σ). Moreover, the suprema are achieved for von Neumann measurements with rank-one projectors $M_i = |i\rangle\langle i|$.

Proof. We leave the justification of the first identity to the reader. It can be obtained by following similar arguments as in the proof of Proposition 7.A.2 (see Ref. 117). Let us show the second identity. Given a POVM $\{M_i\}$, by taking advantage of the definition (176) of the fidelity, the polar decomposition $\sqrt{\rho}\sqrt{\sigma} = U|\sqrt{\rho}\sqrt{\sigma}|$, and the identity $\sum_i M_i = 1$, one gets

$$\sqrt{F(\rho,\sigma)} = \sum_i \mathrm{tr}(U^*\sqrt{\rho}\sqrt{M_i}\sqrt{M_i}\sqrt{\sigma}) \;\le\; \sum_i \sqrt{p_i q_i}. \tag{192}$$

The upper bound comes from the Cauchy-Schwarz inequality. It remains to show that this bound can be attained for an appropriate choice of POVM. The Cauchy-Schwarz inequality holds with equality if and only if $\sqrt{M_i}\sqrt{\rho}\,U = \lambda_i\sqrt{M_i}\sqrt{\sigma}$ with $\lambda_i \in \mathbb{C}$. Assuming $\sigma > 0$ and observing that $\sqrt{\rho}\,U = \sigma^{-\frac{1}{2}}|\sqrt{\rho}\sqrt{\sigma}|$, this identity can be recast as

$$\sqrt{M_i}(R - \lambda_i) = 0 \quad \text{with} \quad R = \sigma^{-\frac{1}{2}}|\sqrt{\rho}\sqrt{\sigma}|\sigma^{-\frac{1}{2}}. \tag{193}$$

Let $R = \sum_i r_i|i\rangle\langle i|$ be a spectral projection of the non-negative matrix R. Taking M_i to be the von Neumann projector $M_i = |i\rangle\langle i|$ and $\lambda_i = r_i$, we find that (193) is satisfied for all i. Thus $\sqrt{F(\rho,\sigma)}$ is equal to the right-hand side of (192). If σ is not invertible it can be approached by invertible density matrices $\sigma_\varepsilon = (1-\varepsilon)\sigma + \varepsilon$, $\varepsilon > 0$, and the result follows by continuity. □

Much as for the quantum relative Rényi entropies (Sec. VI C), one may define another distance on $\mathcal{E}(\mathcal{H})$ which also reduces to the Hellinger distance $d_{\mathrm{H}}^{\mathrm{clas}}$ for commuting matrices, by setting

$$d_{\mathrm{H}}(\rho,\sigma) = d_2(\sqrt{\rho},\sqrt{\sigma}) = \left(2 - 2\sqrt{F_{\frac{1}{2}}^{(\mathrm{n})}(\rho||\sigma)}\right)^{\frac{1}{2}}, \tag{194}$$

where $F_\alpha^{(\mathrm{n})}(\rho||\sigma)$ is the fidelity associated to the normal-ordered α-entropy (142), namely,

$$F_\alpha^{(\mathrm{n})}(\rho||\sigma) = \left(\mathrm{tr}\left[\rho^\alpha\sigma^{1-\alpha}\right]\right)^{\frac{1}{\alpha}} = e^{-\beta S_\alpha^{(\mathrm{n})}(\rho||\sigma)}, \quad \beta = \frac{1-\alpha}{\alpha}. \tag{195}$$

This distance is sometimes called the quantum Hellinger distance. Thanks to Lieb's concavity theorem (Lemma 6.C.2), $F_\alpha^{(\mathrm{n})}(\rho||\sigma)^\alpha$ is jointly concave in (ρ,σ) for all $\alpha \in (0,1)$. Consequently, the square Hellinger distance $d_{\mathrm{H}}(\rho,\sigma)^2$ is jointly convex, just as $d_{\mathrm{B}}(\rho,\sigma)^2$. From Proposition 6.B.2 one then deduces that d_{H} is contractive. It is worth noting that d_{H} does not coincide with the Fubini-study distance (172) for pure states (in fact, one finds $F_{1/2}^{(\mathrm{n})}(\rho_\psi||\sigma_\phi) = |\langle\psi|\phi\rangle|^4$). For any $\rho, \sigma \in \mathcal{E}(\mathcal{H})$, one finds by comparing (173) and (194) that $d_{\mathrm{B}}(\rho,\sigma) \le d_{\mathrm{H}}(\rho,\sigma)$.

D. Comparison of the Bures and trace distances

The next result shows that the Bures and trace distances d_{B} and d_1 are equivalent and gives optimal bounds of d_1 in terms of d_{B}.

Proposition 7.D.1. *For any $\rho, \sigma \in \mathcal{E}(\mathcal{H})$, one has*

$$d_{\mathrm{B}}(\rho,\sigma)^2 \leq d_1(\rho,\sigma) \leq 2\Big\{1-\Big(1-\frac{1}{2}d_{\mathrm{B}}(\rho,\sigma)^2\Big)^2\Big\}^{\frac{1}{2}}. \tag{196}$$

The lower bound has been first proven by Araki[9] in the C^*-algebra setting. We shall justify it from Proposition 7.C.1 as in Ref. 117. The upper bound is saturated for pure states, as shown in the proof below. Note that this bound implies that $d_1(\rho,\sigma) \leq 2d_{\mathrm{B}}(\rho,\sigma)$.

Proof. We first argue that if $\rho_\psi = |\psi\rangle\langle\psi|$ and $\sigma_\phi = |\phi\rangle\langle\phi|$ are pure states, then $d_1(\rho_\psi,\sigma_\phi) = 2\sqrt{1-F(\rho_\psi,\sigma_\phi)}$ and thus the upper bound in (196) is an equality. Actually, let $|\phi\rangle = \cos\theta|\psi\rangle + e^{\mathrm{i}\delta}\sin\theta|\psi^\perp\rangle$, where $\theta, \delta \in [0, 2\pi)$ and $|\psi^\perp\rangle$ is a unit vector orthogonal to $|\psi\rangle$. Since $\rho_\psi - \sigma_\phi$ has non-vanishing eigenvalues $\pm\sin\theta$, one has $d_1(\rho_\psi,\sigma_\phi) = 2|\sin\theta|$. But $F(\rho_\psi,\sigma_\phi) = \cos^2\theta$, hence the aforementioned statement is true. It then follows from Theorem 7.B.2 and from the contractivity of the trace distance with respect to partial trace operations (Proposition 7.A.2) that for arbitrary ρ and $\sigma \in \mathcal{E}(\mathcal{H})$,

$$d_1(\rho,\sigma) \leq 2\sqrt{1-F(\rho,\sigma)}. \tag{197}$$

To bound $d_1(\rho,\sigma)$ from below, we use Proposition 7.C.1 and consider a generalized measurement $\{M_i\}$ such that $\sqrt{F(\rho,\sigma)} = \sum_i \sqrt{p_i q_i}$ with $p_i = \mathrm{tr}(\rho M_i)$ and $q_i = \mathrm{tr}(\sigma M_i)$. This yields

$$d_{\mathrm{B}}(\rho,\sigma)^2 = \sum_i (\sqrt{p_i}-\sqrt{q_i})^2 \leq \sum_i |p_i - q_i| \leq d_1(\rho,\sigma), \tag{198}$$

where the last inequality comes from Proposition 7.C.1 again. □

The following bound on the relative entropy can be obtained from (149) and (145), and Proposition 6.C.4

$$S(\rho||\sigma) \geq -2\ln\Big(1-\frac{1}{2}d_{\mathrm{B}}(\rho,\sigma)^2\Big) \geq -\ln\Big(1-\frac{1}{4}d_1(\rho,\sigma)^2\Big). \tag{199}$$

Remark 7.D.2. By taking advantage of the inequality $F(\rho,\sigma) \geq \mathrm{tr}(\rho\sigma)$, which follows from (176) and the norm inequality $\|A\|_1 \geq \|A\|_2$, one can establish another bound on $S(\rho||\sigma)$ in terms of the fidelity, which reads[152]

$$S(\rho||\sigma) \geq -S(\rho) - \ln F(\rho,\sigma). \tag{200}$$

Remark 7.D.3. The formula

$$F(\rho,\sigma) = \frac{1}{4}\inf_{H>0}\big\{\mathrm{tr}(H\rho)+\mathrm{tr}(H^{-1}\sigma)\big\}^2 = \inf_{H>0}\big\{\mathrm{tr}(H\rho)\,\mathrm{tr}(H^{-1}\sigma)\big\} \tag{201}$$

can be easily proven with the help of Lemma 6.C.3 and Theorem 7.B.2. The last expression is due to Alberti.[5]

Remark 7.D.4. We are now in position to show without much effort several results of Sec. V B.

(a) The upper bound (77) on the optimal success probability $P_{\mathrm{S,u}}^{\mathrm{opt}}$ in unambiguous discrimination of two mixed states can be established from Uhlmann's theorem, formula (70), and the fact that $P_{\mathrm{S,u}}^{\mathrm{opt}}(\{\rho_i,\eta_i\}) \leq P_{\mathrm{S,u}}^{\mathrm{opt}}(\{|\Psi_i\rangle,\eta_i\})$, where $|\Psi_i\rangle$ is a purification of ρ_i for any i.[135]

(b) It is instructive to derive in the special case of $m = 2$ states the lower bound on $P_{\mathrm{S,a}}^{\mathrm{opt}}$ given in Proposition 5.E.1 by using the Helstrom formula (68), the fact that $\mathrm{tr}(|\Lambda|) \geq \sum_i |\langle i|\Lambda|i\rangle|$ for any orthonormal basis $\{|i\rangle\}$, and Proposition 7.C.1 .[26]

(c) The Uhlmann theorem gives an efficient way to calculate the fidelity between the two states (78) (the result is $F(\rho_{\mathrm{eq}},\rho_{\mathrm{diff}}) = |\langle\psi_1|\psi_2\rangle|^2$).

E. Bures and quantum Hellinger metrics, quantum Fisher information

Recall that a Riemannian metric on $\mathcal{E}(\mathcal{H})$ is a map g which associates to each $\rho \in \mathcal{E}(\mathcal{H})$ a scalar product g_ρ on the tangent space to $\mathcal{E}(\mathcal{H})$ at ρ. For any state ρ on $\mathcal{H}$, this tangent space can be identified with the (real) vector space $\mathcal{B}(\mathcal{H})_{\text{s.a.}}$ of self-adjoint operators on $\mathcal{H}$. A metric g defines a Riemannian distance d, which is such that the square distance $\mathrm{d}s^2 = d(\rho, \rho + \mathrm{d}\rho)^2$ between two infinitesimally close states ρ and $\rho + \mathrm{d}\rho$ is given by

$$\mathrm{d}s^2 = g_\rho(\mathrm{d}\rho, \mathrm{d}\rho). \tag{202}$$

The Hilbert-Schmidt distance d_2 is obviously Riemannian: its metric is constant and given by the scalar product (1). In contrast, the trace distance d_1 is not Riemannian.

Let us show that the Bures distance d_{B} is Riemannian and determine its metric g_{B}. It is convenient to introduce a small parameter $t \in \mathbb{R}$. According to Definition 7.B.1 one has

$$d_{\text{B}}(\rho, \rho + t\,\mathrm{d}\rho)^2 = 2 - 2\,\mathrm{tr}(A(t)), \qquad A(t) = \left(\sqrt{\rho}(\rho + t\mathrm{d}\rho)\sqrt{\rho}\right)^{\frac{1}{2}}. \tag{203}$$

The scalar product $(g_{\text{B}})_\rho$ will be given in terms of the eigenvectors $|k\rangle$ and eigenvalues p_k of ρ in the spectral decomposition $\rho = \sum_k p_k |k\rangle\langle k|$. Using the notation $\dot{A}(t) = \mathrm{d}A/\mathrm{d}t$, $\ddot{A}(t) = \mathrm{d}^2A/\mathrm{d}t^2$, and the identity $A(t)^2 = \sqrt{\rho}(\rho + t\mathrm{d}\rho)\sqrt{\rho}$, one finds

$$\begin{aligned} \dot{A}(0)A(0) + A(0)\dot{A}(0) &= \sqrt{\rho}\,\mathrm{d}\rho\,\sqrt{\rho} \\ \ddot{A}(0)A(0) + 2\dot{A}(0)\dot{A}(0) + A(0)\ddot{A}(0) &= 0. \end{aligned} \tag{204}$$

The first equation yields

$$(p_k + p_l)\langle k|\dot{A}(0)|l\rangle = \sqrt{p_k p_l}\langle k|\mathrm{d}\rho|l\rangle. \tag{205}$$

Since $\mathrm{tr}(\mathrm{d}\rho) = 0$, it follows that $\mathrm{tr}[\dot{A}(0)] = 0$. Assume that $A(0) = \rho$ is invertible. Multiplying the second equation in (204) by $A(0)^{-1}$ and taking the trace, one verifies that

$$\mathrm{tr}[\ddot{A}(0)] = -\,\mathrm{tr}\big[\dot{A}(0)^2 A(0)^{-1}\big] = -\sum_{k,l=1}^{n} p_k^{-1}\big|\langle k|\dot{A}(0)|l\rangle\big|^2 = -\sum_{k,l=1}^{n} \frac{p_l|\langle k|\mathrm{d}\rho|l\rangle|^2}{(p_k + p_l)^2}. \tag{206}$$

Thus, going back to (203) we arrive at

$$d_{\text{B}}(\rho, \rho + t\mathrm{d}\rho)^2 = -\,\mathrm{tr}[\ddot{A}(0)]t^2 + \mathcal{O}(t^3) = (g_{\text{B}})_\rho(\mathrm{d}\rho, \mathrm{d}\rho)t^2 + \mathcal{O}(t^3) \tag{207}$$

with[84]

$$(g_{\text{B}})_\rho(A, A) = \frac{1}{2}\sum_{k,l=1}^{n} \frac{|\langle k|A|l\rangle|^2}{p_k + p_l}, \qquad A \in \mathcal{B}(\mathcal{H})_{\text{s.a.}},\ \rho > 0. \tag{208}$$

The last formula defines a scalar product on $\mathcal{B}(\mathcal{H})_{\text{s.a.}}$ by polarization, hence d_{B} is Riemannian with metric g_{B}. One readily obtains from this metric the infinitesimal volume element. The volume of $\mathcal{E}(\mathcal{H})$ and the area of its boundary are determined in Ref. 146.

Definition 7.E.1. *Given a state $\rho \in \mathcal{E}(\mathcal{H})$ and an observable $H \in \mathcal{B}(\mathcal{H})_{\text{s.a.}}$, the non-negative number*

$$\mathcal{F}_Q(\rho, H) = 4(g_{\text{B}})_\rho\big(-\mathrm{i}[H, \rho], -\mathrm{i}[H, \rho]\big) = 2\sum_{k,l,\,p_k+p_l>0} \frac{(p_k - p_l)^2}{p_k + p_l}|\langle k|H|l\rangle|^2 \tag{209}$$

is called the quantum Fisher information of ρ with respect to H.

The quantity $\mathcal{F}_Q(\rho, H)$ has been introduced by Braunstein and Caves[30] as a quantum analog of the Fisher information in statistics. Similarly to the definition of the Bures distance in Sec. VII B, these authors related it to the metric – called the "distinguishability metric" by Wootters[168] – extending the Fubini-Study metric to mixed states. For a pure state $\rho_\Psi = |\Psi\rangle\langle\Psi|$, the quantum Fisher information reduces to the square quantum fluctuation of H, namely,

$$\mathcal{F}_Q(\rho_\Psi, H) = 4\langle(\Delta H)^2\rangle_\Psi = 4\big(\langle\Psi|H^2|\Psi\rangle - \langle\Psi|H|\Psi\rangle^2\big). \tag{210}$$

In general, $\sqrt{\mathcal{F}_Q(\rho, H)}$ gives the speed at which a given state ρ separates from its time-evolved state $\rho(t) = e^{-\mathrm{i}tH}\rho e^{\mathrm{i}tH}$ under the dynamics specified by the Hamiltonian H. In fact, by plugging $\mathrm{d}\rho/\mathrm{d}t = -\mathrm{i}[H, \rho]$ into (207) one checks that

$$\sqrt{\mathcal{F}_Q(\rho, H)} = \left(2\frac{\mathrm{d}^2}{\mathrm{d}t^2}\, d_{\mathrm{B}}(\rho, \rho(t))^2\Big|_{t=0}\right)^{\frac{1}{2}} \approx \sqrt{2}\frac{\delta d_{\mathrm{B}}}{\delta t}\,. \tag{211}$$

We postpone the discussion on the statistical interpretation of $\mathcal{F}_Q(\rho, H)$ to Sec. VIII B below. It will be argued there that $\mathcal{F}_Q(\rho, H)$ measures the amount of quantum correlations in the state ρ that can be used for improving precision in quantum metrology.

Let us now turn to the quantum Hellinger distance (194). We proceed to determine the metric g_α associated to the normal-ordered relative Rényi entropy (142), from which the quantum Hellinger metric g_{H} is obtained by setting $\alpha = 1/2$. We demonstrate that the largest metric g_α for $\alpha \in (0, 1)$ is achieved for $\alpha = 1/2$ and is equal to $g_{\mathrm{H}}/2$, a result that will be needed later on (Sec. VIII A). The metric g_α is defined by

$$\begin{aligned} S_\alpha^{(\mathrm{n})}(\rho + t\mathrm{d}\rho||\rho) &= (1-\alpha)^{-1}\big(1 - F_\alpha^{(\mathrm{n})}(\rho + t\mathrm{d}\rho||\rho)^\alpha\big) + \mathcal{O}(t^3) \\ &= t^2(1-\alpha)^{-1}(g_\alpha)_\rho(\mathrm{d}\rho, \mathrm{d}\rho) + \mathcal{O}(t^3), \end{aligned} \tag{212}$$

where $F_\alpha^{(\mathrm{n})}$ is the α-fidelity, see (195). To determine g_α for all $\alpha \in (0, 1)$, we use (A1) in Appendix A to write

$$\begin{aligned} B_\alpha(t) = \rho^\alpha - (\rho + t\mathrm{d}\rho)^\alpha &= \frac{\sin(\alpha\pi)}{\pi}\int_0^\infty \mathrm{d}x\, x^\alpha\left(\frac{1}{x+\rho+t\mathrm{d}\rho} - \frac{1}{x+\rho}\right) \\ &= \frac{\sin(\alpha\pi)}{\pi}\int_0^\infty \mathrm{d}x\; x^\alpha\left(-\frac{t}{x+\rho}\mathrm{d}\rho\frac{1}{x+\rho} + \frac{t^2}{x+\rho}\mathrm{d}\rho\frac{1}{x+\rho}\mathrm{d}\rho\frac{1}{x+\rho}\right) + \mathcal{O}(t^3)\,. \end{aligned} \tag{213}$$

Introducing as before the spectral decomposition $\rho = \sum_k p_k|k\rangle\langle k|$ and using known integrals, one finds

$$\begin{aligned} 1 - F_\alpha^{(\mathrm{n})}(\rho + t\mathrm{d}\rho||\rho)^\alpha &= \mathrm{tr}[B_\alpha(t)\rho^{1-\alpha}] \\ &= -t\alpha\sum_{k=1}^n \langle k|\mathrm{d}\rho|k\rangle + t^2\sum_{k,l=1}^n \frac{p_k^{1-\alpha}(p_k^\alpha - p_l^\alpha)}{(p_k - p_l)^2}\big|\langle k|\mathrm{d}\rho|l\rangle\big|^2 + \mathcal{O}(t^3). \end{aligned} \tag{214}$$

Because $\mathrm{tr}(\mathrm{d}\rho) = 0$, the linear term in t vanishes as it should be. Plugging (214) into (212) one gets

$$(g_\alpha)_\rho(A, A) = \sum_{k,l=1}^n c_\alpha(p_k, p_l)|\langle k|A|l\rangle|^2, \quad c_\alpha(p, q) = \frac{(p^{1-\alpha} - q^{1-\alpha})(p^\alpha - q^\alpha)}{2(p-q)^2}. \tag{215}$$

It is easy to show that $c_\alpha(p, q) \le c_{1/2}(p, q)$ for any $p, q > 0$, hence

$$\max_{\alpha\in(0,1)} (g_\alpha)_\rho(A, A) = \big(g_{\frac{1}{2}}\big)_\rho(A, A) = \sum_{k,l=1}^n \frac{|\langle k|A|l\rangle|^2}{2(\sqrt{p_k} + \sqrt{p_l})^2}, \quad A \in \mathcal{B}(\mathcal{H})_{\mathrm{s.a.}}, \tag{216}$$

as claimed above. Furthermore, in view of (194) we deduce that the quantum Hellinger distance d_{H} is Riemannian and has a metric $g_{\mathrm{H}} = 2g_{1/2}$.

F. Characterization of the Riemannian contractive distances

The complete characterization of Riemannian contractive distances on $\mathcal{E}(\mathcal{H})$ for finite Hilbert spaces $\mathcal{H}$ has been given by Petz,[128] following a work by Morozova and Chentsov.[113] Such distances are induced by metrics g satisfying

$$g_{\mathcal{M}(\rho)}\big(\mathcal{M}(A), \mathcal{M}(A)\big) \le g_\rho(A, A), \quad A \in \mathcal{B}(\mathcal{H})_{\mathrm{s.a.}}, \tag{217}$$

for any $\rho \in \mathcal{E}(\mathcal{H})$ and any quantum operation $\mathcal{M} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$.

In the classical setting, it is remarkable that the contractivity condition leads to a unique metric (up to a multiplicative constant). Quantum operations correspond classically to Markov mappings $\mathbf{p} \mapsto \mathcal{M}^{\mathrm{clas}}\mathbf{p}$ on the probability simplex $\mathcal{E}_{\mathrm{clas}} = \{\mathbf{p} \in \mathbb{R}^n_+; \sum_i p_i = 1\}$, see (27), with stochastic matrices $\mathcal{M}^{\mathrm{clas}}$ having non-negative elements $\mathcal{M}^{\mathrm{clas}}_{ij}$ such that $\sum_i \mathcal{M}^{\mathrm{clas}}_{ij} = 1$ for any $j = 1, \ldots, n$. The contractive distances d^{clas} on $\mathcal{E}_{\mathrm{clas}}$ satisfy $d^{\mathrm{clas}}(\mathcal{M}^{\mathrm{clas}}\mathbf{p}, \mathcal{M}^{\mathrm{clas}}\mathbf{q}) \leq d^{\mathrm{clas}}(\mathbf{p}, \mathbf{q})$ for any such matrices. According to a result of Cencov,[35] a Riemannian distance on $\mathcal{E}_{\mathrm{clas}}$ with metric g^{clas} is contractive if and only if $g^{\mathrm{clas}}_{\mathbf{p}}(\mathbf{a}, \mathbf{a}) = c \sum_k a_k^2/p_k$ for any $\mathbf{a} \in \mathbb{R}^n$ and some $c > 0$, that is, the infinitesimal distance between a probability vector $\mathbf{p}$ and a neighboring vector $\mathbf{p} + \mathrm{d}\mathbf{p}$ is proportional to

$$\mathrm{d}s^2_{\mathrm{Fisher}} = \sum_{k=1}^{n} \frac{\mathrm{d}p_k^2}{p_k}. \tag{218}$$

The associated metric is known as the Fisher metric and plays an important role in statistics. It induces the Hellinger distance (190) up to a factor of one fourth.

Let us come back to the quantum case. Although g_ρ is in principle defined on the real vector space $\mathcal{B}(\mathcal{H})_{\mathrm{s.a.}}$ (the tangent space of $\mathcal{E}(\mathcal{H})$), one can extend it as a scalar product on the complex Hilbert space $\mathcal{B}(\mathcal{H})$. Without loss of generality, one may require that this scalar product satisfies

$$g_\rho(A, B) = g_\rho(B^*, A^*) = \overline{g_\rho(A^*, B^*)}, \quad A, B \in \mathcal{B}(\mathcal{H}). \tag{219}$$

(for instance, this is the case for the Hilbert-Schmidt product (1)). We first note that one can associate to g a family $\{\mathcal{K}_\rho; \rho \in \mathcal{E}(\mathcal{H})\}$ of positive operators on the Hilbert space $\mathcal{B}(\mathcal{H})$ endowed with the scalar product (1), by setting

$$g_\rho(A, B) = \langle A, \mathcal{K}_\rho^{-1}(B)\rangle, \quad A, B \in \mathcal{B}(\mathcal{H}). \tag{220}$$

Let us write $\rho_{\mathcal{M}} = \mathcal{M}(\rho)$. The monotonicity condition (217) reads $\mathcal{M}^*\mathcal{K}_{\rho_{\mathcal{M}}}^{-1}\mathcal{M} \leq \mathcal{K}_\rho^{-1}$, which means that $\mathcal{K}_\rho^{1/2}\mathcal{M}^*\mathcal{K}_{\rho_{\mathcal{M}}}^{-1}\mathcal{M}\mathcal{K}_\rho^{1/2}$ is a contraction. This is equivalent to $K_{\rho_{\mathcal{M}}}^{-1/2}\mathcal{M}\mathcal{K}_\rho\mathcal{M}^*\mathcal{K}_{\rho_{\mathcal{M}}}^{-1/2}$ being a contraction. Therefore, g is contractive if and only if

$$\mathcal{M}\mathcal{K}_\rho\mathcal{M}^* \leq \mathcal{K}_{\mathcal{M}(\rho)} \tag{221}$$

for any ρ and $\mathcal{M}$.

Lemma 7.F.1.[128] *The contractivity condition (221) is fulfilled by the positive operators*

$$\mathcal{K}_\rho = \mathcal{R}_\rho^{\frac{1}{2}} f(\Delta_\rho) \mathcal{R}_\rho^{\frac{1}{2}}, \tag{222}$$

where $\mathcal{R}_\rho$ stands for the right multiplication by ρ (see (7)), $\Delta_\rho = \Delta_{\rho|\rho}$ is the modular operator defined in (8), and $f : \mathbb{R}_+ \to \mathbb{R}$ is an operator monotone-increasing function with values in $\mathbb{R}_+$.

Proof. Let us recall that the modular operators Δ_ρ and $\Delta_{\rho_{\mathcal{M}}}$ on $\mathcal{B}(\mathcal{H})$ are (self-adjoint and) positive. In analogy with the proof of Theorem 6.B.1, we introduce the contraction $\mathcal{C}_{\mathcal{M}}$ defined by (127). It has been observed in this proof that $\mathcal{C}^*_{\mathcal{M}}\Delta_\rho\mathcal{C}_{\mathcal{M}} \leq \Delta_{\rho_{\mathcal{M}}}$. Since asking that a continuous function $f : \mathbb{R}_+ \to \mathbb{R}$ be operator monotone-increasing and non-negative is the same as asking that f be operator concave (see Appendix A and Ref. 27, Theorem V.2.5), it follows from the Jensen-type inequality (A4) and the monotonicity of f that

$$\mathcal{C}^*_{\mathcal{M}} f(\Delta_\rho)\mathcal{C}_{\mathcal{M}} \leq f(\Delta_{\rho_{\mathcal{M}}}). \tag{223}$$

Multiplying both sides by $B'\rho_{\mathcal{M}}^{\frac{1}{2}}$ and taking the scalar product by the same vector, this is equivalent to

$$\left\langle B',\ \mathcal{M}\mathcal{R}_\rho^{\frac{1}{2}} f(\Delta_\rho)\mathcal{R}_\rho^{\frac{1}{2}}\mathcal{M}^*(B')\right\rangle \leq \left\langle B',\ \mathcal{R}_{\rho_{\mathcal{M}}}^{\frac{1}{2}} f(\Delta_{\rho_{\mathcal{M}}})\mathcal{R}_{\rho_{\mathcal{M}}}^{\frac{1}{2}}(B')\right\rangle \tag{224}$$

for any $B' \in \mathcal{B}(\mathcal{H}')$. Thus the operator $\mathcal{K}_\rho$ defined in (222) satisfies the contractivity condition (221). □

Formulas (220) and (222) yield a family of monotonous metrics, in one-to-one correspondence with non-negative operator monotone functions f. These metrics are given by $g_\rho(A, B)$

$= \langle A\rho^{-\frac{1}{2}},\ f(\Delta_\rho)^{-1}(B\rho^{-\frac{1}{2}})\rangle$ for any $A, B \in \mathcal{B}(\mathcal{H})$. More explicitly, for any ρ with spectral decomposition $\rho = \sum_k p_k |k\rangle\langle k|$ one finds

$$g_\rho(A, A) = \sum_{k,l=1}^{n} c(p_k, p_l)|\langle k|A|l\rangle|^2, \quad A \in \mathcal{B}(\mathcal{H})_{\text{s.a.}}, \tag{225}$$

where $c(p, q)$ is given by

$$c(p, q) = \frac{pf(q/p) + qf(p/q)}{2pqf(p/q)f(q/p)} \tag{226}$$

and satisfies $c(tp, tq) = t^{-1}c(p, q)$ for any $t \in \mathbb{R}$, $t \neq 0$, and $c(p, p) = f(1)^{-1}p^{-1}$. By using $\Delta_\rho(B^*) = (\Delta_\rho^{-1}(B))^*$, it is easy to see that the condition (219) is satisfied if and only if $f(x) = xf(x^{-1})$. In particular, by choosing the following operator monotone functions (see Appendix A):

$$f_{\text{Harm}}(x) = \frac{2x}{x+1} \leq f_{\text{KM}}(x) = \frac{x-1}{\ln x} \leq f_{\text{H}} = \frac{(1+\sqrt{x})^2}{4} \leq f_{\text{B}}(x) = \frac{x+1}{2} \tag{227}$$

one is led to

$$c_{\text{Harm}}(p, q) = \frac{p+q}{2pq} \geq c_{\text{KM}}(p, q) = \frac{\ln p - \ln q}{p-q} \geq c_{\text{H}}(p, q) = \frac{4}{(\sqrt{p}+\sqrt{q})^2} \geq c_{\text{B}}(p, q) = \frac{2}{p+q}. \tag{228}$$

In view of (208) and (216), the last choice f_{B} gives the Bures metrics and f_{H} gives the Hellinger metric up to a factor of one fourth. The second choice corresponds to the so-called Kubo-Mori (or Bogoliubov) metric, which is associated to the relative von Neumann entropy. Actually, by substituting (215) into (212) and taking $\alpha \to 1$ one obtains

$$S(\rho + \mathrm{d}\rho||\rho) = \frac{1}{2}\sum_{k,l=1}^{n} c_{\text{KM}}(p_k, p_l)\big|\langle k|\mathrm{d}\rho|l\rangle\big|^2 = \frac{1}{2}g_{\text{KM}}(\mathrm{d}\rho, \mathrm{d}\rho). \tag{229}$$

According to the formula $S(\rho + t\mathrm{d}\rho) = S(\rho) - t\,\mathrm{tr}(\mathrm{d}\rho \ln \rho) - S(\rho + t\mathrm{d}\rho||\rho)$, one also gets

$$g_{\text{KM}}(\mathrm{d}\rho, \mathrm{d}\rho) = -\left.\frac{\mathrm{d}^2 S(\rho + t\mathrm{d}\rho)}{\mathrm{d}t^2}\right|_{t=0}, \tag{230}$$

S being the von Neumann entropy (since S is concave, the second derivative in the right-hand side is non-positive and defines a scalar product on $\mathcal{B}(\mathcal{H})$). As stressed by Balian, Alhassid, and Reinhardt,[13] this makes the Kubo-Mori metric quite natural from a physical viewpoint.

A result due to Kubo and Ando[96] states that there is a one-to-one correspondence between operator monotone functions f and operator means, that is, maps $m : (R, L) \in \mathcal{B}(\mathcal{H})_+ \times \mathcal{B}(\mathcal{H})_+ \mapsto m(R, L) \in \mathcal{B}(\mathcal{H})$ satisfying

(a) if $0 \leq R \leq T$ and $0 \leq L \leq N$ then $m(R, L) \leq m(T, N)$ (monotonicity);
(b) $C^* m(R, L)C \leq m(C^*RC, C^*LC)$.

This correspondence is given by the formula

$$m_f(R, L) = R^{\frac{1}{2}} f(R^{-\frac{1}{2}} L R^{-\frac{1}{2}}) R^{\frac{1}{2}}. \tag{231}$$

By taking f_{Harm} and f_{B} as in (227) one obtains the harmonic mean $m_{\text{Harm}}(R, L) = (R/2)^{-1} + (L/2)^{-1}$ and the arithmetic mean $m_{\text{B}}(R, L) = (R + L)/2$, respectively, and for $f(x) = \sqrt{x}$ one gets the so-called geometric mean (for more detail see, e.g., Ref. 33). The positive operators (222) can be written as

$$\mathcal{K}_\rho = m_f(\mathcal{R}_\rho, \mathcal{L}_\rho). \tag{232}$$

The theory of Kubo and Ando shows that the harmonic mean m_{Harm} and arithmetic mean m_{B} are respectively the smallest and largest symmetric operator means. Thus the Bures metric g_{B} is the smallest monotone metric among the family of metrics given by (220) and (222) with the

normalization $g_\rho(1,1) = \mathrm{tr}(\rho^{-1})$. It turns out that this family contains all contractive metrics, that is, all such metrics have the form (225).

Theorem 7.F.2. (Petz[128]) *The distances with metrics g given by (225) are contractive for any non-negative operator monotone-increasing function $f(x)$ satisfying $f(x) = xf(x^{-1})$. Conversely, any continuous metric g: $\rho \mapsto g_\rho$ on $\mathcal{E}(\mathcal{H})$ may be obtained from (225) by a choice of a suitable function f with these properties. In particular, there is a one-to-one correspondence between continuous contractive metrics satisfying $g_\rho(1,1) = \mathrm{tr}(\rho^{-1})$ and operator means. The Bures distance is the smallest of all contractive Riemannian distances with metrics satisfying this normalization condition.*

This theorem is of fundamental importance in geometrical approaches to quantum information.

Proof. The first statement has been proven above. Conversely, let g be a continuous contractive metric on $\mathcal{E}(\mathcal{H})$ and let us show that there exists an operator monotone function $f : \mathbb{R}_+ \to \mathbb{R}_+$ such that for any $\rho \in \mathcal{E}(\mathcal{H})$, g_ρ is given by (220) and (222) or, equivalently, by (225) and (226). We first note that g being contractive it is in particular unitary invariant, i.e., $g_{U^*\rho U}(U^*AU, U^*BU) = g_\rho(A, B)$ for any unitary U (see Sec. VII A). More generally, if the quantum operations $\mathcal{M}$ and $\mathcal{T}$ are such that ρ, A, and B are invariant under $\mathcal{T} \circ \mathcal{M}$, then $g_{\mathcal{M}(\rho)}(\mathcal{M}(A), \mathcal{M}(B)) = g_\rho(A, B)$. The main idea of the proof is to combine this invariance property with the uniqueness of the contractive classical distance. Denoting by $(g_\rho)_{ij,\,kl} = g_\rho(|i\rangle\langle j|, |k\rangle\langle l|)$ the matrix elements of the scalar product g_ρ in an orthonormal eigenbasis $\{|k\rangle\}$ of ρ, we need to prove that

$$(g_\rho)_{ij,kl} = \delta_{ik}\delta_{jl}\, c(p_i, p_j), \tag{233}$$

where δ_{ik} is the Kronecker symbol. To show that the matrix elements of g_ρ vanish for $i \neq j$ and $(k,l) \neq (i,j)$, it suffices to establish that

$$g_\rho\big(|i\rangle\langle j| + s|k\rangle\langle l|, |i\rangle\langle j| + s|k\rangle\langle l|\big) = g_\rho\big(|i\rangle\langle j| - s|k\rangle\langle l|, |i\rangle\langle j| - s|k\rangle\langle l|\big) \tag{234}$$

for $s = 1$ and $s = \mathrm{i}$ (the result then follows by polarization). If one of the indices i, j, k, and l is different from the three others, say $i \notin \{j, k, l\}$, this comes from the invariance of g under the unitary $U^{(i)} = \sum_k u_k^{(i)} |k\rangle\langle k|$ with $u_k^{(i)} = -1$ if $k = i$ and 1 otherwise. Hence $(g_\rho)_{ij,\,kl} = 0$ when $i \neq j$ and $(i,j) \neq (k,l), (l,k)$. Similarly, by choosing $u_k^{(i)} = \mathrm{i}$ if $i = k$ and 1 otherwise, this is also true for $i \neq j$ and $(i,j) = (l,k)$. The only non-vanishing matrix elements of g_ρ are thus $(g_\rho)_{ii,\,kk}$ and $(g_\rho)_{ij,\,ij}$ for $i \neq j$. To determine $(g_\rho)_{ii,\,kk}$ we observe that the restriction of g_ρ to the space of matrices commuting with ρ induces a contractive metric on the probability simplex $\mathcal{E}_{\mathrm{clas}}$, defined by $g_{\mathbf{p}}^{\mathrm{clas}}(\mathbf{a}, \mathbf{b}) = g_\rho(\sum_k a_k |k\rangle\langle k|, \sum_k b_k |k\rangle\langle k|)$ for any $\mathbf{a}, \mathbf{b} \in \mathcal{E}_{\mathrm{clas}}$. Indeed, one can associate a quantum operation $\mathcal{M}$ to a stochastic matrix $\mathcal{M}^{\mathrm{clas}}$ by defining $\mathcal{M}(|k\rangle\langle l|) = \delta_{kl} \sum_j \mathcal{M}_{jk}^{\mathrm{clas}} |j\rangle\langle j|$ ($\mathcal{M}$ has the Kraus form (31) as $\mathcal{M}_{jk}^{\mathrm{clas}} \geq 0$ and $\sum_j \mathcal{M}_{jk}^{\mathrm{clas}} = 1$ for any k). Then $\mathcal{M}(\rho) = \sum_j (\mathcal{M}^{\mathrm{clas}}\mathbf{p})_j |j\rangle\langle j|$ where $\mathbf{p}$ is the vector of eigenvalues of ρ, and (217) implies that g^{clas} is contractive under $\mathcal{M}^{\mathrm{clas}}$. According to the uniqueness of the contractive classical metrics, one has

$$(g_\rho)_{ii,kk} = g_{\mathbf{p}}^{\mathrm{clas}}(\delta_i, \delta_k) = c\,\frac{\delta_{ik}}{p_k}, \tag{235}$$

with $c > 0$ and $\delta_i = (\delta_{il})_{l=1}^n$. We now turn to the matrix elements $(g_\rho)_{ij,\,ij}$ for $i \neq j$. By unitary invariance, it is enough to determine $(g_\rho)_{12,\,12}$. To this end, we consider the quantum operations $\mathcal{M}$ from the space $\mathcal{B}(\mathcal{H})$ of $n \times n$ matrices to the space $\mathcal{B}(\mathbb{C}^3)$ of 3×3 matrices and $\mathcal{T} : \mathcal{B}(\mathbb{C}^3) \to \mathcal{B}(\mathcal{H})$ with Kraus operators $\{A_i\}_{i=2}^n$ and $\{B_i\}_{i=2}^n$, respectively, given by

$$A_2 = B_2 = |1\rangle\langle 1| + |2\rangle\langle 2|, \quad A_i = |3\rangle\langle i|, \quad B_i = \frac{\sqrt{p_i}}{\sqrt{1-p_1-p_2}}|i\rangle\langle 3|, \quad i = 3, \ldots, n\,. \tag{236}$$

A simple calculation yields $\mathcal{T} \circ \mathcal{M}(\rho) = \rho$. As stressed above, one can deduce from the contractivity of g_ρ that $(g_\rho)_{12,12} = (g_{\mathcal{M}(\rho)})_{12,12}$, thereby showing that this matrix element depends on p_1 and p_2 only. By unitary invariance, $(g_\rho)_{ij,\,ij}$ only depends on p_i and p_j and one can set $(g_\rho)_{ij,\,ij} = c(p_i, p_j)$ for $i \neq j$, $c(p,q)$ being independent of ρ. This complete the proof of (233), excepted that it remains to justify that $c(p,p) = c/p$.

We proceed by showing that $c(q, p)$ is given by (226) with f having the desired properties. Thanks to (219), we know that $c(p, q)$ is real and symmetric. One verifies that $c(p, p) = c/p$ by the following argument. Let us assume that ρ has a degenerate eigenvalue, say $p_1 = p_2$. Then $\rho = U\rho U^*$ for any unitary U acting trivially on span$\{|3\rangle, \ldots, |n\rangle\}$. By unitary invariance, $g_\rho(|\psi\rangle\langle\psi|, |\psi\rangle\langle\psi|) = (g_\rho)_{11,11} = c/p_1$ for any $|\psi\rangle \in$ span$\{|1\rangle, |2\rangle\}$. Taking, e.g., $|\psi\rangle = (|1\rangle + |2\rangle)/\sqrt{2}$ and using (233), we get $(g_\rho)_{12,12} = c(p_1, p_1) = c/p_1$. In order to establish that $c(p, q)$ is homogeneous we consider the quantum operations $\mathcal{M} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H} \otimes \mathcal{H}_\mathsf{E})$ and $\mathcal{T} : \mathcal{B}(\mathcal{H} \otimes \mathcal{H}_\mathsf{E}) \to \mathcal{B}(\mathcal{H})$ defined by $\mathcal{M}(\rho) = \rho \otimes 1/n_\mathsf{E}$ and $\mathcal{T}(\widehat{\rho}) = \mathrm{tr}_\mathsf{E}(\widehat{\rho})$ (here n_E is the dimension of $\mathcal{H}_\mathsf{E}$). Clearly, $\mathcal{T} \circ \mathcal{M} = 1$, thus by similar arguments as above and by taking advantage of (233), one finds

$$c(p_i, p_j) = (g_\rho)_{ij,ij} = g_{\mathcal{M}(\rho)}(\mathcal{M}(|i\rangle\langle j|), \mathcal{M}(|i\rangle\langle j|)) = n_\mathsf{E}^{-1} c\Big(\frac{p_i}{n_\mathsf{E}}, \frac{p_j}{n_\mathsf{E}}\Big). \tag{237}$$

As this is true for any positive integer n_E and any state ρ, one concludes that $c(tp, tq) = t^{-1}c(p, q)$ for all $p, q \in [0, 1]$ and all rationals t with $tp, tq \in [0, 1]$. This is the point where we need the continuity of the metric to make sure that $c(p, q)$ is continuous. Then the equality holds for all real t. Setting $f(x) = 1/c(x, 1)$ and using the symmetry of $c(p, q)$, one easily derives the identities (226) and $f(x^{-1}) = x^{-1}f(x)$. Furthermore, $f(1)^{-1} = c(1, 1) = c$.

To complete the proof, we have to show that f is operator concave. With this aim, let us consider the inequality (221) which is equivalent to g_ρ being contractive. We choose $\mathcal{M}$ in this inequality to be the partial trace operation $\mathcal{T} : \widehat{\rho} \mapsto \mathrm{tr}_{\mathbb{C}^2}(\widehat{\rho}) \otimes 1/2$ on $\mathcal{B}(\mathcal{H} \otimes \mathbb{C}^2)$ and $\widehat{\rho} = (\rho_0 \otimes |0\rangle\langle 0| + \rho_1 \otimes |1\rangle\langle 1|)/2$. From (221) we find that for any $A \in \mathcal{B}(\mathcal{H})$,

$$\big\langle \mathcal{T}^*(A \otimes 1),\ \mathcal{K}_{\widehat{\rho}}\mathcal{T}^*(A \otimes 1)\big\rangle \le \big\langle A \otimes 1,\ \mathcal{K}_{\mathcal{T}(\widehat{\rho})}(A \otimes 1)\big\rangle\,. \tag{238}$$

But $\mathcal{K}_{\widehat{\rho}}(A \otimes 1) = (\mathcal{K}_{\rho_0}(A) \otimes |0\rangle\langle 0| + \mathcal{K}_{\rho_1}(A) \otimes |1\rangle\langle 1|)/2$. Accordingly, (238) reduces to

$$\frac{1}{2}\big\langle A,\ (\mathcal{K}_{\rho_0} + \mathcal{K}_{\rho_1})A\big\rangle \le \big\langle A,\ \mathcal{K}_{(\rho_0+\rho_1)/2}A\big\rangle\,, \tag{239}$$

thereby showing that the map

$$\rho \mapsto \mathcal{K}_\rho = f(\mathcal{L}_\rho\mathcal{R}_\rho^{-1})\mathcal{R}_\rho \tag{240}$$

is mid-point concave. By a standard argument based on a dyadic decomposition, it follows that this map is concave.[33] Using the $*$-isomorphism between the C^*-algebras $\mathcal{B}(\mathcal{B}(\mathcal{H}))$ and $\mathcal{B}(\mathcal{H} \otimes \mathcal{H})$ (Sec. II A), this is equivalent to say that the map

$$A \mapsto f\big(A \otimes (A^T)^{-1}\big)\, 1 \otimes A^T \tag{241}$$

is concave. One easily deduces from this that the map $(A, B) \mapsto f(A \otimes (B^T)^{-1})\, 1 \otimes B^T$ is jointly concave. In particular, $A \mapsto f(A)$ is concave. This shows that f is operator concave. □

VIII. STATE DISCRIMINATION AND PARAMETER ESTIMATION IN LARGE SYSTEMS

In this section we examine two problems related to the state discrimination talk discussed in Sec. V, namely, the quantum hypothesis testing and parameter estimation. In the first problem, one wants to determine asymptotically the probability of error in discriminating two states when one has N independent copies of those states, for $N \to \infty$. In the second problem, the goal is to estimate as precisely as possible a real parameter from measurements performed on a large number of particles in a state depending smoothly on this parameter.

A. Quantum hypothesis testing: Discriminating two states from many identical copies

An important issue in classical information theory is to discriminate two probability measures $\mathbf{p}_1$ and $\mathbf{p}_2$ on a measurable space $(\Omega, \mathcal{F})$, given the outcomes of N independent identically distributed (i.i.d.) random variables, whose law is either $\mathbf{p}_1$ or $\mathbf{p}_2$. Since one has to decide among two hypothesis – the first (second) one being that the observed data are distributed according to $\mathbf{p}_1$ ($\mathbf{p}_2$) – this discrimination task bears the name of "hypothesis testing." For a given test function, i.e., a random

variable $M_{\rm clas}$ with values in [0, 1], the probability of error is $P_{{\rm err},N} = \eta_1 \mathbf{p}_1^{(N)}(M_{\rm clas}) + \eta_2 \mathbf{p}_2^{(N)}(1 - M_{\rm clas})$, where $\mathbf{p}_i^{(N)} = \mathbf{p}_i^{\otimes N}$ is the N-fold product measure and η_i the prior probability attached to $\mathbf{p}_i$. It is easy to convince oneself that the minimal error is achieved for the maximum likelihood test function defined by

$$M_{\rm clas}^{\rm opt} = 1_{\{\eta_2 \rho_2^{(N)} - \eta_1 \rho_1^{(N)} \geq 0\}}\,, \tag{242}$$

$\rho_i^{(N)} = {\rm d}\mathbf{p}_i^{(N)}/{\rm d}\boldsymbol{\mu}^{(N)}$ being the density of $\mathbf{p}_i^{(N)}$ with respect to the measure $\boldsymbol{\mu}^{(N)} = \mathbf{p}_1^{(N)} + \mathbf{p}_2^{(N)} = \boldsymbol{\mu}^{\otimes N}$. Here 1_A stands for the indicator function on $A \subset \Omega$, i.e., $1_A(\omega) = 1$ if $\omega \in A$ and 0 otherwise. The corresponding error is

$$\begin{aligned} P_{{\rm err},N}^{\rm opt}(\{\mathbf{p}_i^{(N)}, \eta_i\}) &= \min_{0 \leq M_{\rm clas} \leq 1} \left\{ \int_{\Omega^N} {\rm d}\boldsymbol{\mu}^{(N)} \left(\eta_1 \rho_1^{(N)} M_{\rm clas} + \eta_2 \rho_2^{(N)} (1 - M_{\rm clas}) \right) \right\} \\ &= \int_{\Omega^N} {\rm d}\boldsymbol{\mu}^{(N)} \min\left\{ \eta_1 \rho_1^{(N)}, \eta_2 \rho_2^{(N)} \right\}. \end{aligned} \tag{243}$$

One is typically interested in the limit of a large number of tests, i.e., $N \to \infty$. One can show that the error probability decays exponentially like $P_{{\rm err},N}^{\rm opt} \sim e^{-N\xi(\mathbf{p}_1, \mathbf{p}_2)}$, with an exponent given by the Chernoff bound[40]

$$\xi(\mathbf{p}_1, \mathbf{p}_2) = -\lim_{N\to\infty} \frac{1}{N} \ln P_{{\rm err},N}^{\rm opt}(\{\mathbf{p}_i^{(N)}, \eta_i\}) = -\inf_{\alpha \in (0,1)} \left\{ \ln \left(\int_\Omega {\rm d}\boldsymbol{\mu}\, \rho_1^\alpha \rho_2^{1-\alpha} \right) \right\}, \tag{244}$$

where we have set $\rho_i = \rho_i^{(1)}$. One recognizes in the infimum in the right-hand side the classical Rényi divergence (148) multiplied by $(\alpha - 1)$.

In quantum mechanics, the hypothesis testing can be rephrased as the discrimination of two N-fold tensor product states $\rho_1^{\otimes N}$ and $\rho_2^{\otimes N}$. The corresponding minimal error probability is given by the Helstrom formula (68),

$$P_{{\rm err},N}^{\rm opt}(\{\rho_i^{\otimes N}, \eta_i\}) = \frac{1}{2}\big(1 - {\rm tr}\,|\Lambda_N|\big), \quad \Lambda_N = \eta_1 \rho_1^{\otimes N} - \eta_2 \rho_2^{\otimes N}\,, \tag{245}$$

and the optimal measurement consists of the orthogonal projectors $M_\pm^{\rm opt}$ on the supports of the positive and negative parts of Λ_N. Note that if ρ_1 and ρ_2 commute then $M_-^{\rm opt}$ can be identified with the maximum likelihood test function and one recovers the classical formula (243) from (245). Surprisingly, the generalization of the Chernoff bound (244) to the quantum setting has been settled out only recently. It has been highlighted in Sec. VI C that the Rényi divergences appearing in this bound have several natural quantum extensions, according to the choice of operator ordering. It was proven by Audenaert *et al.*[7] and by Nussbaum and Szkola[118] that the right extension is the normal-ordered relative Rényi entropy $S_\alpha^{(\rm n)}(\rho||\sigma)$ defined in (142).

Proposition 8.A.1. (Quantum Chernoff bound[7,118]) *One has*

$$-\lim_{N\to\infty} \frac{1}{N} \ln P_{{\rm err},N}^{\rm opt}(\{\rho_i^{\otimes N}, \eta_i\}) = -\inf_{\alpha\in(0,1)} \left\{ \ln\big({\rm tr}[\rho_1^\alpha \rho_2^{1-\alpha}]\big) \right\} = \sup_{\alpha\in(0,1)} \left\{ (1-\alpha) S_\alpha^{(\rm n)}(\rho_1||\rho_2) \right\}. \tag{246}$$

This limit defines a jointly convex function $\xi_Q(\rho_1, \rho_2)$ with values in $\mathbb{R}_+ \cup \{+\infty\}$, which is contractive under quantum operations. Moreover, ξ_Q induces the quantum Hellinger metric up to a factor of one half, that is, if ρ and $\rho + d\rho$ are infinitesimally close then $\xi_Q(\rho + {\rm d}\rho, \rho) = g_{\rm H}(d\rho, {\rm d}\rho)/2$ is given by (216).

The infimum in (246) is attained for a unique $\alpha \in (0, 1)$ satisfying ${\rm tr}(\rho_1^\alpha \rho_2^{1-\alpha}(\ln \rho_1 - \ln \rho_2)) = 0$.[7] Actually, for any fixed ρ and σ, the function $\alpha \mapsto F_\alpha^{(\rm n)}(\rho||\sigma)^\alpha = {\rm tr}[\rho^\alpha \sigma^{1-\alpha}]$ is convex (this is a simple consequence of the convexity of $\alpha \mapsto p^\alpha q^{1-\alpha}$ for $p, q > 0$) and $F_\alpha^{(\rm n)}(\rho||\sigma) \leq F_{0,1}^{(\rm n)}(\rho||\sigma) = 1$ by the Hölder inequality (3). Before entering into the proof, let us also mention that $\xi_Q(\rho, \sigma) < \infty$ whenever ρ and σ do not have orthogonal supports. If $\rho = |\psi\rangle\langle\psi|$ is pure, the quantum Chernoff bound is related to the fidelity by $\xi_Q(\rho, \sigma) = -\ln F(\rho, \sigma) = -\ln \langle\psi|\sigma|\psi\rangle$ (in fact, then $F_\alpha^{(\rm n)}(\rho||\sigma)^\alpha = \langle\psi|\sigma^{1-\alpha}|\psi\rangle$ is minimum for $\alpha = 0$).

Proof. To shorten notation we write $P^{\rm opt}_{{\rm eff},N}$ when referring to $P^{\rm opt}_{{\rm err},N}(\{\rho^{\otimes N},\eta,\sigma^{\otimes N},1-\eta\})$. The fact that

$$\limsup_{N\to\infty}\frac{1}{N}\ln P^{\rm opt}_{{\rm eff},N}\le -\xi_Q(\rho,\sigma)=\inf_{\alpha\in(0,1)}\left\{\ln({\rm tr}[\rho^\alpha\sigma^{1-\alpha}])\right\} \tag{247}$$

follows from (245) and the trace inequality

$$\frac{1}{2}\Big({\rm tr}(A)+{\rm tr}(B)-{\rm tr}\,|A-B|\Big)\le {\rm tr}(A^\alpha B^{1-\alpha})\,, \tag{248}$$

where A and B are non-negative operators and $\alpha\in[0,1]$. This inequality has been first established in Ref. 7. A simple proof due to N. Ozawa is reported in Appendix B. The reverse inequality to (247) is a consequence of the classical Chernoff bound. This can be justified as follows.[118] Let us observe that the optimal measurement is a von Neumann measurement $\{\Pi^{\rm opt},1-\Pi^{\rm opt}\}$ with $\Pi^{\rm opt}$ a projector, so that

$$\begin{aligned}P^{\rm opt}_{{\rm err},N}=1-P^{\rm opt}_{{\rm S},N}&=\eta\,{\rm tr}\big((1-\Pi^{\rm opt})\rho^{\otimes N}\big)+(1-\eta)\,{\rm tr}\big(\Pi^{\rm opt}\sigma^{\otimes N}\big)\\&=\sum_{\underline{k},\underline{l}}\Big(\eta p_{\underline{k}}\big|\langle\Phi_{\underline{l}}|(1-\Pi^{\rm opt})|\Psi_{\underline{k}}\rangle\big|^2+(1-\eta)q_{\underline{l}}\big|\langle\Psi_{\underline{k}}|\Pi^{\rm opt}|\Phi_{\underline{l}}\rangle\big|^2\Big),\end{aligned} \tag{249}$$

where $\{|\Psi_{\underline{k}}\rangle\}$ and $\{|\Phi_{\underline{l}}\rangle\}$ are orthonormal eigenbases of $\rho^{\otimes N}$ and $\sigma^{\otimes N}$, respectively, and $p_{\underline{k}}$ and $q_{\underline{l}}$ are the corresponding eigenvalues. We may without loss of generality assume that $\eta\le 1/2$. By using the inequality $|a|^2+|b|^2\ge|a+b|^2/2$ one gets

$$P^{\rm opt}_{{\rm err},N}\ge\eta\sum_{\underline{k},\underline{l}}\frac{1}{2}\min\{p_{\underline{k}},q_{\underline{l}}\}\big|\langle\Phi_{\underline{l}}|\Psi_{\underline{k}}\rangle\big|^2. \tag{250}$$

But $\rho^{\otimes N}$ corresponds to N independent copies of the state $\rho=\sum_k p_k|\psi_k\rangle\langle\psi_k|$, hence its eigenvalues $p_{\underline{k}}$ and eigenvectors $|\Psi_{\underline{k}}\rangle$ are products of N eigenvalues p_k and N eigenvectors $|\psi_k\rangle$ of ρ, respectively, and similarly for $\sigma^{\otimes N}$ with the eigenvalues q_l and eigenvectors $|\phi_l\rangle$ of σ. This means that $p_{\underline{k}}|\langle\Phi_{\underline{l}}|\Psi_{\underline{k}}\rangle|^2$ can be viewed as the N-fold product of the probability $\boldsymbol{\pi}_1$ on $\{1,\ldots,n\}^2$ defined by $(\boldsymbol{\pi}_1)_{kl}=p_k|\langle\phi_l|\psi_k\rangle|^2$. Analogously, $q_{\underline{l}}|\langle\Phi_{\underline{l}}|\Psi_{\underline{k}}\rangle|^2$ is the N-fold product of $\boldsymbol{\pi}_2$ with $(\boldsymbol{\pi}_2)_{kl}=q_l|\langle\phi_l|\psi_k\rangle|^2$. Consequently, the sum in (250) is the minimal error probability $P^{\rm opt}_{{\rm err},N}(\{\boldsymbol{\pi}_i^{(N)},1/2\})$ for discriminating $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ with equal prior probabilities (see (243)). One then deduces from the classical Chernoff bound (244) that

$$\liminf_{N\to\infty}\frac{1}{N}\ln P^{\rm opt}_{\rm err}\ge\inf_{\alpha\in(0,1)}\left\{\ln\left(\sum_{k,l=1}^n(\boldsymbol{\pi}_1)_{kl}^\alpha(\boldsymbol{\pi}_2)_{kl}^{1-\alpha}\right)\right\}=-\xi_Q(\rho,\sigma). \tag{251}$$

Together with (247) this proves the quantum Chernoff bound.

It is nevertheless instructive to show (251) directly from (250), without relying on the classical result, by using the theory of large deviations for sums of i.i.d. random variables and the relative modular operator $\Delta_{\sigma|\rho}$ (see Sec. II), which appears here quite naturally.[87] Indeed, let us set $\xi=\rho^{\frac{1}{2}}$ and note that for any real function $f:(0,\infty)\to\mathbb{R}$, according to (8) and by the functional calculus, it holds

$$\langle\xi,\,f(\Delta_{\sigma|\rho})\,\xi\rangle=\sum_{k,l=1}^n p_k f\Big(\frac{q_l}{p_k}\Big)|\langle\phi_l|\psi_k\rangle|^2\,. \tag{252}$$

In particular, $\langle\xi,\,\ln(\Delta_{\sigma|\rho})\,\xi\rangle={\rm tr}[\rho(\ln\sigma-\ln\rho)]=-S(\rho||\sigma)$, as already observed in Sec. VI B. Let $\mathbf{m}_{\sigma|\rho}$ be the spectral measure of $-\ln\Delta_{\sigma|\rho}$ with respect to the vector ξ. This is a probability measure (ξ is normalized), which is related to the relative entropy by $S(\rho||\sigma)=\int{\rm d}\mathbf{m}_{\sigma|\rho}(t)\,t$. Taking $f(x)=\min\{x,1\}=g(-\ln x)$ with $g(t)=\min\{e^{-t},1\}$ in (252), one finds

$$\sum_{k,l=1}^n\min\{p_k,q_l\}|\langle\phi_l|\psi_k\rangle|^2=\big\langle\xi,\,g(-\ln\Delta_{\sigma|\rho})\,\xi\big\rangle=\int_{\mathbb{R}}{\rm d}\mathbf{m}_{\sigma|\rho}(t)\,g(t)\;\ge\;\mathbf{m}_{\sigma|\rho}(\mathbb{R}_-). \tag{253}$$

A similar inequality holds for the sum in the right-hand side of (250): it suffices to substitute $\Delta_{\sigma|\rho}$ by $\Delta_{\sigma^{\otimes N}|\rho^{\otimes N}} = \Delta_{\sigma|\rho}^{\otimes N}$. The spectral measure of $-\ln \Delta_{\sigma|\rho}^{\otimes N}$ is a product measure $\mathbf{m}_{\sigma|\rho}^{(N)}$ and thus $-\ln \Delta_{\sigma|\rho}^{\otimes N}$ can be interpreted as a sum of i.i.d. random variables $-\ln \Delta_{\sigma|\rho}^{(\nu)}$ with law $\mathbf{m}_{\sigma|\rho}$. The large deviation principle ensures that if $e'_{\sigma|\rho}(0) < \theta < e'_{\sigma|\rho}(1)$ then[53]

$$\lim_{N\to\infty} \frac{1}{N} \ln\left(\mathbf{m}_{\sigma|\rho}^{(N)}\left(-\sum_{\nu=1}^{N} \ln \Delta_{\sigma|\rho}^{(\nu)} \le -\theta N\right)\right) = -\sup_{\alpha\in[0,1]} \left\{\alpha\theta - e_{\sigma|\rho}(\alpha)\right\} \tag{254}$$

is up to a minus sign the Legendre transform of

$$e_{\sigma|\rho}(\alpha) = \ln\left(\int_{\mathbb{R}} \mathrm{d}\mathbf{m}_{\sigma|\rho}(t) e^{-t\alpha}\right) = \ln\left(\langle \xi,\ \Delta_{\sigma|\rho}^{\alpha}\, \xi\rangle\right) = \ln\left(\mathrm{tr}[\rho^{1-\alpha}\sigma^{\alpha}]\right) . \tag{255}$$

If $\rho \neq \sigma$ then $e'_{\sigma|\rho}(0) = -S(\rho||\sigma) < 0$ and $e'_{\sigma|\rho}(1) = S(\sigma||\rho) > 0$ (the second identity follows from the first one by symmetry $e_{\sigma|\rho}(1-\alpha) = e_{\rho|\sigma}(\alpha)$). Thus the large deviation bound (254) holds for $\theta = 0$. Taking advantage of (250) and (253) one is led to

$$\begin{aligned} \liminf_{N\to\infty} \frac{1}{N} \ln P_{\mathrm{err},N}^{\mathrm{opt}} &\ge \liminf_{N\to\infty} \frac{1}{N} \ln\left(\sum_{\underline{k},\underline{l}} \min\{p_{\underline{k}}, q_{\underline{l}}\} \left|\langle \Phi_{\underline{l}} | \Psi_{\underline{k}}\rangle\right|^2\right) \\ &\ge \lim_{N\to\infty} \frac{1}{N} \ln \mathbf{m}_{\sigma|\rho}^{(N)}\left(-\sum_{\nu=1}^{N} \ln \Delta_{\sigma|\rho}^{(\nu)} \le 0\right) = \inf_{\alpha\in[0,1]} \left\{e_{\sigma|\rho}(\alpha)\right\} = -\xi_Q(\rho,\sigma), \end{aligned} \tag{256}$$

in agreement with (251). Note that these arguments justify in particular that the second member in the classical Chernoff bound (244) is bounded from above by the third one, as a consequence of the large deviation principle. Applying (247) for commuting matrices ρ and σ, this gives a full proof of this classical bound.

The joint convexity of $\xi_Q(\rho,\sigma)$ mentioned in the proposition results from the joint convexity of the relative entropies $S_{\alpha}^{(\mathrm{n})}(\rho||\sigma)$ for $\alpha \in (0,1)$, which follows from the Lieb concavity theorem, see Sec. VI C. One then gets the contractivity of ξ_Q with respect to quantum operations from Proposition 6.B.2. This concludes the proof. □

Remark 8.A.2. The quantum Chernoff bound (246) can be generalized to the case where the two states $\rho_{i,N} \in \mathcal{E}(\mathcal{H}^{\otimes N})$ to discriminate are not product states (i.e., for dependent copies).

Actually, the large deviation principle used in the proof is not restricted to sums of i.i.d. random variables. It must be assumed that the limit $e(\alpha) = \lim_{N\to\infty} N^{-1} \ln \mathrm{tr}[\rho_{1,N}^{\alpha} \rho_{2,N}^{1-\alpha}]$ exists, is continuous in α on $[0,1]$ and differentiable on $(0,1)$, and its right derivative $e'(0)$ is smaller than its left derivative $e'(1)$ (see Ref. 87).

Remark 8.A.3. In asymmetric hypothesis testing one is interested by the minimal error probability of identifying the second state under the constraint that the error on the identification of the first state is smaller than ε,

$$P_{\mathrm{err},N,\varepsilon}^{\mathrm{asym}} = \min_{0\le M\le 1} \left\{\mathrm{tr}[M\rho_2^{\otimes N}]\,;\ \mathrm{tr}[(1-M)\rho_1^{\otimes N}] \le \varepsilon\right\} . \tag{257}$$

The quantum Stein's lemma[76,121] shows that this probability decays exponentially with a rate given by the relative von Neumann entropy, i.e.,

$$-\lim_{N\to\infty} \frac{1}{N} \ln P_{\mathrm{err},N,\varepsilon}^{\mathrm{asym}} = S(\rho_1||\rho_2) . \tag{258}$$

The limit one gets by replacing the fixed parameter $\varepsilon > 0$ by e^{-rN} (that is, asking for an exponentially decaying error on the identification of ρ_1) is, in turn, given by the Hoeffding bound (see, e.g., Ref. 87 for more detail).

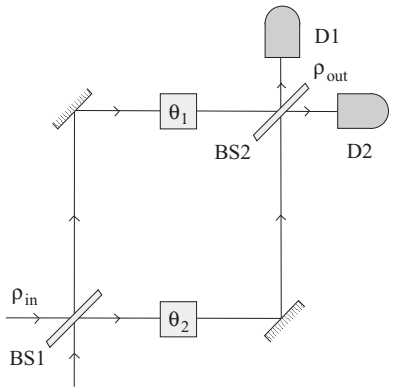

FIG. 2. In a Mach-Zehnder interferometer, the light entering in one of the two input modes is split into two beams by a beam splitter (represented by the rectangle BS1 inclined by 45°). The photons in the first and second beams acquire some phase shifts θ_1 and θ_2, respectively. They then go through a second beam splitter (rectangle BS2) and finally into the detectors D1 and D2, which count the number of photons in the two output modes.

An interesting link between the quantum hypothesis testing and fluctuation theorems in quantum statistical physics has been found by Jakšić *et al.*[87] They have shown that the quantum Chernoff bound for discriminating the forward and backward time-evolved states $\rho_{\pm T/2}$ as $T \to \infty$ appears in the large deviation principle for the full counting statistics of measurements of the energy/entropy flow over the time interval $[0, T]$.

B. Parameter estimation in quantum metrology

The parameter estimation problem is a kind of continuous version of quantum state discrimination, in which the system state $\rho(\theta)$ depends on a continuous parameter θ. One aims at estimating this unknown parameter with the highest possible precision $\Delta\theta$ by performing measurements on $\rho(\theta)$. This precision is limited by our ability to distinguish the states $\rho(\theta)$ for values of θ differing by $\Delta\theta$.

1. Phase estimation in Mach-Zehnder interferometers

An important example is phase estimation in the Mach-Zehnder interferometer represented in Fig. 2. An input photon passes through a beam splitter[28] which transforms its state into a superposition of two modes propagating along different paths. These two modes acquire distinct phases θ_1 and θ_2 during the propagation and are finally recombined in a second beam splitter to read out interference fringes, from which the phase difference $\theta = \theta_1 - \theta_2$ is inferred. The interferometric sequence can be described by means of rotation matrices acting on the two-mode photon state. We shall assume at this point that the reader is familiar with second quantization (a good mathematical introduction to this formalism can be found in Ref. 29.) The generators of the aforementioned rotations are the angular momentum operators J_x, J_y, and J_z related to the bosonic annihilation and creation operators b_j and b_j^* of a photon in mode $j = 1, 2$ by $J_x = (b_1^* b_2 + b_2^* b_1)/2$, $J_y = -\mathrm{i}(b_1^* b_2 - b_2^* b_1)/2$, and $J_z = (b_1^* b_1 - b_2^* b_2)/2$ (Schwinger representation). These operators act on the bosonic Fock space $\mathcal{F}_\mathrm{b}(\mathbb{C}^2)$ associated to the single photon space $\mathcal{H} \simeq \mathbb{C}^2$. The output state of the interferometer is given in terms of the input state ρ_in by[172]

$$\rho_\mathrm{out}(\theta) = e^{-\mathrm{i}\theta J_\mathbf{n}} \rho_\mathrm{in} e^{\mathrm{i}\theta J_\mathbf{n}} , \tag{259}$$

where θ is the phase to be estimated and $J_\mathbf{n} = n_x J_x + n_y J_y + n_z J_z$ the angular momentum in the direction specified by the unit vector $\mathbf{n} \in \mathbb{R}^3$.

One can also realize a Mach-Zehnder interferometer with ultracold atoms forming a Bose-Einstein condensate in an optical trap, instead of photons. Then the two modes correspond to two distinct atomic energy levels and the total number of atoms $N_\mathrm{p} = N_1 + N_2$ in these modes is fixed.

In such a case the Hilbert space of the system has finite dimension $N_{\rm p}+1$ (one deals here with indistinguishable particles). Atom interferometry in Bose-Einstein condensates is very promising due to the tunable interactions between atoms, which make it possible to generate dynamically entangled states involving a large number of particles (in contrast, because of the absence of direct interactions between photons it is difficult to generate large numbers of photons having multipartite entanglement). We will see below that using such entangled states as inputs leads to smaller errors $\Delta\theta$ in the phase estimation than for separable inputs. For independent (i.e., separable) particles the precision is of the order of $(\Delta\theta)_{\rm SN}\approx 1/\sqrt{N_{\rm p}}$ (shot noise limit). Higher precisions than $(\Delta\theta)_{\rm SN}$ have been reported experimentally.[65,133] Important potential applications of these ultra-precise interferometers include atomic clocks and magnetic sensors with enhanced sensitivities.[163,140]

2. Quantum Cramér-Rao bound

In the more general setting, the problem of estimating an unknown parameter θ from a θ-dependent state evolution and measurements on the output states can be described as follows. For simplicity we assume that the evolution is given by a self-adjoint operator H (equal to $J_{\mathbf{n}}$ in the above Mach-Zehnder interferometer), i.e.,

$$\rho(\theta)=e^{-{\rm i}\theta H}\,\rho\,e^{{\rm i}\theta H}\ , \tag{260}$$

where $\rho=\rho(0)=\rho_{\rm in}$ is the input state. One performs generalized measurements given by a POVM $\{M_i\}_{i=1}^m$ on the output state $\rho(\theta)=\rho_{\rm out}$. The probability to get the outcome i is $p_{i|\theta}={\rm tr}[M_i\,\rho(\theta)]$ (Sec. III C). After N independent measurements on copies of $\rho(\theta)$ yielding the outcomes $i_1, i_2, \ldots, i_N$, the parameter θ is estimated by using a statistical estimator depending on these outcomes, that is, a function $\theta_{\rm est}(i_1, i_2, \ldots, i_N)$ (in practice the experiment is repeated N times, starting from the same initial state ρ and in similar conditions, so that the quantum evolution can be considered to be the same at each run). The precision of the estimation is defined by the variance

$$\Delta\theta=\left\langle\left(\left|\frac{\partial\langle\theta_{\rm est}\rangle_\theta}{\partial\theta}\right|^{-1}\theta_{\rm est}-\theta\right)^2\right\rangle_\theta^{\frac{1}{2}}\ , \tag{261}$$

where $\langle\,\cdot\,\rangle_\theta$ denotes the average for the product probability measure $\{p_{i_1|\theta}\ldots p_{i_N|\theta}\}_{i_1,\ldots,i_N=1}^m$ of the independent outcomes. The factor $|\partial\langle\theta_{\rm est}\rangle_\theta/\partial\theta|^{-1}$ is put in front of $\theta_{\rm est}$ to remove some possible differences in physical units between θ and its estimator $\theta_{\rm est}$ (see Ref. 30). We restrict our attention to unbiased estimators satisfying $|\partial\langle\theta_{\rm est}\rangle_\theta/\partial\theta|^{-1}\langle\theta_{\rm est}\rangle_\theta=\theta$. For a given input state ρ, one looks for the smallest error $\Delta\theta$ that can be achieved. This involves two different optimization steps, associated to the optimization over (i) all possible estimators $\theta_{\rm est}$ and (ii) all possible measurements. The step (i) relies on a classical result in statistics known as the Cramér-Rao bound,

$$\left\langle(\Delta\theta_{\rm est})^2\right\rangle_\theta\geq\frac{1}{N\mathcal{F}(\{p_{i|\theta}\})}\left(\frac{\partial\langle\theta_{\rm est}\rangle_\theta}{\partial\theta}\right)^2, \tag{262}$$

where $\Delta\theta_{\rm est}=\theta_{\rm est}-\langle\theta_{\rm est}\rangle_\theta$ and

$$\mathcal{F}(\{p_{i|\theta}\})=\sum_{i=1}^m\frac{1}{p_{i|\theta}}\left(\frac{\partial p_{i|\theta}}{\partial\theta}\right)^2 \tag{263}$$

is the Fisher information. The inequality (262) is saturated asymptotically for $N\to\infty$ by the maximum-likelihood estimator. The second optimization step (ii) has been solved in Ref. 30, leading to the following important statement. Recall that the quantum Fisher information is defined as (see Sec. VII E)

$$\mathcal{F}_Q(\rho,H)=4(g_{\rm B})_\rho(-{\rm i}[H,\rho],-{\rm i}[H,\rho])=4d_{\rm B}(\rho,\rho+{\rm d}\rho)^2\ , \tag{264}$$

where $g_{\rm B}$ is the Bures metric and ${\rm d}\rho=(\partial\rho/\partial\theta){\rm d}\theta=-{\rm i}[H,\rho]{\rm d}\theta$.

Proposition 8.B.1. (Braunstein and Caves[30]) *The smallest error $\Delta\theta$ that can be achieved in the parameter estimation is*

$$(\Delta\theta)_{\text{best}} = \frac{1}{\sqrt{N}\sqrt{\mathcal{F}_Q(\rho, H)}}\,, \tag{265}$$

where N is the number of measurements and $\mathcal{F}_Q(\rho, H)$ is the quantum Fisher information. Thus $\Delta\theta \geq (\Delta\theta)_{\text{best}}$ and the equality $\Delta\theta = (\Delta\theta)_{\text{best}}$ can be reached asymptotically as $N \to \infty$.

It is worth noting that (265) can be interpreted as a generalized uncertainty principle.[30] In fact, if $\rho = |\Psi\rangle\langle\Psi|$ is a pure state, in view of the relation (210) between $\mathcal{F}_Q(\rho, H)$ and the square fluctuation $\langle(\Delta H)^2\rangle_\Psi$ of H, the bound $\Delta\theta \geq (\Delta\theta)_{\text{best}}$ can be written as

$$\Delta\theta\, \langle(\Delta H)^2\rangle_\Psi^{\frac{1}{2}} \geq \frac{1}{2\sqrt{N}}\,. \tag{266}$$

In this uncertainty relation H plays the role of the variable conjugated to the parameter θ.

Proof. We present here a direct proof of (265) based on the results of Sec. VII (see Ref. 30 for an independent proof). Before that, let us explain how the classical Cramér-Rao bound is derived. By differentiating with respect to θ the identity

$$0 = \langle\Delta\theta_{\text{est}}\rangle_\theta = \sum_{i_1,\ldots,i_N} p_{i_1|\theta}\ldots p_{i_N|\theta}\Delta\theta_{\text{est}}(i_1,\ldots,i_N) \tag{267}$$

one obtains

$$0 = \sum_{i_1,\ldots,i_N} p_{i_1|\theta}\ldots p_{i_N|\theta}\sum_{\nu=1}^{N}\frac{\partial \ln p_{i_\nu|\theta}}{\partial\theta}\Delta\theta_{\text{est}}(i_1,\ldots,i_N) - \frac{\partial\langle\theta_{\text{est}}\rangle_\theta}{\partial\theta}. \tag{268}$$

Then the Cramér-Rao bound (262) readily follows from the Cauchy-Schwarz inequality. Of course, the interesting point is that equality can be achieved in the limit $N \to \infty$, but we will not dwell into that. Going back to the quantum problem, we rearrange (262) as

$$\frac{(\mathrm{d}\theta)^2}{N} \leq (\Delta\theta)^2\sum_{i=1}^{m}\frac{(\mathrm{tr}[M_i\mathrm{d}\rho(\theta)])^2}{\mathrm{tr}[M_i\rho(\theta)]} \tag{269}$$

with $\mathrm{d}\rho(\theta) = (\partial\rho/\partial\theta)\mathrm{d}\theta$. Now, by using Proposition 7.C.1 and performing an expansion up to the second order in $\mathrm{d}\rho$, one finds

$$\mathcal{F}_Q(\rho(\theta), H)(\mathrm{d}\theta)^2 = \sup_{\{M_i\}}\left\{\sum_{i=1}^{m}\frac{(\mathrm{tr}[M_i\mathrm{d}\rho(\theta)])^2}{\mathrm{tr}[M_i\rho(\theta)]}\right\}. \tag{270}$$

Here, the supremum is over all POVMs $\{M_i\}$ and we have used $\sum_i \mathrm{tr}[M_i\mathrm{d}\rho(\theta)] = \mathrm{tr}[\mathrm{d}\rho(\theta)] = 0$. But $\mathcal{F}_Q(\rho(\theta), H) = \mathcal{F}_Q(\rho, H)$ as a consequence of (209), since $\rho(\theta)$ and ρ are related by a unitary evolution generated by H. Comparing (269) and (270), we conclude that $\inf_{\{M_i\}}\Delta\theta \geq (\Delta\theta)_{\text{best}}$, with equality as $N \to \infty$ for the maximum likelihood estimator, as stated in the proposition. □

Before proceeding to derive upper bounds on $\mathcal{F}_Q(\rho, H)$, let us observe that the monotonicity of the Bures metric g_{B} implies:[61]

Corollary 8.B.2. *The quantum Fisher information $\mathcal{F}_Q(\rho, H)$ is convex in ρ.*

Proof. Given two states ρ_0 and ρ_1 on $\mathcal{H}$ and η_0, $\eta_1 \geq 0$, $\eta_0 + \eta_1 = 1$, we introduce the state $\widehat{\rho} = \eta_0\rho_0 \otimes |0\rangle\langle 0| + \eta_1\rho_1 \otimes |1\rangle\langle 1|$ on $\mathcal{H}\otimes\mathbb{C}^2$ as in the proof of Theorem 7.F.2. From the expression of $\mathcal{F}_Q$ in the right-hand side of (209) one deduces that

$$\mathcal{F}_Q(\widehat{\rho}, H\otimes 1) = \eta_0\mathcal{F}_Q(\rho_0, H) + \eta_1\mathcal{F}_Q(\rho_1, H). \tag{271}$$

Let $\mathcal{T} : \widehat{\sigma} \mapsto \mathrm{tr}_{\mathbb{C}^2}(\widehat{\sigma})$ denote the partial trace on $\mathbb{C}^2$. Then $\mathcal{T}(\widehat{\rho}) = \rho = \eta_0\rho_0 + \eta_1\rho_1$ and $\mathcal{T}([H\otimes 1, \widehat{\rho}]) = [H, \rho]$. As $\mathcal{T}$ is a quantum operation, it results from the contractivity of the Bures metric

that

$$(g_{\mathrm{B}})_{\widehat{\rho}}\big(-\mathrm{i}[H\otimes 1,\widehat{\rho}],-\mathrm{i}[H\otimes 1,\widehat{\rho}]\big)\geq (g_{\mathrm{B}})_{\rho}\big(-\mathrm{i}[H,\rho],-\mathrm{i}[H,\rho]\big). \tag{272}$$

Collecting together (271) and (272) yields $\mathcal{F}_Q(\rho,H)\leq \eta_0\mathcal{F}_Q(\rho_0,H)+\eta_1\mathcal{F}_Q(\rho_1,H)$. □

3. *Interferometer precision and inter-particle entanglement*

We now show by relying on Proposition 8.B.1 that if the input state has N_{p} particles in a maximally entangled state, the precision $(\Delta\theta)_{\mathrm{best}}$ is smaller by a factor $1/\sqrt{N_{\mathrm{p}}}$ with respect to the precision obtained with separable input states. The Hilbert space of the particles is $\mathcal{H}^{(N_{\mathrm{p}})}=\mathcal{H}_1\otimes\cdots\otimes\mathcal{H}_{N_{\mathrm{p}}}$, $\mathcal{H}_\nu$ being the Hilbert space of the νth particle. Assuming that the particles do not interact between themselves, the Hamiltonian reads

$$H=\sum_{\nu=1}^{N_{\mathrm{p}}}1\otimes\cdots\otimes H_\nu\otimes\cdots\otimes 1, \tag{273}$$

where H_ν acts on $\mathcal{H}_\nu$. To simplify the discussion we suppose that the single particle Hamiltonians H_ν have the same highest eigenvalue $\lambda_{\max}$ and the same lowest eigenvalue $\lambda_{\min}$. This is the case, for instance, if H is the angular momentum $J_{\mathbf{n}}$ in the interferometer of Sec. VIII B 1 (then $H_\nu=(n_x\sigma_{x\nu}+n_y\sigma_{y\nu}+n_z\sigma_{z\nu})/2$ with $|\mathbf{n}|=1$ and $\sigma_{x\nu}$, $\sigma_{y\nu}$, and $\sigma_{z\nu}$ the three Pauli matrices acting on $\mathcal{H}_\nu\simeq\mathbb{C}^2$, so that $\lambda_{\max}=-\lambda_{\min}=1/2$). Let us recall that the quantum Fisher information $\mathcal{F}_Q(|\Psi\rangle,H)$ of a pure state $|\Psi\rangle$ is given by the square fluctuation $\langle(\Delta H)^2\rangle_\Psi=\langle\Psi|H^2|\Psi\rangle-\langle\Psi|H|\Psi\rangle^2$ up to a factor of four (see Sec. VII E). We first observe that the maximum of $\langle(\Delta H_\nu)^2\rangle_{\psi_\nu}$ over all pure states $|\psi_\nu\rangle\in\mathcal{H}_\nu$ is equal to $(\Delta h)^2=(\lambda_{\max}-\lambda_{\min})^2/4$, the maximum being attained when $|\psi_\nu\rangle=(|\phi_{\nu,\max}\rangle+|\phi_{\nu,\min}\rangle)/\sqrt{2}$, where $|\phi_{\nu,\max}\rangle$ and $|\phi_{\nu,\min}\rangle$ are the eigenvectors of H_ν with eigenvalues $\lambda_{\max}$ and $\lambda_{\min}$, respectively. Let the N_{p} particles be in a separable state ρ_{sep} and let $\{|\Psi_i\rangle,\eta_i\}$ be a decomposition of ρ_{sep} into pure product states $|\Psi_i\rangle=|\psi_{i1}\rangle\otimes\cdots\otimes|\psi_{iN_{\mathrm{p}}}\rangle\in\mathcal{H}^{(N_{\mathrm{p}})}$. A simple calculation gives[63]

$$\mathcal{F}_Q(|\Psi_i\rangle,H)=4\big\langle(\Delta H)^2\big\rangle_{\Psi_i}=4\sum_{\nu=1}^{N_{\mathrm{p}}}\langle(\Delta H_\nu)^2\rangle_{\psi_{i\nu}}\ \leq\ 4(\Delta h)^2N_{\mathrm{p}}\,. \tag{274}$$

By applying Corollary 8.B.2 we get

$$\rho_{\mathrm{sep}}\ \text{ separable }\ \Rightarrow\ \ \mathcal{F}_Q(\rho_{\mathrm{sep}},H)\leq 4(\Delta h)^2N_{\mathrm{p}}\,. \tag{275}$$

According to Proposition 8.B.1 the phase precision of the interferometer satisfies for separable inputs

$$\Delta\theta\geq(\Delta\theta)_{\mathrm{SN}}=\frac{1}{2\Delta h\sqrt{NN_{\mathrm{p}}}}\,. \tag{276}$$

This means that separable input states cannot do better than N_{p} independent particles sent one-by-one through the interferometer, henceforth producing an error of the order of $1/\sqrt{N_{\mathrm{p}}}$. Note that (275) provides a sufficient condition $\mathcal{F}_Q(\rho,H)>4(\Delta h)^2N_{\mathrm{p}}$ for entanglement of ρ.[130] There are, however, entangled states which do not satisfy this criterion.[130] Such entangled states are not useful for interferometry, in the sense that they produce phase errors larger than the shot noise value $(\Delta\theta)_{\mathrm{SN}}$.

We now argue that much higher Fisher informations, of the order of N_{p}^2, can be achieved for entangled states. By the same observation as above, $\langle(\Delta H)^2\rangle_\Psi$ has a maximum given by the square of the half difference of the maximal and minimal eigenvalues of H. For the Hamiltonian (273), one immediately finds

$$\mathcal{F}_Q(|\Psi\rangle,H)\leq 4(\Delta h)^2N_{\mathrm{p}}^2\,. \tag{277}$$

This upper bound is often called the *Heisenberg bound* in the literature. It is saturated for the entangled states[63]

$$|\Psi_{\mathrm{ent}}^{\pm}\rangle=\frac{1}{\sqrt{2}}\Big(|\phi_{1,\max}\rangle|\phi_{2,\max}\rangle\ldots|\phi_{N_{\mathrm{p}},\max}\rangle\pm|\phi_{1,\min}\rangle|\phi_{2,\min}\rangle\ldots|\phi_{N_{\mathrm{p}},\min}\rangle\Big). \tag{278}$$

For large N_p such states deserve the name of *macroscopic superpositions*, as they are formed by a superposition of two macroscopically distinct states in which each particle is in the highest energy eigenstate of the single particle Hamiltonian (for the first component of the superposition) or in the lowest energy eigenstate (for the second component). If one uses these superpositions as input states of the interferometer, an error of $\Delta\theta = 1/(2\Delta h\sqrt{N}N_\mathrm{p}) = (\Delta\theta)_\mathrm{SN}/\sqrt{N_\mathrm{p}}$ can be achieved asymptotically for $N \to \infty$ on the unknown phase. According to (265) and (277), this is the best possible precision.

IX. MEASURES OF ENTANGLEMENT IN BIPARTITE SYSTEMS

Even if it would be better for many computational and communication tasks to work with maximally entangled pure states, in practice the coupling of the system with its environment transforms such states into non-maximally entangled mixed states because of the induced decoherence processes.[31,64,70] It is thus important to quantify the amount of entanglement in an arbitrary quantum state. Unfortunately, this amount of entanglement is not a directly measurable quantity. It is quantified by an entanglement measure, which vanishes if and only if the state is separable and cannot increase under local operations on each subsystems and classical communication (entanglement monotonicity). All measures satisfying these two requirements are not equivalent, i.e., a state ρ can be more entangled than a state σ for one measure and less entangled for the other. In this section, we investigate the properties of entanglement measures, give their general form for pure states, and study more especially two of the most popular ones, the entanglement of formation and the concurrence. We restrict our attention to bipartite entanglement (see Refs. 67 and 82 for generalizations to entanglement in systems with more than two parties).

A. Entanglement as correlations between local measurements

Let $|\Psi\rangle$ be a pure state of a bipartite system AB. In view of the discussion in Sec. II D, it seems natural physically to characterize the entanglement in $|\Psi\rangle$ by maximizing the correlator $G_{AB}(|\Psi\rangle)$ in (17) over all local observables $A \in \mathcal{B}(\mathcal{H}_\mathsf{A})_\mathrm{s.a.}$ and $B \in \mathcal{B}(\mathcal{H}_\mathsf{B})_\mathrm{s.a.}$ and to define

$$G(|\Psi\rangle) = \max_{A=A^*,\|\Delta A\|_{\infty,\Psi}\le 1} \max_{B=B^*,\|\Delta B\|_{\infty,\Psi}\le 1} \left\{ \left| G_{AB}(|\Psi\rangle) \right| \right\} . \tag{279}$$

One must face with some arbitrariness on the choice of the norm used to bound $\Delta A = A - \langle A \otimes 1\rangle_\Psi$ and $\Delta B = B - \langle 1 \otimes B\rangle_\Psi$. In order to obtain an entanglement measure with the required properties, we take the Ψ-dependent norm $\|\Delta A\|_{\infty,\Psi} = \max_{i,j} |\langle\alpha_i|\Delta A|\alpha_j\rangle|$, where $\{|\alpha_i\rangle\}$ is an orthonormal eigenbasis of the reduced state $[\rho_\Psi]_\mathsf{A}$, and similarly for $\|\Delta B\|_{\infty,\Psi}$ with the eigenbasis $\{|\beta_k\rangle\}$ of $[\rho_\Psi]_\mathsf{B}$. These norms correspond to the infinity norms of the vectors in $\mathcal{H}_\mathsf{AA}$ and $\mathcal{H}_\mathsf{BB}$ associated to ΔA and ΔB via the isometry (5). By using the Schmidt decomposition (9) and setting $A_{ij} = \langle\alpha_i|A|\alpha_j\rangle$ and $B_{ij} = \langle\beta_i|B|\beta_j\rangle$, one finds

$$G_{AB}(|\Psi\rangle) = \langle \Delta A \otimes \Delta B\rangle_\Psi = \sum_{i=1}^n \mu_i (\Delta A)_{ii}(\Delta B)_{ii} + \sum_{i\ne j}^n \sqrt{\mu_i\mu_j} A_{ij} B_{ij} . \tag{280}$$

The Cauchy-Schwarz inequality immediately yields

$$G(|\Psi\rangle) = \max_{\|\Delta\mathbf{a}\|_\infty \le 1} \left\{ \overline{(\Delta\mathbf{a})^2} \right\} + C(|\Psi\rangle), \tag{281}$$

where the overline stands for the average with respect to the Schmidt coefficients μ_i (e.g., $\bar{\mathbf{a}} = \sum_i \mu_i a_i$), $\Delta\mathbf{a} = \mathbf{a} - \bar{\mathbf{a}}$ with $\mathbf{a} = (A_{11}, \ldots, A_{mn})$, $\|\Delta\mathbf{a}\|_\infty = \max_i |(\Delta\mathbf{a})_i|$, and

$$C(|\Psi\rangle) = \sum_{i\ne j}^n \sqrt{\mu_i\mu_j} = \left(\mathrm{tr}\sqrt{[\rho_\Psi]_\mathsf{A}} \right)^2 - 1 . \tag{282}$$

Thus $C(|\Psi\rangle) = 0$ (similarly, $G(|\Psi\rangle) = 0$) is equivalent to $\mu_i = 0$ save for one index i, that is, to $|\Psi\rangle$ being separable. Furthermore, $C(|\Psi\rangle) \le n - 1$ with equality if and only if $\mu_i = 1/n$ for all i, that is, if and only if $|\Psi\rangle$ is maximally entangled (Sec. II D). This last property is not true if one uses the

operator norm instead of $\|\cdot\|_{\infty,\Psi}$ in (279), except in the two-qubit case $n = 2$. Finally, we note that G and C are invariant under local unitaries, i.e., $G(U_A \otimes U_B|\Psi\rangle) = G(|\Psi\rangle)$ for any unitaries U_A and U_B on $\mathcal{H}_\mathsf{A}$ and $\mathcal{H}_\mathsf{B}$. For two qubits one obtains

$$G(|\Psi\rangle) = \mu_{\max}^{-1} - 1 + C(|\Psi\rangle), \quad C(|\Psi\rangle) = 2\sqrt{\mu_0\mu_1} \tag{283}$$

with $\mu_{\max} = \max\{\mu_0, \mu_1\}$. It is easy to show that $C(|\Psi\rangle) = |\langle\Psi|\sigma_y \otimes \sigma_y J|\Psi\rangle|$, where $\sigma_y = \mathrm{i}(|0\rangle\langle 1| - |1\rangle\langle 0|)$ is the y-Pauli matrix and J the complex conjugation in the canonical basis. This quantity has been first introduced by Wootters[169] and is known as the *concurrence*.

One may wonder how the correlator G_{AB} could be generalized for mixed states. The first guess would be to replace the expectation value $\langle\cdot\rangle_\Psi$ by $\langle\cdot\rangle_\rho = \mathrm{tr}(\rho\,\cdot)$, but one easily sees that then $G(\rho)$ can be non-zero even for separable mixed states, because this correlator contains both the quantum and classical (i.e., statistical) correlations in the density matrix ρ. Noting that

$$G_{AB}(|\Psi\rangle) = \frac{1}{2}\big\langle\big(\Delta(A\otimes 1 + 1\otimes B)\big)^2\big\rangle_\Psi - \frac{1}{2}\big\langle\big(\Delta(A\otimes 1)\big)^2\big\rangle_\Psi - \frac{1}{2}\big\langle\big(\Delta(1\otimes B)\big)^2\big\rangle_\Psi, \tag{284}$$

it is tempting to define a correlator for ρ in terms of the quantum Fisher information (209), i.e., of the Bures metric g_B,

$$\begin{aligned} G_{AB}(\rho) &= \frac{1}{8}\Big(\mathcal{F}_Q(\rho, A\otimes 1 + 1\otimes B) - \mathcal{F}_Q(\rho, A\otimes 1) - \mathcal{F}_Q(\rho, 1\otimes B)\Big) \\ &= \mathrm{Re}\,\big\{(g_\mathrm{B})_\rho\big(-\mathrm{i}[A\otimes 1, \rho], -\mathrm{i}[1\otimes B, \rho]\big)\big\}. \end{aligned}$$

By inspection on (210), $G_{AB}(\rho)$ reduces for pure states to the previous correlator. However, the maximum of $|G_{AB}(\rho)|$ over all A and B does not fulfill the axioms of an entanglement measure. We will see in Sec. IX D another way to define the concurrence C for mixed states, by using on a convex roof construction.

B. LOCC operations

The main physical postulate on entanglement measures is that they must be monotonous with respect to certain state transformations. Such transformations that cannot increase entanglement are called *Local Operations and Classical Communication* (LOCC) and can be described as follows.[23,82] Let us consider an entangled state ρ shared by two observers Alice and Bob. Alice and Bob can perform any quantum operations $\mathcal{M}_\mathsf{A} : \mathcal{B}(\mathcal{H}_\mathsf{A}) \to \mathcal{B}(\mathcal{H}'_\mathsf{A})$ and $\mathcal{M}_\mathsf{B} : \mathcal{B}(\mathcal{H}_\mathsf{B}) \to \mathcal{B}(\mathcal{H}'_\mathsf{B})$ on their respective subsystems A and B. Here, the final spaces $\mathcal{H}'_\mathsf{A}$ and $\mathcal{H}'_\mathsf{B}$ may include local ancillae, or may be some subspaces of $\mathcal{H}_\mathsf{A}$ and $\mathcal{H}_\mathsf{B}$, respectively. The corresponding transformations on the system AB are called *local quantum operations*. They are of the form $\mathcal{M}_\mathrm{loc} = \mathcal{M}_\mathsf{A} \otimes \mathcal{M}_\mathsf{B}$ and are given by families $\{A_i \otimes B_j\}$ of Kraus operators, where A_i and B_j are local observables on A and B. Local operations are performed physically by coupling each subsystem to a local ancilla and by making joint unitary evolutions and von Neumann measurements on the subsystem and its ancilla (see Sec. III B). Such processes can clearly not increase the amount of entanglement between A and B. In addition to performing local generalized measurements, Alice and Bob can communicate their measurement outcomes to each other via a classical communication channel (two-way communication). No transfer of quantum systems between them is allowed. Thanks to classical communication, the observers can increase the classical correlations between A and B, but not the AB-entanglement. A *LOCC operation* is a quantum operation on $\mathcal{B}(\mathcal{H}_\mathsf{AB})$ obtained through a succession of the aforementioned actions of Alice and Bob, taken in arbitrary order. For example, if Alice performs a measurement on A and Bob a measurement on B depending on Alice's outcome i (one way communication), the post-measurement state in the absence of readout is

$$\mathcal{M}_{1-\mathrm{way}}(\rho) = \sum_i 1\otimes\mathcal{M}_\mathsf{B}^{(i)}(A_i\otimes 1\rho A_i^*\otimes 1). \tag{285}$$

This defines a LOCC operation with Kraus operators $A_i \otimes B_j^{(i)}$, where $\sum_i A_i^* A_i = \sum_j (B_j^{(i)})^* B_j^{(i)} = 1$.

Any LOCC operation can be obtained by composing local operations $\mathcal{M}_{\rm loc}$ with the maps

$$\mathcal{M}^{\mathsf{A}}_{\rm LOCC}(\rho)=\sum_i \big(A_i\otimes 1\,\rho\,A_i^*\otimes 1\big)\otimes|\kappa_i\rangle\langle\kappa_i|,\quad \mathcal{M}^{\mathsf{B}}_{\rm LOCC}(\rho)=\sum_j\big(1\otimes B_j\,\rho\,1\otimes B_j^*\big)\otimes|\epsilon_j\rangle\langle\epsilon_j|, \tag{286}$$

where $\sum_i A_i^*A_i=\sum_j B_j^*B_j=1$ and $\{|\kappa_i\rangle\}$ (respectively, $\{|\epsilon_j\rangle\}$) is an orthonormal basis for Bob's ancilla (respectively, Alice's ancilla).[82] A strictly larger but much simpler class of transformations, known as the *separable quantum operations*,[160] is the set of all operations with Kraus operators $A_i\otimes B_i$, i.e.,

$$\mathcal{M}_{\rm sep}(\rho)=\sum_i A_i\otimes B_i\rho A_i^*\otimes B_i^* \tag{287}$$

with $A_i\in\mathcal{B}(\mathcal{H}_{\mathsf{A}},\mathcal{H}'_{\mathsf{A}})$, $B_i\in\mathcal{B}(\mathcal{H}_{\mathsf{B}},\mathcal{H}'_{\mathsf{B}})$, and $\sum_i A_i^*A_i\otimes B_i^*B_i=1$. The local operations and maps (286) being separable, any LOCC operation is separable. A result from Ref. 24 shows, however, that certain separable operations are not LOCCs.

It is clear that the set $\mathcal{S}_{\mathsf{AB}}$ of separable states is invariant under separable operations. It is also true that every separable state can be converted into any other separable state by a separable operation. Actually, any separable state can be obtained from the classical state $\rho_{\rm clas}=\sum_{jk}p_{jk}|j\rangle\langle j|\otimes|k\rangle\langle k|$ by such an operation (take $A_{ijk}=\sqrt{\eta_i}|\psi_i\rangle\langle j|$ and $B_{ijk}=|\phi_i\rangle\langle k|$ with η_i, $|\psi_i\rangle$, and $|\phi_i\rangle$ as in (19)). Furthermore, an arbitrary state ρ can be transformed into a classical state $\rho_{\rm clas}$ by a measurement in the product basis $\{|j\rangle|k\rangle\}$, which is a local operation.

When one restricts LOCC transformations to pure states, a great simplification comes from the following observation. If the space dimensions of A and B are such that $n_{\mathsf{A}}\geq n_{\mathsf{B}}$, any measurement by Bob can be simulated by a measurement by Alice followed by a unitary transformation by Bob conditioned to Alice's outcome (such a conditioning is allowed as Alice and Bob can communicate classically). In fact, let $\{|\alpha_i\rangle\}_{i=1}^{n_{\mathsf{A}}}$ and $\{|\beta_i\rangle\}_{i=1}^{n_{\mathsf{B}}}$ be orthonormal eigenbasis of the reduced states $[\rho_\Psi]_{\mathsf{A}}$ and $[\rho_\Psi]_{\mathsf{B}}$, and let B_i be the Kraus operators describing Bob's measurement. Consider the measurement done by Alice with Kraus operators $A_i=\sum_{j,l}(B_i)_{lj}|\alpha_l\rangle\langle\alpha_j|$, where $(B_i)_{lj}=\langle\beta_l|B_i|\beta_j\rangle$. The unnormalized post-measurement states

$$|\widetilde{\Phi}_i\rangle=1\otimes B_i|\Psi\rangle=\sum_{j,l}\sqrt{\mu_j}(B_i)_{lj}|\alpha_j\rangle|\beta_l\rangle,\quad |\widetilde{\Phi}'_i\rangle=A_i\otimes 1|\Psi\rangle=\sum_{j,l}\sqrt{\mu_j}(B_i)_{lj}|\alpha_l\rangle|\beta_j\rangle \tag{288}$$

have the same Schmidt coefficients because $\mathrm{tr}_{\mathsf{B}}(|\widetilde{\Phi}_i\rangle\langle\widetilde{\Phi}_i|)$ and $\mathrm{tr}_{\mathsf{A}}(|\widetilde{\Phi}'_i\rangle\langle\widetilde{\Phi}'_i|)$ are related by an isometry $\mathcal{H}_{\mathsf{A}}\to\mathcal{H}_{\mathsf{B}}$. Thus $|\widetilde{\Phi}'_i\rangle=U_i\otimes V_i|\widetilde{\Phi}_i\rangle$ for some local unitaries U_i on $\mathcal{H}_{\mathsf{A}}$ and V_i on $\mathcal{H}_{\mathsf{B}}$. Consequently, Bob performing the measurement $\{B_i\}$ is equivalent to Alice performing the measurement $\{U_i^*A_i\}$ and Bob performing the unitary transformation V_i^* when Alice gets the outcome i. Applying this result to all Bob's measurements, we conclude that a LOCC acting on a pure state $|\Psi\rangle$ may always be simulated by a one-way communication protocol involving only three steps: (1) Alice first performs a generalized measurement on subsystem A; (2) she sends her measurement result to Bob; (3) Bob performs a unitary evolution on B conditional to Alice's result.

Based on this observation, we say that a pure state $|\Psi\rangle\in\mathcal{H}_{\mathsf{AB}}$ can be transformed by a LOCC into the pure state $|\Phi\rangle\in\mathcal{H}_{\mathsf{AB}}$ if there are families of Kraus operators $\{A_i\}$ on $\mathcal{H}_{\mathsf{A}}$ and unitaries $\{V_i\}$ on $\mathcal{H}_{\mathsf{B}}$ such that all unnormalized conditional states $A_i\otimes V_i|\Psi\rangle$ are proportional to $|\Phi\rangle$, irrespective of the measurement outcome i. Note that this is equivalent to $\mathcal{M}_{\rm LOCC}(|\Psi\rangle\langle\Psi|)$ being equal to $|\Phi\rangle\langle\Phi|$, with $\mathcal{M}_{\rm LOCC}$ the LOCC operation with Kraus family $\{A_i\otimes V_i\}$. One defines in this way an order relation on the set of pure states. Nielsen[116] discovered a nice relation between this order and the theory of majorization for n-dimensional vectors.[27] Let $\mathbf{x}=(x_1,\ldots,x_n)$ and $\mathbf{y}=(y_1,\ldots,y_n)$ be two vectors in $\mathbb{R}^n$. We denote by $\mathbf{x}^\downarrow$ the vector formed by the components of $\mathbf{x}$ in decreasing order, and similarly for $\mathbf{y}^\downarrow$. One says that $\mathbf{x}$ is majorized by $\mathbf{y}$ and write $\mathbf{x}\prec\mathbf{y}$ if $\sum_{i=1}^k x_i^\downarrow\leq\sum_{i=1}^k y_i^\downarrow$ for any $k=1,\ldots,n$, with equality instead of inequality for $k=n$.

Proposition 9.B.1. (Nielsen[116]) *A pure state* $|\Psi\rangle$ *of the bipartite system* AB *can be transformed into another pure state* $|\Phi\rangle$ *of* AB *by a LOCC if and only if* $\boldsymbol{\mu}_\Psi \prec \boldsymbol{\mu}_\Phi$, *where* $\boldsymbol{\mu}_\Psi$ *and* $\boldsymbol{\mu}_\Phi$ *are the vectors formed by the Schmidt coefficients of* $|\Psi\rangle$ *and* $|\Phi\rangle$, *respectively*.

A detailed proof of this result can be found in Ref. 117 (Sec. 12.5), so we omit it here. This proof relies on the following theorem: if $\boldsymbol{\lambda}_H$ and $\boldsymbol{\lambda}_K$ are vectors formed by the eigenvalues of two Hermitian matrices H and K, respectively, then $\boldsymbol{\lambda}_H \prec \boldsymbol{\lambda}_K$ if and only if $H = \sum_i \eta_i U_i K U_i^*$ with $\{\eta_i\}$ a set of probabilities and U_i some unitary matrices.

Remark 9.B.2. Even if $|\Psi\rangle$ *cannot be transformed into* $|\Phi\rangle$ *by a LOCC, it may still happen that* $|\Psi\rangle \otimes |\kappa\rangle$ *can be transformed into* $|\Phi\rangle \otimes |\kappa\rangle$ *by a LOCC (here the state of the ancilla does not change during the transformation, i.e., it acts as catalysts in chemical reactions).*[90]

C. Axioms on entanglement measures

We are now in position to formulate the physical postulates on entanglement measures.[23,160,161]

Definition 9.C.1. *An entanglement measure of a bipartite system* AB *is a function* $E : \mathcal{E}(\mathcal{H}_{\mathsf{AB}}) \to \mathbb{R}$ *such that*

(i) $E(\rho) = 0$ *if and only if* ρ *is separable;*
(ii) E *is convex*;
(iii) E *cannot increase under LOCCs, i.e., if* $\mathcal{M}_{\mathrm{LOCC}}$ *is a LOCC operation then* $E(\mathcal{M}_{\mathrm{LOCC}}(\rho)) \leq E(\rho)$.

As any two separable states can be transformed one into each other by means of a LOCC operation, the monotonicity (iii) implies that E is constant on the set of separable states $\mathcal{S}_{\mathsf{AB}}$. Taking this constant equal to zero yields $\rho \in \mathcal{S}_{\mathsf{AB}} \Rightarrow E(\rho) = 0$, so that only the reverse implication is needed in (i). Furthermore, any state ρ can be converted into a separable state by a LOCC, thus $E(\rho)$ is minimum for separable states and $E(\rho) \geq 0$. The convexity condition (ii) is motivated by the following observation.[161] Assume that Alice and Bob share m pairs of particles in the states ρ_1, $\ldots$, ρ_m. By classical communication, they can agree to keep the ith pair with probability η_i, thus preparing the ensemble $\{\rho_i, \eta_i\}_{i=1}^m$. By erasing the information about which state ρ_i was kept, the state becomes $\rho = \sum \eta_i \rho_i$ (see Sec. II C). The inequality $E(\rho) \leq \sum_i \eta_i E(\rho_i)$ means that this local loss of information does not increase the average entanglement.

It results from the monotonicity (iii) that entanglement measures are invariant under conjugations by local unitaries, i.e., $E(U_{\mathsf{A}} \otimes U_{\mathsf{B}}\, \rho\, U_{\mathsf{A}}^* \otimes U_{\mathsf{B}}^*) = E(\rho)$. For pure states $|\Psi\rangle$, this implies that $E(|\Psi\rangle)$ only depends on the Schmidt coefficients μ_i of $|\Psi\rangle$. Consequently, $E(|\Psi\rangle) = f([\rho_\Psi]_{\mathsf{A}})$ is a unitary-invariant function of the reduced state $[\rho_\Psi]_{\mathsf{A}} = \mathrm{tr}_{\mathsf{B}}(|\Psi\rangle\langle\Psi|)$ (or, equivalently, of $[\rho_\Psi]_{\mathsf{B}} = \mathrm{tr}_{\mathsf{A}}(|\Psi\rangle\langle\Psi|)$). Given that a pure state is separable if and only if it has a single non-vanishing Schmidt coefficient, one deduces from axiom (i) that $f(\rho_{\mathsf{A}})$ vanishes if and only if ρ_{A} is of rank one. The result below due to Vidal[161] characterizes all entanglement measures on pure states satisfying a slightly stronger condition than (iii). This shows in particular that there are many measures of entanglement fulfilling the three physical requirements (i-iii) of Definition 9.C.1, given by concave functions f.

Proposition 9.C.2. (Vidal[161]) *Let* $f : \mathcal{E}(\mathcal{H}_{\mathsf{A}}) \to \mathbb{R}$ *be concave, unitary invariant, and such that* $f(\rho_{\mathsf{A}}) = 0$ *if and only if* ρ_{A} *is a pure state. Then*

$$E_f(|\Psi\rangle) = f([\rho_\Psi]_{\mathsf{A}}) \tag{289}$$

defines an entanglement measure on the set of pure states of AB, *which satisfies the monotonicity condition*

(iii′) $\sum_i p_i E(|\Phi_i\rangle) \leq E(|\Psi\rangle)$, *where* $p_i = \|A_i \otimes B_i|\Psi\rangle\|^2$ and $|\Phi_i\rangle = p_i^{-1/2} A_i \otimes B_i|\Psi\rangle$ *and the probabilities and conditional states of a separable measurement with Kraus operators* $A_i \otimes B_i$.

Conversely, any entanglement measure on pure states fulfilling (iii′) is given by (289) for some function f satisfying the above assumptions.

It should be noted that asking $E(|\Phi_i\rangle) \leq E(|\Psi\rangle)$ for all outcomes i would put a too strong condition on E. Indeed, local measurements can in principle create entanglement on some conditional states, but not on average (see below).

Proof. Let f be like in the proposition. We have already argued above that E_f fulfills axiom (i), and (ii) is empty because of the restriction to pure states. Recall that for such states any measurement on B can be simulated by a measurement on A followed by a unitary operation on B conditioned to the measurement result. Hence it suffices to show the monotonicity (iii′) for $B_i = V_i$ unitary. Let us set $\rho_{\mathsf{B}|i} = \mathrm{tr}_{\mathsf{A}}(|\Phi_i\rangle\langle\Phi_i|)$. Then $\{V_i^* \rho_{\mathsf{B}|i} V_i, p_i\}$ is a pure state decomposition of $[\rho_\Psi]_{\mathsf{B}}$, i.e., $\sum_i p_i V_i^* \rho_{\mathsf{B}|i} V_i = [\rho_\Psi]_{\mathsf{B}}$. This can be interpreted by saying that a local measurement on A does not modify the state of B when B has no information on the measurement outcomes (if this would not be true, information could be sent faster than light in contradiction with Einstein's principle of relativity[127]). The concavity and unitary invariance of f imply

$$\sum_i p_i E_f(|\Phi_i\rangle) = \sum_i p_i f(V_i^* \rho_{\mathsf{B}|i} V_i) \leq f([\rho_\Psi]_{\mathsf{B}}) = E_f(|\Psi\rangle) \,. \tag{290}$$

This shows (iii′). Thus E_f is an entanglement measure.

Reciprocally, let E be an entanglement measure on pure states satisfying (iii′). From the discussion before the proposition we know that $E(|\Psi\rangle) = f([\rho_\Psi]_{\mathsf{A}}) = f([\rho_\Psi]_{\mathsf{B}})$ for some unitary-invariant function f vanishing on pure states only. It remains to show that f is concave. We may assume that the space dimensions of A and B are such that $n_{\mathsf{A}} \leq n_{\mathsf{B}}$ (otherwise one can exchange the role of A and B in the arguments below). Let ρ_{A} be an arbitrary state of A and $\sigma_{\mathsf{A}}^{(1)}$, $\sigma_{\mathsf{A}}^{(2)}$ be such that $\rho_{\mathsf{A}} = p_1 \sigma_{\mathsf{A}}^{(1)} + p_2 \sigma_{\mathsf{A}}^{(2)}$ with $p_1 + p_2 = 1$. As $n_{\mathsf{A}} \leq n_{\mathsf{B}}$, one may find a purification $|\Psi\rangle$ of ρ_{A} on $\mathcal{H}_{\mathsf{AB}}$ (Sec. II C). If one can exhibit a measurement on B with outcome probabilities p_i and conditional states $|\Phi_i\rangle$ having marginals $\mathrm{tr}_{\mathsf{B}}(|\Phi_i\rangle\langle\Phi_i|) = \sigma_{\mathsf{A}}^{(i)}$ for $i = 1, 2$, then the concavity of f can be deduced from (iii′) thanks to the bound

$$f(\rho_{\mathsf{A}}) = E(|\Psi\rangle) \geq p_1 E(|\Phi_1\rangle) + p_2 E(|\Phi_2\rangle) = p_1 f(\sigma_{\mathsf{A}}^{(1)}) + p_2 f(\sigma_{\mathsf{A}}^{(2)}) \,. \tag{291}$$

The measurement we are looking for is just the square root measurement associated to $\{\sigma_{\mathsf{A}}^{(i)}, p_i\}$ (Sec. IV C). Indeed, let $\{|\alpha_j\rangle\}_{j=1}^{n_{\mathsf{A}}}$ and $\{|\beta_k\rangle\}_{k=1}^{n_{\mathsf{B}}}$ be eigenbases of $[\rho_\Psi]_{\mathsf{A}}$ and $[\rho_\Psi]_{\mathsf{B}}$ and M_i^{lsm}, $i = 1, 2$, be the operators on $\mathcal{H}_{\mathsf{B}}$ with matrix elements given by (compare with (57))

$$\langle\beta_j|M_i^{\mathrm{lsm}}|\beta_l\rangle = \begin{cases} p_i \langle\alpha_l|\rho_{\mathsf{A}}^{-\frac{1}{2}} \sigma_{\mathsf{A}}^{(i)} \rho_{\mathsf{A}}^{-\frac{1}{2}}|\alpha_j\rangle & \text{if } j, l = 1, \ldots, n_{\mathsf{A}} \\ 0 & \text{otherwise.} \end{cases} \tag{292}$$

If $n_{\mathsf{B}} > n_{\mathsf{A}}$ we add a third measurement operator, equal to the projector onto $\mathrm{span}\{|\beta_k\rangle; n_{\mathsf{A}} < k \leq n_{\mathsf{B}}\}$. Then $M_1^{\mathrm{lsm}} + M_2^{\mathrm{lsm}} + M_3^{\mathrm{lsm}} = 1$. With the help of the Schmidt decomposition (9) one finds that $\langle\Psi|1 \otimes M_i^{\mathrm{lsm}}|\Psi\rangle$ equals p_i for $i = 1, 2$ and zero for $i = 3$, and the conditional state $|\Phi_i\rangle = p_i^{-1/2} 1 \otimes \sqrt{M_i^{\mathrm{lsm}}}|\Psi\rangle$ has marginal $\mathrm{tr}_{\mathsf{B}}(|\Phi_i\rangle\langle\Phi_i|) = \sigma_{\mathsf{A}}^{(i)}$ for $i = 1, 2$. This concludes the proof. □

Proposition 9.C.2 can be partially justified with the help of Proposition 9.B.1. More precisely, the latter implies that $E_f(|\Psi\rangle) \geq E_f(|\Phi\rangle)$ if $|\Phi\rangle\langle\Phi| = \mathcal{M}_{\mathrm{LOCC}}(|\Psi\rangle\langle\Psi|)$, that is, if there exists a LOCC measurement on $|\Psi\rangle$ with all conditional states $|\Phi_i\rangle$ equal to $|\Phi\rangle$. This comes from the fact that, by unitary invariance, $f([\rho_\Psi]_{\mathsf{A}})$ is a symmetric function of the eigenvalues $(\mu_\Psi)_1, \ldots, (\mu_\Psi)_n$ of $[\rho_\Psi]_{\mathsf{A}}$. But concave symmetric functions $f : \mathbb{R}^n \to \mathbb{R}$ are Schur-concave, i.e., $\mathbf{x} \prec \mathbf{y} \Rightarrow f(\mathbf{x}) \geq f(\mathbf{y})$ (see Ref. 27, Theorem II.3.3).

Many entanglement measures satisfying the axioms (i-iii) of Definition 9.C.1 have been defined in the literature. Their restrictions to pure states are all given by (289) for specific concave functions f. We present in Subsection IX D a few of these measures, namely the entanglement of formation, the

concurrence, and the Schmidt number. An integer-valued entanglement measure has been introduced in Ref. 139 by using a symplectic geometry approach, but this goes beyond the scope of this article.

D. Entanglement of formation

1. Entanglement of formation for pure states

A natural choice for the function f is the von Neumann entropy. We set

$$E_{\mathrm{EoF}}(|\Psi\rangle) = S\big([\rho_\Psi]_{\mathsf{A}}\big) = S\big([\rho_\Psi]_{\mathsf{B}}\big) = -\sum_i \mu_i \ln \mu_i \ . \tag{293}$$

Then $E_{\mathrm{EoF}}(|\Psi\rangle) = 0$ if and only if $|\Psi\rangle$ is separable and $E_{\mathrm{EoF}}(|\Psi\rangle)$ is maximum (and equal to $\ln n$ with $n = \min\{n_{\mathsf{A}}, n_{\mathsf{B}}\}$) if and only if $|\Psi\rangle$ is maximally entangled. Since the von Neumann entropy is concave, Proposition 9.C.2 ensures that E_{EoF} is an entanglement measure on pure states.

An important result due to Bennett *et al.*[22] relates $E_{\mathrm{EoF}}(|\Psi\rangle)$ to entanglement distillation and entanglement cost, which consist in the following problems. The EPR two-qubit state $|\Phi_+\rangle = (|0\rangle|0\rangle + |1\rangle|1\rangle)/\sqrt{2} \in \mathbb{C}^4$ corresponds to an e-bit of information shared by Alice and Bob. One such e-bit is required, for instance, if Alice wants to teleport an unknown quantum state to Bob.[117] Entanglement distillation is the transformation of N copies of $|\Psi\rangle$ onto $M < N$ copies of $|\Phi_+\rangle$. It was demonstrated by Bennett *et al.*[22] that in the large N limit, $E_{\mathrm{EoF}}(|\Psi\rangle)$ is equal to the maximal rate of distillation M/N, the maximum being over all LOCC operations. Stated differently, $E_{\mathrm{EoF}}(|\Psi\rangle)$ is the highest number of e-bits per input copy of $|\Psi\rangle$ that can be distilled from $|\Psi\rangle$ via LOCCs. Conversely, $E_{\mathrm{EoF}}(|\Psi\rangle)$ is the smallest number of e-bits per unit copy of $|\Psi\rangle$ from which $|\Psi\rangle$ may be obtained via LOCCs. The precise mathematical statement is given in the proposition below.

Proposition 9.D.1. (Bennett *et al.*[22])

$$\frac{E_{\mathrm{EoF}}(|\Psi\rangle)}{\ln 2} = \sup\Big\{ r \ ; \ \lim_{N\to\infty} \Big(\inf_{\mathrm{LOCC}} \big\| \mathcal{M}^{(N)}_{\mathrm{LOCC}}(|\Psi^{\otimes N}\rangle\langle\Psi^{\otimes N}|) - |\Phi_+^{\otimes rN}\rangle\langle\Phi_+^{\otimes rN}| \big\|_1 \Big) = 0 \Big\} \tag{294}$$

$$= \inf\Big\{ r \ ; \ \lim_{N\to\infty} \Big(\inf_{\mathrm{LOCC}} \big\| |\Psi^{\otimes N}\rangle\langle\Psi^{\otimes N}| - \mathcal{M}^{(N)}_{\mathrm{LOCC}}(|\Phi_+^{\otimes rN}\rangle\langle\Phi_+^{\otimes rN}|) \big\|_1 \Big) = 0 \Big\} \ . \tag{295}$$

Let us stress that these identities are no longer valid for mixed states: then the right-hand sides of (294) and (295) are, in general, not equal. They define two measures of entanglement called the distillable entanglement and the entanglement cost (see Ref. 82 and references therein). The fact that these quantities coincide with $E_{\mathrm{EoF}}(|\Psi\rangle)$ for pure states basically indicates that, among all the possible entanglement measures, only one (namely, $E_{\mathrm{EoF}}(|\Psi\rangle)$) becomes relevant asymptotically when dealing with many copies of $|\Psi\rangle$.

Proof. A simple and illuminating proof due to Nielsen[116] is based on Proposition 9.B.1 and the Shannon equipartition theorem. It runs as follows. Let μ_i be the Schmidt coefficients of $|\Psi\rangle$. Consider N i.i.d. random variables with distribution $\{\mu_i\}$ and values in $I = \{1, \ldots, n\}$. The joint probabilities of these random variables are $p(\underline{i}) = \mu_{i_1} \ldots \mu_{i_N}$ with $\underline{i} = (i_1, \ldots, i_N) \in I^N$. Given $\varepsilon > 0$, the "most likely set" $\mathcal{A}_{N,\varepsilon} \subset I^N$ is by definition the set of all $\underline{i} \in I^N$ such that $2^{-N(H+\varepsilon)} \le p(\underline{i}) \le 2^{-N(H-\varepsilon)}$, where H is the Shannon entropy of $\{\mu_i\}$ (in our case, $H = E_{\mathrm{EoF}}(|\Psi\rangle)/\ln 2$) which is defined here using the binary logarithm. The Shannon equipartition theorem[143] tells us that $\mathcal{A}_{N,\varepsilon}$ has probability $P_{N,\varepsilon} > 1 - \varepsilon$ and cardinality $|\mathcal{A}_{N,\varepsilon}|$ satisfying $(1-\varepsilon)2^{N(H-\varepsilon)} \le |\mathcal{A}_{N,\varepsilon}| \le 2^{N(H+\varepsilon)}$ for sufficiently large N. The idea of Nielsen's proof is to approximate

$$\begin{aligned} |\Psi^{\otimes N}\rangle &= \sum_{\underline{i}\in I^N} \sqrt{p(\underline{i})}\, |\alpha_{i_1}\rangle \ldots |\alpha_{i_N}\rangle \otimes |\beta_{i_1}\rangle \ldots |\beta_{i_N}\rangle \\ &\simeq |\Phi_{N,\varepsilon}\rangle = \sum_{\underline{i}\in\mathcal{A}_{N,\varepsilon}} \sqrt{q(\underline{i})}\, |\alpha_{i_1}\rangle \ldots |\alpha_{i_N}\rangle \otimes |\beta_{i_1}\rangle \ldots |\beta_{i_N}\rangle \end{aligned} \tag{296}$$

with $q(\underline{i}) = p(\underline{i})/P_{N,\varepsilon}$ and $|\alpha_i\rangle$, $|\beta_i\rangle$ as in Theorem 2.B.1. Observe that the fidelity $|\langle\Psi^{\otimes N}|\Phi_{N,\epsilon}\rangle|^2 = P_{N,\varepsilon}$ is almost one for small ε. For any $\mathcal{A} \subset |\mathcal{A}_{N,\varepsilon}|$, one has

$$\frac{(1-\varepsilon)|\mathcal{A}|\,2^{-2N\varepsilon}}{|\mathcal{A}_{N,\varepsilon}|} \le \sum_{\underline{i}\in\mathcal{A}} q(\underline{i}) \le \frac{|\mathcal{A}|\,2^{2N\varepsilon}}{(1-\varepsilon)|\mathcal{A}_{N,\varepsilon}|}\ . \tag{297}$$

The second inequality implies that $\mathbf{q} = (q(\underline{i}))_{i\in\mathcal{A}_{N,\varepsilon}} \prec (2^{-M},\dots,2^{-M},0,\dots,0)$ with

$$M = \ln_2(|\mathcal{A}_{N,\varepsilon}|(1-\varepsilon)) - 2N\varepsilon\ . \tag{298}$$

By Proposition 9.B.1, this means that $|\Phi_{N,\varepsilon}\rangle$ can be transformed by a LOCC into the M-qubit state

$$|\Phi_+^{\otimes M}\rangle = \sum_{\underline{j}\in\{0,1\}^M} 2^{-\frac{M}{2}}|j_1\rangle\dots|j_M\rangle\otimes|j_1\rangle\dots|j_M\rangle\ . \tag{299}$$

We conclude that for N sufficiently large there exists a LOCC operation $\mathcal{M}_{\mathrm{LOCC}}^{(N,\varepsilon)}$ from $\mathcal{B}(\mathcal{H}_{\mathsf{AB}}^{\otimes N})$ into $\mathcal{B}(\mathbb{C}^{\otimes 2M})$ such that

$$\begin{aligned}\left\|\mathcal{M}_{\mathrm{LOCC}}^{(N,\varepsilon)}(|\Psi^{\otimes N}\rangle\langle\Psi^{\otimes N}|) - |\Phi_+^{\otimes M}\rangle\langle\Phi_+^{\otimes M}|\right\|_1 &\le \left\||\Psi^{\otimes N}\rangle\langle\Psi^{\otimes N}| - |\Phi_{N,\varepsilon}\rangle\langle\Phi_{N,\varepsilon}|\right\|_1 \\ &\le 2\big(1-\big|\langle\Psi^{\otimes N}|\Phi_{N,\varepsilon}\rangle\big|^2\big)^{\frac{1}{2}} \ \le\ 2\sqrt{\varepsilon}\end{aligned} \tag{300}$$

(we have used Propositions 7.A.2 and 7.D.1 to get the first and second inequalities, respectively). In addition, the distillation rate M/N is bounded from below by $H - 3\varepsilon + 2N^{-1}\ln(1-\varepsilon)$. Taking, e.g., $\varepsilon = 1/\sqrt{N}$, this proves that $E_{\mathrm{EoF}}(|\Psi\rangle) \le E_D(|\Psi\rangle)$, where $E_D(|\Psi\rangle)$ denotes the right-hand side of (294).

Similarly, the first inequality in (297) implies that $|\Phi_{N,\varepsilon}\rangle$ can be obtained asymptotically by transforming M' copies of $|\Phi_+\rangle$ with LOCCs, more precisely it shows the existence of a LOCC operation $\mathcal{M}_{\mathrm{LOCC}}^{(N,\varepsilon)'}$ such that

$$\left\||\Psi^{\otimes N}\rangle\langle\Psi^{\otimes N}| - \mathcal{M}_{\mathrm{LOCC}}^{(N,\varepsilon)'}(|\Phi_+^{\otimes M'}\rangle\langle\Phi_+^{\otimes M'}|\right\|_1 \le 2\sqrt{\varepsilon} \tag{301}$$

for N large enough, with

$$M' = \ln_2(|\mathcal{A}_{N,\varepsilon}|/(1-\varepsilon)) + 2N\varepsilon\ . \tag{302}$$

The production rate M'/N is bounded from above by $H + 3\varepsilon - N^{-1}\ln(1-\varepsilon)$. This establishes that $E_{\mathrm{EoF}}(|\Psi\rangle) \ge E_C(|\Psi\rangle)$, where $E_C(|\Psi\rangle)$ denotes the right-hand side of (295). But $E_D(|\Psi\rangle) \le E_C(|\Psi\rangle)$, as otherwise one could transform asymptotically by a LOCC $r'N$ e-bits into rN e-bits with $r' < r$, which is impossible. Hence $E_{\mathrm{EoF}}(|\Psi\rangle = E_D(|\Psi\rangle) = E_C(|\Psi\rangle)$. □

2. Convex roof constructions

The extension of E_{EoF} to mixed states is done via a convex roof construction.[23]

Definition 9.D.2. *The entanglement of formation of a mixed state $\rho \in \mathcal{E}(\mathcal{H}_{\mathsf{AB}})$ is*

$$E_{\mathrm{EoF}}(\rho) = \min_{\{|\Psi_i\rangle,\eta_i\}}\left\{\sum_i \eta_i E_{\mathrm{EoF}}(|\Psi_i\rangle)\right\}\ , \tag{303}$$

where the minimum is over all pure state decompositions $\rho = \sum_i \eta_i|\Psi_i\rangle\langle\Psi_i|$ of ρ.

Proposition 9.D.3. (Vidal[161]) *$E_{\mathrm{EoF}}(\rho)$ is an entanglement measure with values in the interval $[0, \ln n]$. It satisfies the monotonicity condition (which is stronger than (iii))*

(iii″) $\sum_i p_i E_{\mathrm{EoF}}(p_i^{-1}\mathcal{M}_{\mathrm{loc}}^{(i)}(\rho)) \le E_{\mathrm{EoF}}(\rho)$ *with* $p_i = \mathrm{tr}[\mathcal{M}_{\mathrm{loc}}^{(i)}(\rho)]$, *for any family of CP local maps* $\mathcal{M}_{\mathrm{loc}}^{(i)}$ *with Kraus operators* $\{A_{ij}\otimes B_{ik}\}_{j,k}$ *such that* $\sum_{i,j,k} A_{ij}^* A_{ij}\otimes B_{ik}^* B_{ik} = 1$.

Note that the maps $\mathcal{M}_{\mathrm{loc}}^{(i)}$ are not required to be trace preserving (but $\mathrm{tr}[\mathcal{M}_{\mathrm{loc}}^{(i)}(\rho)] \le 1$). Modulo a state normalization, they describe wavepacket reduction processes, see (37).

Proof. One has clearly $0 \le E_{\mathrm{EoF}}(\rho) \le \ln n$. We now argue that E_{EoF} satisfies all the axioms (i-iii) of an entanglement measure. In fact, E_{EoF} is convex by construction. Moreover, it follows from the aforementioned properties of $E_{\mathrm{EoF}}(|\Psi\rangle)$ and the definition of mixed state entanglement (Sec. II D) that $E_{\mathrm{EoF}}(\rho) = 0$ if and only if $\rho \in \mathcal{S}_{\mathsf{AB}}$. Finally, the monotonicity with respect to LOCC operations is a consequence of the convexity and can be shown as follows. Let $\rho = \sum_i \eta_i |\Psi_i\rangle\langle\Psi_i|$ be the pure state decomposition minimizing the average entanglement in the right-hand side of (303). Let $\mathcal{M}$ be a separable operation with Kraus operators $A_j \otimes B_j$. We denote by $\eta_{j|i} = \|A_j\otimes B_j|\Psi_i\rangle\|^2$ the probability of outcome j given that the state is $|\Psi_i\rangle$. From the convexity of E_{EoF} and its monotonicity (iii′) for pure states (which holds by Proposition 9.C.2) one finds

$$\begin{aligned} E_{\mathrm{EoF}}\big(\mathcal{M}(\rho)\big) &\le \sum_i \eta_i E_{\mathrm{EoF}}\big(\mathcal{M}(|\Psi_i\rangle\langle\Psi_i|)\big) \le \sum_{ij} \eta_i \eta_{j|i} E_{\mathrm{EoF}}\big(\eta_{j|i}^{-\frac{1}{2}} A_j\otimes B_j|\Psi_i\rangle\big) \\ &\le \sum_i \eta_i E_{\mathrm{EoF}}(|\Psi_i\rangle\langle\Psi_i|) = E_{\mathrm{EoF}}(\rho). \end{aligned} \tag{304}$$

Thus E_{EoF} is an entanglement measure. A similar reasoning shows that E_{EoF} satisfies (iii″). □

More generally, one can construct entanglement measures by extending to mixed states any entanglement measure on pure states via a convex roof construction analog to (303). One gets in this way a family of measures E_f depending on the choice of the function f in Proposition 9.C.2. Conversely, any entanglement measure E satisfying the axiom (iii″) above coincides with E_f on pure states for some function f fulfilling the assumptions of Proposition 9.C.2.[161] In particular, this suggests to define the concurrence for mixed states as

$$C(\rho) = \min_{\{|\Psi_i\rangle,\eta_i\}} \left\{ \sum_i \eta_i C(|\Psi_i\rangle) \right\}, \tag{305}$$

where $C(|\Psi_i\rangle)$ is given by (282). It is known that $\rho_{\mathsf{A}} \mapsto \|\rho_{\mathsf{A}}\|_{1/2} = (\mathrm{tr}[\rho_{\mathsf{A}}^{1/2}])^2$ is concave (see (B1) in Appendix B), whence $C(\rho)$ is an entanglement measure. Another measure of entanglement of common use for pure states is the Schmidt number obtained by choosing $f(\rho_{\mathsf{A}}) = 1/\mathrm{tr}(\rho_{\mathsf{A}}^2)$ in Proposition 9.C.2.

As stated above, (iii″) means that separable measurements cannot increase the average entanglement, but entanglement can increase if one considers conditional expectations over subgroups of outcomes, i.e., one may have $E_{\mathrm{EoF}}(p_i^{-1}\mathcal{M}_{\mathrm{loc}}^{(i)}(\rho)) \ge E_{\mathrm{EoF}}(\rho)$ for some i. An example is given by the qutrit-qutrit system in the state

$$\rho = \frac{1}{2}|\Phi_+\rangle\langle\Phi_+| + \frac{1}{2}|2\rangle\langle 2|\otimes|2\rangle\langle 2|, \quad |\Phi_+\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle|0\rangle + |1\rangle|1\rangle\big). \tag{306}$$

Assume that Alice and Bob perform each a von Neumann measurement with projectors Π_1 onto $\mathrm{span}\{|0\rangle, |1\rangle\}$ and Π_2 onto $\mathbb{C}|2\rangle$. The conditional states $\rho_{\mathsf{AB}|11} = |\Phi_+\rangle\langle\Phi_+|$ and $\rho_{\mathsf{AB}|22} = |2\rangle\langle 2|\otimes|2\rangle\langle 2|$ have entanglement of formations $\ln 2$ and 0, respectively. The first value is larger than $E_{\mathrm{EoF}}(\rho)$, which is equal to $\ln 2/2$ according to the following result.

Corollary 9.D.4. *Let* ρ_1 *and* ρ_2 *be two states on* $\mathcal{H}_{\mathsf{AB}}$ *with bi-orthogonal supports* $\mathrm{ran}\,\rho_i \subset \mathcal{V}_i^{\mathsf{A}}\otimes\mathcal{V}_i^{\mathsf{B}}$, *where* $\mathcal{V}_i^{\mathsf{A}} \subset \mathcal{H}_{\mathsf{A}}$ *and* $\mathcal{V}_i^{\mathsf{B}} \subset \mathcal{H}_{\mathsf{B}}$ *are such that* $\mathcal{V}_2^{\mathsf{A}} = (\mathcal{V}_1^{\mathsf{A}})^\perp$ *and* $\mathcal{V}_2^{\mathsf{B}} = (\mathcal{V}_1^{\mathsf{B}})^\perp$. *Let* $\rho = \eta_1\rho_1 + \eta_2\rho_2$ *with* $\eta_i \ge 0$, $\eta_1 + \eta_2 = 1$. *Then* $E_{\mathrm{EoF}}(\rho) = \eta_1 E_{\mathrm{EoF}}(\rho_1) + \eta_2 E_{\mathrm{EoF}}(\rho_2)$.

Proof. The inequality $E_{\rm EoF}(\rho) \leq \eta_1 E_{\rm EoF}(\rho_1) + \eta_2 E_{\rm EoF}(\rho_2)$ follows from convexity. The reverse inequality is a consequence of the monotonicity property (iii″) applied to the maps

$$\mathcal{M}^{(i)}_{\rm loc}(\rho) = \pi_i^{\mathsf{A}} \otimes \pi_i^{\mathsf{B}}\, \rho\, \pi_i^{\mathsf{A}} \otimes \pi_i^{\mathsf{B}}\,,\ i=1,2, \quad \mathcal{M}^{(3)}_{\rm loc}(\rho) = \pi_1^{\mathsf{A}} \otimes \pi_2^{\mathsf{B}}\, \rho\, \pi_1^{\mathsf{A}} \otimes \pi_2^{\mathsf{B}} + \pi_2^{\mathsf{A}} \otimes \pi_1^{\mathsf{B}}\, \rho\, \pi_2^{\mathsf{A}} \otimes \pi_1^{\mathsf{B}}, \tag{307}$$

where π_i^{A} and π_i^{B} are the projectors onto $\mathcal{V}_i^{\mathsf{A}}$ and $\mathcal{V}_i^{\mathsf{B}}$, respectively. □

It is worth realizing the link between $E_{\rm EoF}(\rho)$ and the classical mutual information $I_{X:\,Y}$, where $X = \{\eta_i\}$ is associated to a pure state decomposition $\{|\Psi_i\rangle, \eta_i\}$ of ρ and Y to the outcomes of a local measurement on A (Sec. V F). Indeed, the maximum of $I_{X:\,Y}$ over all pure state decompositions and all POVMs on A is bounded by

$$\max_{\{|\Psi_i\rangle,\eta_i\},\{M_i^{\mathsf{A}}\}} \left\{ I_{X:Y} \right\} \leq S(\rho_{\mathsf{A}}) - E_{\rm EoF}(\rho)\,. \tag{308}$$

This inequality is a direct consequence of the Holevo bound (113) and the definition (303) of $E_{\rm EoF}(\rho)$.

3. The Wootters formula for two qubits

The main problem with the convex-roof construction (303) is that finding the pure state decomposition minimizing the average entanglement is a non-trivial task. Nevertheless, an astonishing formula enabling to evaluate $E_{\rm EoF}(\rho)$ explicitly for two qubits was found by Wootters.[169] It reads

$$E_{\rm EoF}(\rho) = h(C(\rho)), \tag{309}$$

where $C(\rho)$ is given by (305) and h: $[0,1] \to [0, \ln n]$ is the convex increasing function

$$h(C) = -\frac{1+\sqrt{1-C^2}}{2} \ln\!\left(\frac{1+\sqrt{1-C^2}}{2}\right) - \frac{1-\sqrt{1-C^2}}{2} \ln\!\left(\frac{1-\sqrt{1-C^2}}{2}\right). \tag{310}$$

The main point is that $C(\rho)$ can be calculated explicitly as follows. Let $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$ be the square roots of the eigenvalues of $\rho\sigma_y \otimes \sigma_y\, \overline{\rho}\, \sigma_y \otimes \sigma_y$ (here σ_y is the y-Pauli matrix and $\overline{\rho} = J\rho J$ the complex conjugate of ρ in the canonical basis). Then

$$C(\rho) = \max\{0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4\}. \tag{311}$$

For pure states this yields $C(|\Psi\rangle) = |\langle\Psi|\sigma_y \otimes \sigma_y J|\Psi\rangle|^2$, in agreement with the result of Sec. IX A. The proof of (309) is somehow tricky but relies on simple linear algebra arguments (see Ref. 169).

E. Maximally entangled states

One may expect intuitively that the most entangled states are extremal states in $\mathcal{E}(\mathcal{H}_{\mathsf{AB}})$, that is, they are the pure maximally entangled states described in Sec. II D. If one uses as a criterion for being mostly entangled the property of having the highest entanglement of formation, this is indeed correct when the dimensions of $\mathcal{H}_{\mathsf{A}}$ and $\mathcal{H}_{\mathsf{B}}$ are such that $n_{\mathsf{A}}/2 < n_{\mathsf{B}} < 2n_{\mathsf{A}}$. When $n_{\mathsf{B}} \geq 2n_{\mathsf{A}}$, convex combinations of pure maximally entangled states with reduced B-states living on orthogonal subspaces of $\mathcal{H}_{\mathsf{B}}$ are also maximally entangled (a similar statement holds of course by exchanging A and B).

Proposition 9.E.1. *Assume that $n = n_{\mathsf{A}} \leq n_{\mathsf{B}}$ and let $r = 1, 2, \ldots$ be such that $rn_{\mathsf{A}} \leq n_{\mathsf{B}} < (r+1)n_{\mathsf{A}}$. Then the states $\rho \in \mathcal{E}(\mathcal{H}_{\mathsf{AB}})$ having a maximal entanglement of formation $E_{\rm EoF}(\rho) = \ln n$ are convex combinations of the r orthogonal maximally entangled states*

$$|k\rangle = n^{-\frac{1}{2}} \sum_{i=1}^{n} |\alpha_i^{(k)}\rangle \otimes |\beta_i^{(k)}\rangle, \quad k = 1, \ldots, r\,, \tag{312}$$

with $\langle\alpha_i^{(k)}|\alpha_j^{(k)}\rangle = \delta_{ij}$ and $\langle\beta_i^{(k)}|\beta_j^{(l)}\rangle = \delta_{kl}\delta_{ij}$.

Proof. Let ρ be a state with $E_{\rm EoF}(\rho) = \ln n$. According to Definition 9.D.2 and given that $E_{\rm EoF}(|\Psi\rangle) \leq \ln n$ with equality if and only if $|\Psi\rangle$ is maximally entangled, this means that any

pure state decomposition of ρ is made of maximally entangled states. This is the case in particular for the spectral decomposition $\rho = \sum_k p_k |k\rangle\langle k|$, from which one can obtain all other pure state decompositions $\{|\Psi_i\rangle, \eta_i\}$ by the formula $\sqrt{\eta_i}|\Psi_i\rangle = \sum_k u_{ik}\sqrt{p_k}|k\rangle$ with $\eta_i = \sum_k |u_{ik}|^2 p_k$ (see (16)). Let us set $D_{kl} = \mathrm{tr}_{\mathsf{B}}(|k\rangle\langle l|)$. We would like to show that $D_{kl} = n^{-1}\delta_{kl}$ if $p_k p_l \neq 0$. We already know that $D_{kk} = 1/n$ if $p_k \neq 0$, since $|k\rangle$ is maximally entangled. By plugging the above expression of $\sqrt{\eta_i}|\Psi_i\rangle$ into $\mathrm{tr}_{\mathsf{B}}(|\Psi_i\rangle\langle\Psi_i|) = 1/n$, one is led to

$$\sum_{k,l,k\neq l} \sqrt{p_k p_l}\, u_{ik}\overline{u}_{il} D_{kl} = 0. \tag{313}$$

This equality holds for any i and any unitary matrix (u_{ik}), hence $\sqrt{p_k p_l} D_{kl} = 0$ if $k \neq l$ and the above claim is true. One deduces from $D_{kk} = 1/n$ that the eigenvectors $|k\rangle$ with eigenvalues $p_k > 0$ have Schmidt decompositions given by (312). For $k \neq l$, $D_{kl} = 0$ is then equivalent to $\mathcal{V}_{\mathsf{B}}^{(k)} \perp \mathcal{V}_{\mathsf{B}}^{(l)}$ with $\mathcal{V}_{\mathsf{B}}^{(k)} = \mathrm{span}\{|\beta_i^{(k)}\rangle\}_{i=1}^n \subset \mathcal{H}_{\mathsf{B}}$. If $n_{\mathsf{B}} < (r+1)n$ then at most r subspaces $\mathcal{V}_{\mathsf{B}}^{(k)}$ may be pairwise orthogonal. Thus at most r eigenvalues p_k are non-zero. □

X. THE QUANTUM DISCORD

The quantum discord was introduced by Ollivier and Zurek[120] and Henderson and Vedral[75] as an indicator of the "degree of quantumness" of mixed states. For pure states it coincides with the entanglement of formation. Certain separable mixed states have, however, a non-zero discord. These states are obtained by preparing locally mixtures of non-orthogonal states, which cannot be perfectly discriminated by local measurements. Such separable states cannot be classified as "classical" and actually contain quantum correlations that are not captured by the entanglement measures reviewed in Sec. IX. Apart from this observation, a motivation for the quantum discord came out in the last decade from the claim that it could play the role of a resource in certain quantum algorithms and quantum communication protocols.[49,97,123,107,66,47] In particular, it has been suggested[49,97,123] that the discord might capture the quantum correlations at the origin of the quantum speedup in the deterministic quantum computation with one qubit (DQC1) of Knill and Laflamme.[93] The DQC1 algorithm computes the trace of a $2^N \times 2^N$ unitary matrix exponentially faster than all known classical algorithms. The entanglement produced during the computation with $(N+1)$ qubits is bounded independently of N, for any bipartition of the $(N+1)$ qubits.[48] This means that the total amount of bipartite entanglement is a negligible fraction of the maximal entanglement possible. However, a non-vanishing quantum discord between the control qubit and the N target qubits appears during the computation,[49] save for particular unitaries.[46] The DCQ1 algorithm is singled out by the fact that it uses mixed states, the N target qubits being initially in a Gibbs state at infinite temperature. In contrast, for quantum computations using pure states, Jozsa and Linden[92] have shown that in order to offer an exponential speedup over classical computers, the computation must produce entanglement which is not restricted to qubit blocks of fixed size as the problem size increases.

A. Definition of the quantum discord

Let us first consider some classical discrete random variables A and B with joint probabilities p_{ij} and marginals $p_A(i) = \sum_j p_{ij}$ and $p_B(j) = \sum_i p_{ij}$. The correlations between A and B are measured by the mutual information $I_{A:B} = H(A) + H(B) - H(A,B)$. We recall from Sec. V F that

$$I_{A:B} = H(B) - H(B|A), \tag{314}$$

where $H(B|A) = \sum_i p_A(i) H(B|i)$ is the conditional entropy, see (110). This conditional entropy describes the amount of information on B left after the value $A = i$ has been measured, averaged over all possible outcomes i.

In the quantum setting, the analog of the random variables A and B is a bipartite quantum system AB in a state ρ. The marginals are the reduced states $\rho_{\mathsf{A}} = \mathrm{tr}_{\mathsf{B}}(\rho)$ and $\rho_{\mathsf{B}} = \mathrm{tr}_{\mathsf{A}}(\rho)$. The

generalization of the mutual information reads

$$I_{\mathsf{A}:\mathsf{B}}(\rho) = S(\rho_{\mathsf{A}}) + S(\rho_{\mathsf{B}}) - S(\rho), \tag{315}$$

where $S(\cdot)$ is the von Neumann entropy (114). Similarly to the classical case, one has $I_{\mathsf{A}:\mathsf{B}}(\rho) \geq 0$ and $I_{\mathsf{A}:\mathsf{B}}(\rho) = 0$ if and only if ρ is a product state, i.e., $\rho = \rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}}$ (this is nothing but the subadditivity property of S, see Sec. VI A). It is easy to verify that $I_{\mathsf{A}:\mathsf{B}}(\rho)$ is related to the relative entropy (122) by

$$I_{\mathsf{A}:\mathsf{B}}(\rho) = S(\rho || \rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}}). \tag{316}$$

By the monotonicity of the relative entropy (Theorem 6.B.1), $I_{\mathsf{A}:\mathsf{B}}(\mathcal{M}_{\rm loc}(\rho)) \leq I_{\mathsf{A}:\mathsf{B}}(\rho)$ for any local operation $\mathcal{M}_{\rm loc} = \mathcal{M}_{\mathsf{A}} \otimes \mathcal{M}_{\mathsf{B}}$, where the operations $\mathcal{M}_{\mathsf{A}} : \mathcal{B}(\mathcal{H}_{\mathsf{A}}) \to \mathcal{B}(\mathcal{H}'_{\mathsf{A}})$ and $\mathcal{M}_{\mathsf{B}} : \mathcal{B}(\mathcal{H}_{\mathsf{B}}) \to \mathcal{B}(\mathcal{H}'_{\mathsf{B}})$ may have different initial and final spaces (for instance, $\mathcal{M}_{\mathsf{A}}$ can be the partial trace over a part of A).

However, there is no quantum analog of the identity (314). Let us define a conditional entropy of B given a von Neumann measurement $\{\pi_i^{\mathsf{A}}\}$ on A by $S_{\mathsf{B}|\mathsf{A}}(\rho|\{\pi_i^{\mathsf{A}}\}) = \sum_i \eta_i S(\rho_{\mathsf{B}|i})$, where

$$\rho_{\mathsf{B}|i} = \eta_i^{-1} \operatorname{tr}_{\mathsf{A}}(\pi_i^{\mathsf{A}} \otimes 1\, \rho), \qquad \eta_i = \operatorname{tr}(\pi_i^{\mathsf{A}} \otimes 1\, \rho)\,. \tag{317}$$

Here η_i is the probability of the measurement outcome i and $\rho_{\mathsf{B}|i} = \operatorname{tr}_{\mathsf{A}}(\rho_{\mathsf{A}\mathsf{B}|i})$ is the corresponding conditional state of B (see Sec. III). The ensemble $\{\rho_{\mathsf{B}|i}, \eta_i\}$ defines a convex decomposition of ρ_{B} (i.e., $\rho_{\mathsf{B}} = \sum_i \eta_i \rho_{\mathsf{B}|i}$) describing a state preparation of the subsystem B realized by the measurement on A. The quantum version of the right-hand side of (314) is the maximal reduction of entropy of B due to a von Neumann measurement on A,

$$J^{\text{v.N.}}_{\mathsf{B}|\mathsf{A}}(\rho) = S(\rho_{\mathsf{B}}) - \min_{\{\pi_i^{\mathsf{A}}\}} \left\{ \sum_i \eta_i S(\rho_{\mathsf{B}|i}) \right\}, \tag{318}$$

the minimum being over all orthonormal families of projectors on $\mathcal{H}_{\mathsf{A}}$. This quantity represents the classical correlations between A and B (see the discussion after Proposition 10.A.2 below). Note that $J^{\text{v.N.}}_{\mathsf{B}|\mathsf{A}}(\rho)$ places an upper bound on the classical mutual information between the ensemble $\{\rho_{\mathsf{B}|i}, \eta_i\}$ and the outcome probabilities when performing measurements on B to discriminate the states $\rho_{\mathsf{B}|i}$ (Sec. V F). Actually, $J^{\text{v.N.}}_{\mathsf{B}|\mathsf{A}}(\rho)$ coincides with the corresponding Holevo quantity (113). By concavity of the von Neumann entropy, one has $J^{\text{v.N.}}_{\mathsf{B}|\mathsf{A}}(\rho) \geq 0$. Furthermore, (121) entails $J^{\text{v.N.}}_{\mathsf{B}|\mathsf{A}}(\rho) \leq \max_{\{\pi_i^{\mathsf{A}}\}} H(\{\eta_i\})$.

It also follows from the concavity of S that the minimum in (318) is achieved for rank-one projectors. In fact, by decomposing each projector π_i^{A} of rank r_i as a sum of r_i rank-one projectors π_{ik}^{A}, one finds that $\rho_{\mathsf{B}|i} = \sum_k (\eta_{ik}/\eta_i) \rho_{\mathsf{B}|ik}$ is a convex combination of the states $\rho_{\mathsf{B}|ik} = \eta_{ik}^{-1} \operatorname{tr}_{\mathsf{A}}(\pi_{ik}^{\mathsf{A}} \otimes 1\, \rho)$ if $\eta_i = \sum_k \eta_{ik} > 0$. Thereby $\sum_i \eta_i S(\rho_{\mathsf{B}|i}) \geq \sum_{ik} \eta_{ik} S(\rho_{\mathsf{B}|ik})$.

Ollivier and Zurek[120] and Henderson and Vedral[75] proposed in two independent works published in 2001 to characterize the amount of non-classicality in the state ρ by forming the difference between the total correlations given by $I_{\mathsf{A}:\mathsf{B}}(\rho)$ and the classical correlations given by $J^{\text{v.N.}}_{\mathsf{B}|\mathsf{A}}(\rho)$.

Definition 10.A.1. *The quantum discord of the bipartite system* AB *in state* ρ *is*

$$\delta^{\text{v.N.}}_{\mathsf{A}}(\rho) = I_{\mathsf{A}:\mathsf{B}}(\rho) - J^{\text{v.N.}}_{\mathsf{B}|\mathsf{A}}(\rho) = S(\rho_{\mathsf{A}}) - S(\rho) + \min_{\{\pi_i^{\mathsf{A}}\}} \left\{ \sum_i \eta_i S(\rho_{\mathsf{B}|i}) \right\}. \tag{319}$$

In Ref. 75, the minimization is done over generalized measurements given by POVMs $\{M_i^{\mathsf{A}}\}$ on $\mathcal{H}_{\mathsf{A}}$, instead of von Neumann measurements. The conditional states and outcome probabilities are then (Sec. III)

$$\rho_{\mathsf{B}|i} = \eta_i^{-1} \operatorname{tr}_{\mathsf{A}}(M_i^{\mathsf{A}} \otimes 1\, \rho), \qquad \eta_i = \operatorname{tr}(M_i^{\mathsf{A}} \otimes 1\, \rho). \tag{320}$$

We denote the corresponding discord by $\delta_{\mathsf{A}}(\rho)$. As in the case of von Neumann measurements, the minimum is achieved for rank-one measurement operators M_i^{A}. In general, the inequality $\delta_{\mathsf{A}}(\rho) < \delta^{\text{v.N.}}_{\mathsf{A}}(\rho)$ is strict (see, e.g., Refs. 68 and 62 for a comparison of the von Neumann and POVM discords for two qubits.) Nevertheless, by the Neumark extension theorem, δ_{A} coincides with $\delta^{\text{v.N.}}_{\mathsf{A}}$ up to an

enlargement of the space $\mathcal{H}_{\mathsf{A}}$. More precisely, by plugging $M_i^{\mathsf{A}} = \langle\epsilon_0|\Pi^{\mathsf{AE}}|\epsilon_0\rangle$ (see Remark 3.C.3) into (320) and using the additivity of S under tensor products, a simple calculation gives

$$\delta_{\mathsf{A}}(\rho) = \delta_{\mathsf{AE}}^{\text{v.N.}}(\rho \otimes |\epsilon_0\rangle\langle\epsilon_0|, \tag{321}$$

the right-hand side being independent of the ancilla state $|\epsilon_0\rangle \in \mathcal{H}_{\mathsf{E}}$.

The discords $\delta_{\mathsf{A}}^{\text{v.N.}}(\rho)$ and $\delta_{\mathsf{A}}(\rho)$ thus measure the amount of total correlations between A and B which cannot be accessed by local measurements on the subsystem A. Note that they are asymmetric under the exchange $\mathsf{A} \leftrightarrow \mathsf{B}$. One can define similarly the discords $\delta_{\mathsf{B}}^{\text{v.N.}}(\rho)$ and $\delta_{\mathsf{B}}(\rho)$ by performing the measurements on the subsystem B.

For pure states $\rho_\Psi = |\Psi\rangle\langle\Psi|$, the mutual information $I_{\mathsf{A}:\mathsf{B}}(\rho_\Psi)$ is equal to $2S([\rho_\Psi]_{\mathsf{B}})$, see (117), and the measurement minimizing the conditional entropy of B is the measurement in the eigenbasis $\{|\alpha_i\rangle\}$ of the reduced state $[\rho_\Psi]_{\mathsf{A}}$. In fact, according to (9) the corresponding post-measurement states $\rho_{\mathsf{B}|i} = |\beta_i\rangle\langle\beta_i|$ are pure and thus have zero entropy. Then (318) yields $J_{\mathsf{B}|\mathsf{A}}(\rho_\Psi) = S([\rho_\Psi]_{\mathsf{B}})$. As a result, the discords coincide for pure states with the entanglement of formation,

$$\delta_{\mathsf{A}}(|\Psi\rangle) = \delta_{\mathsf{A}}^{\text{v.N.}}(|\Psi\rangle) = \delta_{\mathsf{B}}(|\Psi\rangle) = \delta_{\mathsf{B}}^{\text{v.N.}}(|\Psi\rangle) = E_{\text{EoF}}(|\Psi\rangle). \tag{322}$$

For mixed states, it was pointed out in Ref. 120 that if the measurement operators M_i^{A} are of rank one then

$$\sum_i \eta_i S(\rho_{\mathsf{B}|i}) = S\big(\mathcal{M}_{\mathsf{A}} \otimes 1(\rho)\big) - S\big([\mathcal{M}_{\mathsf{A}} \otimes 1(\rho)]_{\mathsf{A}}\big) = -I_{\mathsf{A}:\mathsf{B}}\big(\mathcal{M}_{\mathsf{A}} \otimes 1(\rho)\big) + S(\rho_{\mathsf{B}}), \tag{323}$$

where $\mathcal{M}_{\mathsf{A}}$ is the quantum operation on A associated to the measurement. Actually, consider the family of Kraus operators for $\mathcal{M}_{\mathsf{A}}$ given by $\{A_i = |i\rangle\langle\widetilde{\mu}_i|\}$, where $|\widetilde{\mu}_i\rangle$ are unnormalized vectors such that $M_i^{\mathsf{A}} = |\widetilde{\mu}_i\rangle\langle\widetilde{\mu}_i|$ and $\{|i\rangle\}$ is an orthonormal basis of a pointer space $\mathcal{H}_{\mathsf{P}}$. Then $\mathcal{M}_{\mathsf{A}} \otimes 1(\rho) = \sum_i \eta_i |i\rangle\langle i| \otimes \rho_{\mathsf{B}|i}$ and the reduced state $[\mathcal{M}_{\mathsf{A}} \otimes 1(\rho)]_{\mathsf{A}} = \sum_i \eta_i |i\rangle\langle i|$ has entropy $-\sum_i \eta_i \ln \eta_i$. A simple calculation yields the first equality in (323). The second equality is clear once one notices that $[\mathcal{M}_{\mathsf{A}} \otimes 1(\rho)]_{\mathsf{B}} = \rho_{\mathsf{B}}$.

Therefore, by combining (318), (319), and (316) one obtains the following result.

Proposition 10.A.2.[105] *The discord $\delta(\rho) = I_{\mathsf{A}:\mathsf{B}}(\rho) - J_{\mathsf{B}|\mathsf{A}}(\rho)$ is the minimal difference of mutual information of AB before and after a measurement on A, i.e.,*

$$J_{\mathsf{B}|\mathsf{A}}(\rho) = \max_{\{M_i^{\mathsf{A}}\}} \big\{ I_{\mathsf{A}:\mathsf{B}}(\mathcal{M}_{\mathsf{A}} \otimes 1(\rho)) \big\}, \tag{324}$$

where the maximum is over all POVMs on A with rank-one operators M_i^{A} and $\mathcal{M}_{\mathsf{A}}$ is the associated quantum operation on $\mathcal{B}(\mathcal{H}_{\mathsf{A}})$. As a result,

$$\delta_{\mathsf{A}}(\rho) = \min_{\{M_i^{\mathsf{A}}\}} \Big\{ S\big(\rho||\rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}}\big) - S\big(\mathcal{M}_{\mathsf{A}} \otimes 1(\rho)||\mathcal{M}_{\mathsf{A}}(\rho_{\mathsf{A}}) \otimes \rho_{\mathsf{B}}\big) \Big\}. \tag{325}$$

Similarly, $J_{\mathsf{B}|\mathsf{A}}^{\text{v.N.}}(\rho)$ is given by maximizing $I_{\mathsf{A}:\mathsf{B}}(\mathcal{M}_{\pi^{\mathsf{A}}} \otimes 1(\rho))$ over all von Neumann measurements $\mathcal{M}_{\pi^{\mathsf{A}}}$ on A of the form (26) with rank-one projectors π_i^{A}.

Observing that a measurement on A with no readout removes the quantum correlations between A and B, the right-hand side of (324) can be interpreted as the amount of classical correlations between the two subsystems. These subsystems are not correlated classically, i.e., $J_{\mathsf{B}|\mathsf{A}}(\rho) = 0$, if and only if $\rho = \rho_{\mathsf{A}} \otimes \rho_{\mathsf{B}}$ is a product state. This result holds for $J_{\mathsf{B}|\mathsf{A}}^{\text{v.N.}}(\rho)$ as well. Actually, by (324), $J_{\mathsf{B}|\mathsf{A}}(\rho) = 0$ is equivalent to $\mathcal{M}_{\mathsf{A}} \otimes 1(\rho)$ being a product state for any collection of operators $M_i^{\mathsf{A}} = |\widetilde{\mu}_i\rangle\langle\widetilde{\mu}_i|$ forming a POVM. This implies $\eta_i \rho_{\mathsf{B}|i} = \langle\widetilde{\mu}_i|\rho|\widetilde{\mu}_i\rangle = \eta_i \rho_{\mathsf{B}}$ for all i (see the discussion before Proposition 10.A.2). Choosing the $|\widetilde{\mu}_i\rangle$ to be the eigenvectors of the observable A, one obtains that $\langle A \otimes B\rangle_\rho = \langle A \otimes 1\rangle_\rho \langle 1 \otimes B\rangle_\rho$ for any $A \in \mathcal{B}(\mathcal{H}_{\mathsf{A}})_{\text{s.a.}}$ and $B \in \mathcal{B}(\mathcal{H}_{\mathsf{B}})_{\text{s.a.}}$, with $\langle\cdot\rangle_\rho = \text{tr}(\cdot\rho)$.

Let us emphasize that finding the optimal measurement which maximizes the post-meas mutual information, and hence calculating the discords $\delta_{\mathsf{A}}^{\text{v.N.}}(\rho)$ and $\delta_{\mathsf{A}}(\rho)$, is a difficult problem in general. Even for two qubits, this problem has been solved so far for a restricted family of states only,

namely, the states ρ with maximally mixed marginals $\rho_{\mathsf{A}} = \rho_{\mathsf{B}} = 1/2$.[104] In other cases the discords must be evaluated numerically (however, $\delta_{\mathsf{A}}(\rho)$ can be determined analytically for low-rank density matrices with the help of the monogamy relation, see Sec. X D and Ref. 110). An incorrect work[6] claiming to extend the result of Ref. 104 to the larger family of the so-called *X*-states has generated a profusion of articles. Comparing with numerical evaluations, the result of Ref. 6 apparently gives good approximations of the discord for randomly chosen *X*-states (see the discussion in Ref. 110).

B. The A-classical states

The monotonicity property of the relative entropy and formula (325) imply that $\delta_{\mathsf{A}}(\rho)$ is non-negative. The states with vanishing discord can be determined with the help of Theorem 6.B.1, leading to the following result.

Proposition 10.B.1. *The quantum discord is non-negative and* $\delta_{\mathsf{A}}(\sigma) = 0$ *if and only if*

$$\sigma = \sum_{i=1}^{n_{\mathsf{A}}} q_i |\varphi_i\rangle\langle\varphi_i| \otimes \sigma_{\mathsf{B}|i}, \tag{326}$$

where $\{|\varphi_i\rangle\}_{i=1}^{n_{\mathsf{A}}}$ *is an orthonormal basis of* $\mathcal{H}_{\mathsf{A}}$, $\sigma_{\mathsf{B}|i}$ *are some (arbitrary) states of* B *depending on the index* i, *and* $q_i \geq 0$ *are some probabilities*, $\sum_i q_i = 1$.

The non-negativity of $\delta_{\mathsf{A}}(\rho)$ means that one cannot gain more information on a bipartite system AB by performing a measurement on the subsystem A than the entropy of A, namely, $S(\rho_{\mathsf{AB}}) - \sum \eta_i S(\rho_{\mathsf{AB}|i}) \leq S(\rho_{\mathsf{A}})$ for any $\rho_{\mathsf{AB}} \in \mathcal{E}(\mathcal{H}_{\mathsf{AB}})$. The important point is that if ρ_{AB} is not of the form (326), then any measurement on A gives *less* information on AB than $S(\rho_{\mathsf{A}})$. Stated differently, one cannot retrieve all the information on A by a local measurement, because of the presence of quantum correlations between A and B.

In Ref. 120, the authors argue that the non-negativity of $\delta_{\mathsf{A}}^{\text{v.N.}}(\rho)$ is a direct consequence of (323) and the concavity of $S(\rho) - S(\rho_{\mathsf{A}})$ with respect to ρ. I do not see how such a claim could be justified and believe that the simplest proof of Proposition 10.B.1 is to rely on Theorem 6.B.1. Alternatively, the non-negativity of the discord can be justified with the help of the strong subadditivity of the von Neumann entropy (which is closely related to Theorem 6.B.1, see Sec. VI B), as shown in Ref. 106.

Proof. It remains to show the second affirmation. It is easy to convince oneself that the states (326) have a vanishing discord. In fact, one finds $I_{\mathsf{A}:\mathsf{B}}(\sigma) = S(\sigma_{\mathsf{B}}) - \sum_i q_i S(\sigma_{\mathsf{B}|i}) \leq J_{\mathsf{B}|\mathsf{A}}^{\text{v.N.}}(\sigma)$ (the inequality follows by noting that $\sigma_{\mathsf{B}|i}$ and q_i are the conditional state and outcome probability for a measurement on A in the basis $\{|\varphi_i\rangle\}$). Hence $\delta_{\mathsf{A}}(\sigma) = \delta_{\mathsf{A}}^{\text{v.N.}}(\sigma) = 0$ as a consequence of the non-negativity of δ_{A}. Reciprocally, let $\sigma \in \mathcal{E}(\mathcal{H}_{\mathsf{AB}})$ be such that $\delta_{\mathsf{A}}^{\text{v.N.}}(\sigma) = 0$. As we shall see below it is enough to work with the von Neumann discord, the result for δ_{A} will then follow from (321). According to (325) and Theorem 6.B.1, $\delta_{\mathsf{A}}^{\text{v.N.}}(\sigma) = 0$ if and only if there exists a von Neumann measurement $\mathcal{M}_{\mathsf{A}}$ on A with rank-one projectors $\pi_i^{\mathsf{A}} = |\varphi_i\rangle\langle\varphi_i|$ such that $\sigma = \mathcal{R}_{\mathsf{A}}\mathcal{M}_{\mathsf{A}} \otimes 1(\sigma)$, where $\mathcal{R}_{\mathsf{A}} = \mathcal{R}_{\mathcal{M}_{\mathsf{A}}\otimes 1,\sigma_0}$ is the transpose operation of $\mathcal{M}_{\mathsf{A}} \otimes 1$ for the state $\sigma_0 = \sigma_{\mathsf{A}} \otimes \sigma_{\mathsf{B}}$. Without loss of generality we may assume $\eta_i = \langle\varphi_i|\sigma_{\mathsf{A}}|\varphi_i\rangle > 0$ for all i. Thanks to (123) and to the identity $\mathcal{M}_{\mathsf{A}} \otimes 1(\sigma_0) = \sum_i \eta_i |\varphi_i\rangle\langle\varphi_i| \otimes \sigma_{\mathsf{B}}$, the transpose operation $\mathcal{R}_{\mathsf{A}}$ has Kraus operators $R_i = \eta_i^{-1/2}\sqrt{\sigma_{\mathsf{A}}}|\varphi_i\rangle\langle\varphi_i| \otimes 1$. We now argue that this implies that $\sigma = \widehat{\mathcal{M}}_{\mathsf{A}} \otimes 1(\sigma)$ with $\widehat{\mathcal{M}}_{\mathsf{A}}$ the von Neumann measurement with projectors $\widehat{\pi}_k^{\mathsf{A}}$ onto the subspaces $\text{span}\{|\varphi_i\rangle; i \in I_k\}$, where $\{I_1, \ldots, I_d\}$ is a partition of $\{1, \ldots, n_{\mathsf{A}}\}$. Actually, the condition $\sigma = \mathcal{R}_{\mathsf{A}}\mathcal{M}_{\mathsf{A}} \otimes 1(\sigma)$ reads

$$\langle\varphi_i|\sigma|\varphi_j\rangle = \sum_{l=1}^{n_{\mathsf{A}}} \eta_l^{-1}(\sqrt{\sigma_{\mathsf{A}}})_{il}(\sqrt{\sigma_{\mathsf{A}}})_{lj}\langle\varphi_l|\sigma|\varphi_l\rangle, \quad i, j = 1, \ldots, n_{\mathsf{A}} \tag{327}$$

with $(\sqrt{\sigma_{\mathsf{A}}})_{ij} = \langle\varphi_i|\sqrt{\sigma_{\mathsf{A}}}|\varphi_j\rangle \in \mathbb{R}$. Let us set $\sigma_{\mathsf{B}|i} = \eta_i^{-1}\langle\varphi_i|\sigma|\varphi_i\rangle$ and $\eta_{l|i} = |(\sqrt{\sigma_{\mathsf{A}}})_{il}|^2/\eta_i$. This defines, respectively, a state on $\mathcal{H}_{\mathsf{B}}$ and a probability distribution for any fixed i. With this notation,

(327) can be rewritten for $i = j$ as

$$\sigma_{\mathsf{B}|i} = \sum_{l=1}^{n_\mathsf{A}} \eta_{l|i}\sigma_{\mathsf{B}|l}, \quad i = 1, \ldots, n_\mathsf{A}. \tag{328}$$

Let $I_i = \{j; \sigma_{\mathsf{B}|j} = \sigma_{\mathsf{B}|i}\} \subset \{1, \ldots, n_\mathsf{A}\}$. Clearly, the sets I_i are either equal or disjoint. Hence, one can extract from them a partition $\{I_{i_1}, I_{i_2}, \ldots, I_{i_d}\}$ of $\{1, \ldots, n_\mathsf{A}\}$. We claim that (328) implies $\eta_{l|i} = 0$ for $l \notin I_i$. This is a consequence of the following lemma.

Lemma 10.B.2. *Let* $\mathbf{x} = (x_1, \ldots, x_d)$ *be a vector of* $\mathcal{X}^d$ *with distinct components* x_k*, where* $\mathcal{X}$ *is a real vector space, and* $\{\xi_{k|m}\}_{k=1}^d$ *be some probability distributions such that* $\xi_{k|m} = 0 \Leftrightarrow \xi_{m|k} = 0$ *and the components of* $\mathbf{x}$ *have convex decompositions*

$$x_m = \sum_{k=1}^d \xi_{k|m} x_k \quad \forall \ m = 1, \ldots, d. \tag{329}$$

Then $\xi_{k|m} = \delta_{km}$ *for any* $k, m = 1, \ldots, d$.

We postpone the proof of this result to the next paragraph. By rewriting (328) as

$$\sigma_{\mathsf{B}|i_m} = \sum_{k=1}^d \xi_{k|m}\sigma_{\mathsf{B}|i_k} \quad \text{with} \quad \xi_{k|m} = |I_{i_m}|^{-1} \sum_{(l,i)\in I_{i_k}\times I_{i_m}} \eta_{l|i}, \tag{330}$$

one concludes from Lemma 10.B.2 that $\xi_{k|m} = 0$ for $k \neq m$, i.e., $\eta_{l|i} = (\sqrt{\sigma_\mathsf{A}})_{il} = 0$ for any (i, l) such that $l \notin I_i$. One then obtains from (327)

$$\sigma = \sum_{i,j=1}^{n_\mathsf{A}} \sum_{l\in I_i\cap I_j} (\sqrt{\sigma_\mathsf{A}})_{il}(\sqrt{\sigma_\mathsf{A}})_{lj}|\varphi_i\rangle\langle\varphi_j| \otimes \sigma_{\mathsf{B}|l} = \sum_{k=1}^d \sum_{i,j\in I_{i_k}} (\sigma_\mathsf{A})_{ij}|\varphi_i\rangle\langle\varphi_j| \otimes \sigma_{\mathsf{B}|i_k}. \tag{331}$$

This gives

$$\sigma = \sum_{k=1}^d \widehat{\pi}_k^\mathsf{A}\sigma_\mathsf{A}\widehat{\pi}_k^\mathsf{A} \otimes \sigma_{\mathsf{B}|i_k}, \quad \widehat{\pi}_k^\mathsf{A} = \sum_{i\in I_{i_k}} |\varphi_i\rangle\langle\varphi_i|. \tag{332}$$

The last expression is of the form (326) (note that the vectors $|\varphi_i\rangle$ in the latter formula are the eigenvectors of $\widehat{\pi}_k^\mathsf{A}\sigma_\mathsf{A}\widehat{\pi}_k^\mathsf{A}$, so that they are in general linear combinations of the vectors $|\varphi_i\rangle$ defined above). To get the result for the discord δ_A we take advantage of (321). From the foregoing result, $\delta_\mathsf{A}(\sigma) = 0$ is equivalent to $\sigma \otimes |\epsilon_0\rangle\langle\epsilon_0|$ being of the form (326) for some orthonormal basis $\{|\varphi_i^{\mathsf{AE}}\rangle\}$ of $\mathcal{H}_{\mathsf{AE}}$. This straightforwardly implies $|\varphi_i^{\mathsf{AE}}\rangle = |\varphi_i\rangle|\epsilon_0\rangle$ with $\{|\varphi_i\rangle\}$ an orthonormal basis of $\mathcal{H}_\mathsf{A}$. □

Proof of Lemma 10.B.2. One proceeds by induction on d. The result is trivial for $d = 2$. Let us assume that it holds true for $d \geq 2$ and that one can find a vector $\mathbf{x} \in \mathcal{X}^{d+1}$ and some probabilities $\{\xi_{k|m}\}_{k=1}^{d+1}$ like in the lemma such that $\xi_{k_0|k_0} < 1$ for some $k_0 \in \{1, \ldots, d+1\}$. We are going to show that this leads to a contradiction. By plugging $x_{k_0} = (1 - \xi_{k_0|k_0})^{-1}\sum_{k\neq k_0} \xi_{k|k_0}x_k$ into the p other convex decompositions, one gets $x_m = \sum_{k\neq k_0} \zeta_{k|m}x_k$ for $k \neq k_0$, with $\zeta_{k|m} = \xi_{k|m} + (1 - \xi_{k_0|k_0})^{-1}\xi_{k_0|m}\xi_{k|k_0}$. As $\{\zeta_{k|m}\}_{k\neq k_0}$ is a probability distribution satisfying $\zeta_{k|m} = 0 \Leftrightarrow \zeta_{m|k} = 0$, by the induction hypothesis one has $\zeta_{k|m} = \delta_{km}$ for any $k, m \in \{1, \ldots, d+1\}\backslash\{k_0\}$. Now $\xi_{m_0|k_0} > 0$ for some index $m_0 \neq k_0$ (because $\xi_{k_0|k_0} < 1$). One deduces from the above identities and the hypothesis on $\xi_{k|m}$ that the only non-vanishing probabilities are $\xi_{k_0|m_0}$, $\xi_{m_0|k_0}$, and $\xi_{k|k}$, $k = 1, \ldots, p+1$. The problem then reduces to the case $p = 2$. Thus $\xi_{k_0|k_0} = \xi_{m_0|m_0} = 1$, in contradiction with our assumption. □

Definition 10.B.3. *The zero-discord states of the form (326) are called the* A*-classical states. We denote by* $\mathcal{C}_\mathsf{A}$ *the set of all* A*-classical states. Similarly,* $\mathcal{C}_\mathsf{B}$ *is the set of all* B*-classical states, namely*

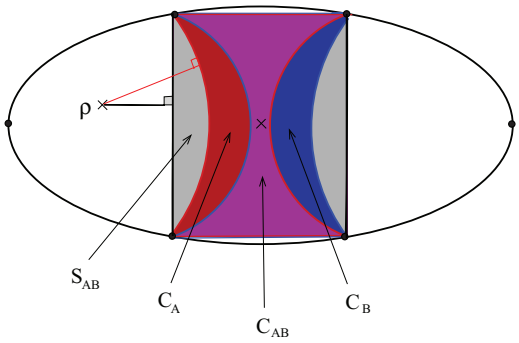

FIG. 3. Schematic view of the set of quantum states $\mathcal{E}_{\mathsf{AB}} = \mathcal{E}(\mathcal{H}_{\mathsf{AB}})$ of a bipartite system AB. The subset $\mathcal{C}_{\mathsf{AB}}$ of classical states (in magenta) is the intersection of the subsets $\mathcal{C}_{\mathsf{A}}$ and $\mathcal{C}_{\mathsf{B}}$ of A- and B-classical states (in red and blue). The convex hull of $\mathcal{C}_{\mathsf{A}}$ (or $\mathcal{C}_{\mathsf{B}}$) is the subset $\mathcal{S}_{\mathsf{AB}}$ of separable states (gray square). All these subsets intersect the border of $\mathcal{E}_{\mathsf{AB}}$ (pure states of AB) at the pure product states, represented by the four vertices of the square. The maximally mixed state $\rho_{\mathsf{AB}} = 1/(n_{\mathsf{A}} n_{\mathsf{B}})$ lies at the center (cross). The two points at the left and right extremities of the ellipse represent the maximally entangled pure states, which are the most distant states from $\mathcal{S}_{\mathsf{AB}}$ (and also from $\mathcal{C}_{\mathsf{A}}$, $\mathcal{C}_{\mathsf{B}}$, and $\mathcal{C}_{\mathsf{AB}}$). The closest distances of a state ρ to $\mathcal{S}_{\mathsf{AB}}$ (black line) and of ρ to $\mathcal{C}_{\mathsf{A}}$ (red line) define the square roots of the geometric measure of entanglement $E_{\mathrm{Bu}}(\rho)$ and of the geometric discord $D_{\mathsf{A}}(\rho)$, respectively. Note that this picture is for illustrative purposes and does not reflect all geometrical aspects (in particular, $\mathcal{C}_{\mathsf{A}}$, $\mathcal{C}_{\mathsf{B}}$, and $\mathcal{C}_{\mathsf{AB}}$ typically have a lower dimensionality than $\mathcal{E}_{\mathsf{AB}}$ and $\mathcal{S}_{\mathsf{AB}}$).

the states with vanishing B*-discord. A classical state is a state which is both* A- *and* B*-classical. We write* $\mathcal{C}_{\mathsf{AB}} = \mathcal{C}_{\mathsf{A}} \cap \mathcal{C}_{\mathsf{B}}$.

Our terminology can be justified by noting that if AB is in a state of the form (326) then the subsystem A is in one of the orthogonal states $|\varphi_i\rangle$ with probability q_i, whence A behaves as a classical system being in state i with probability q_i. Alternatively, a state σ is A-classical if and only if there exists a von Neumann measurement on A with rank-one projectors $\pi_i^{\mathsf{A}} = |\varphi_i\rangle\langle\varphi_i|$ which does not perturb it in the absence of readout, i.e., $\sigma = \mathcal{M}_{\{\pi_i^{\mathsf{A}}\}} \otimes 1(\sigma)$. The unfortunate name "classical-quantum states" has become popular in the literature to refer to the A-classical states, the B-classical states being called "quantum-classical." Using the spectral decompositions of the $\sigma_{\mathsf{B}|i}$'s, any A-classical state $\sigma_{\mathsf{A-cl}} \in \mathcal{C}_{\mathsf{A}}$ can be decomposed as

$$\sigma_{\mathsf{A-cl}} = \sum_{i=1}^{n_{\mathsf{A}}} \sum_{j=1}^{n_{\mathsf{B}}} q_{ij} |\varphi_i\rangle\langle\varphi_i| \otimes |\chi_{j|i}\rangle\langle\chi_{j|i}| \,, \tag{333}$$

where $q_{ij} \geq 0$, $\sum_{i,j} q_{ij} = 1$ and, for any i, $\{|\chi_{j|i}\rangle\}_{j=1}^{n_{\mathsf{B}}}$ is an orthonormal basis of $\mathcal{H}_{\mathsf{B}}$ (note that the $|\chi_{j|i}\rangle$ need not be orthogonal for distinct i's). A classical state $\sigma_{\mathrm{clas}} \in \mathcal{C}_{\mathsf{A}} \cap \mathcal{C}_{\mathsf{B}}$ possesses an eigenbasis $\{|\varphi_i\rangle \otimes |\chi_j\rangle\}_{i=1,\,j=1}^{n_{\mathsf{A}},n_{\mathsf{B}}}$ of product vectors. It is fully classical, in the sense that any quantum system in this state can be "simulated" by a classical apparatus being in the state (i, j) with probability q_{ij}.

Let us point out that $\mathcal{C}_{\mathsf{A}}$, $\mathcal{C}_{\mathsf{B}}$, and $\mathcal{C}_{\mathsf{AB}}$ are not convex. Their convex hull is the set $\mathcal{S}_{\mathsf{AB}}$ of separable states. It is also important to realize that for pure states, A-classical, B-classical, classical, and separable states all coincide. Actually, according to (333) the pure A-classical (and, similarly, the pure B-classical) states are product states. In contrast, one can find mixed separable states which are not A-classical. An example for two qubits is

$$\rho = \frac{1}{4}\big(|+\rangle\langle+| \otimes |0\rangle\langle 0| + |-\rangle\langle-| \otimes |1\rangle\langle 1| + |0\rangle\langle 0| \otimes |-\rangle\langle-| + |1\rangle\langle 1| \otimes |+\rangle\langle+|\big) \tag{334}$$

with $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. It is clear that $\rho \in \mathcal{S}_{\mathsf{AB}}$, but ρ is neither A-classical nor B-classical. A schematic picture of the sets $\mathcal{S}_{\mathsf{AB}}$, $\mathcal{C}_{\mathsf{A}}$, $\mathcal{C}_{\mathsf{B}}$, and $\mathcal{C}_{\mathsf{AB}}$ for a general bipartite system AB is displayed in Fig. 3.

C. Properties of the quantum discord

1. Invariance and monotonicity properties

Unlike entanglement measures, the quantum discord is *not* monotonous with respect to LOCCs. In particular, local operations on the measured subsystem A can create discord. For instance, consider the classical state

$$\sigma = \frac{1}{2}\Big(|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |1\rangle\langle 1| \otimes |1\rangle\langle 1|\Big) \tag{335}$$

of two qubits. One can transform this state by a local operation $\mathcal{M}_{\mathsf{A}}$ on A into

$$\rho = \mathcal{M}_{\mathsf{A}} \otimes 1(\sigma) = \frac{1}{2}\Big(|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |+\rangle\langle +| \otimes |1\rangle\langle 1|\Big), \tag{336}$$

where $\mathcal{M}_{\mathsf{A}}$ has Kraus operators $A_0 = |0\rangle\langle 0|$ and $A_1 = |+\rangle\langle 1|$. The final state ρ has less total correlations than σ, its mutual information $I_{\mathsf{A}:\mathsf{B}}(\rho) = -p\ln p - (1-p)\ln(1-p)$ being smaller than $I_{\mathsf{A}:\mathsf{B}}(\sigma) = \ln 2$ (here $p = 1/2 + \sqrt{2}/4$). However, it has a positive discord $\delta_{\mathsf{A}}(\rho) > \delta_{\mathsf{A}}(\sigma) = 0$. This means that the loss of classical correlations $J_{\mathsf{B}|\mathsf{A}}(\sigma) - J_{\mathsf{B}|\mathsf{A}}(\rho)$ is larger than the loss of total correlations $I_{\mathsf{A}:\mathsf{B}}(\sigma) - I_{\mathsf{A}:\mathsf{B}}(\rho)$.

In contrast, as far as local operations on B are concerned everything goes as expected, as shown by the following result.

Proposition 10.C.1. *The quantum discord δ_{A} and classical correlations $J_{\mathsf{B}|\mathsf{A}}(\rho)$ are invariant with respect to unitary conjugations $\mathcal{U}_{\mathsf{A}} : \rho_{\mathsf{A}} \mapsto U_{\mathsf{A}}\rho_{\mathsf{A}}U_{\mathsf{A}}^*$ on A and monotonous with respect to quantum operations $\mathcal{M}_{\mathsf{B}}$ on B, namely,*

$$\begin{aligned} \delta_{\mathsf{A}}(\mathcal{U}_{\mathsf{A}} \otimes 1(\rho)) &= \delta_{\mathsf{A}}(\rho) \ , \ \delta_{\mathsf{A}}(1 \otimes \mathcal{M}_{\mathsf{B}}(\rho)) \le \delta_{\mathsf{A}}(\rho) \\ J_{\mathsf{B}|\mathsf{A}}(\mathcal{U}_{\mathsf{A}} \otimes 1(\rho)) &= J_{\mathsf{B}|\mathsf{A}}(\rho) \ , \ J_{\mathsf{B}|\mathsf{A}}(1 \otimes \mathcal{M}_{\mathsf{B}}(\rho)) \le J_{\mathsf{B}|\mathsf{A}}(\rho) \end{aligned} \tag{337}$$

and similarly for $\delta_{\mathsf{A}}^{\text{v.N.}}$ and $J_{\mathsf{B}|\mathsf{A}}^{\text{v.N.}}$.

Proof. The unitary invariance is trivial. The monotonicity of $J_{\mathsf{B}|\mathsf{A}}(\rho)$ with respect to operations on B comes from the monotonicity of the relative entropy and the formula

$$J_{\mathsf{B}|\mathsf{A}}(\rho) = \max_{\{M_i^{\mathsf{A}}\}}\Big\{\sum_i \eta_i S(\rho_{\mathsf{B}|i}||\rho_{\mathsf{B}})\Big\}, \tag{338}$$

which is a consequence of the definition (318) and of $\rho_{\mathsf{B}} = \sum_i \eta_i \rho_{\mathsf{B}|i}$. A simple justification of the monotonicity of δ_{A} with respect to operations on B uses the following reasoning.[131] Let us consider a generalized measurement $\{M_i^{\mathsf{A}}\}$ on A with associated quantum operation $\mathcal{M}_{\mathsf{A}}$. By invoking the Stinespring theorem, one can represent $\mathcal{M}_{\mathsf{A}}$ as $\mathcal{M}_{\mathsf{A}} \otimes 1(\rho) = \text{tr}_{\mathsf{E}}(\sigma_{\mathsf{ABE}})$ with $\sigma_{\mathsf{ABE}} = U_{\mathsf{AE}}\rho \otimes |\epsilon_0\rangle\langle\epsilon_0|U_{\mathsf{AE}}^*$ pertaining to an enlarged space $\mathcal{H}_{\mathsf{ABE}}$ and U_{AE} a unitary on $\mathcal{H}_{\mathsf{AE}}$. Thanks to the additivity and unitary invariance of the von Neumann entropy and to the relation $\text{tr}_{\mathsf{AE}}(\sigma_{\mathsf{ABE}}) = \rho_{\mathsf{B}}$, one finds

$$I_{\mathsf{A}:\mathsf{B}}(\rho) = I_{\mathsf{AE}:\mathsf{B}}(\sigma_{\mathsf{ABE}}), \quad I_{\mathsf{A}:\mathsf{B}}(\mathcal{M}_{\mathsf{A}} \otimes 1(\rho)) = I_{\mathsf{A}:\mathsf{B}}(\sigma_{\mathsf{AB}}). \tag{339}$$

Plugging these expressions into (325) gives the following expression of $\delta_{\mathsf{A}}(\rho)$ in terms of the conditional mutual information

$$\delta_{\mathsf{A}}(\rho) = \min_{\{M_i^{\mathsf{A}}\}}\big\{I_{\mathsf{AE}:\mathsf{B}}(\sigma_{\mathsf{ABE}}) - I_{\mathsf{A}:\mathsf{B}}(\sigma_{\mathsf{AB}})\big\} = \min_{\{M_i^{\mathsf{A}}\}}\big\{I_{\mathsf{AB}:\mathsf{E}}(\sigma_{\mathsf{ABE}}) - I_{\mathsf{A}:\mathsf{E}}(\sigma_{\mathsf{AE}})\big\}. \tag{340}$$

The monotonicity of δ_{A} then follows from the monotonicity of the mutual information with respect to local operations (Sec. X A). □

2. States with the highest discord

As stated at the beginning of this section, the quantum discord $\delta_{\mathsf{A}}(\rho)$ is an indicator of the degree of quantumness of ρ. It is thus natural to ask whether the "most quantum" states having the

highest discord are the maximally entangled states characterized in Proposition 9.E.1. The answer is affirmative when $n_{\mathsf{A}} \leq n_{\mathsf{B}}$.

Proposition 10.C.2. *For any state ρ of the bipartite system* AB, *one has*

$$\delta_{\mathsf{A}}(\rho) \leq \delta_{\mathsf{A}}^{\text{v.N.}}(\rho) \leq S(\rho_{\mathsf{A}}) \leq \ln n_{\mathsf{A}}. \tag{341}$$

If $n_{\mathsf{A}} \leq n_{\mathsf{B}}$ then the maximal value of $\delta_{\mathsf{A}}(\rho)$ over all states $\rho \in \mathcal{E}(\mathcal{H}_{\mathsf{AB}})$ is equal to $\ln n_{\mathsf{A}}$ and $\delta_{\mathsf{A}}(\rho) = \ln n_{\mathsf{A}}$ if and only if ρ has highest entanglement of formation. Thus, the states ρ_{ent} with highest discord are the maximally entangled states given by Proposition 9.E.1, which satisfy

$$\delta_{\mathsf{A}}(\rho_{\text{ent}}) = \delta_{\mathsf{A}}^{\text{v.N.}}(\rho_{\text{ent}}) = E_{\text{EoF}}(\rho_{\text{ent}}) = \ln n_{\mathsf{A}}. \tag{342}$$

The statements in this proposition are probably well known in the literature, although I have not found an explicit reference.

Proof. Let $\rho = \sum_k p_k |k\rangle\langle k|$ be the spectral decomposition of ρ and $r = \text{rank}(\rho)$. As mentioned earlier, the von Neumann measurement minimizing the conditional entropy $\sum_i \eta_i S(\rho_{\mathsf{B}|i})$ consists of rank-one projectors $\pi_i^{\mathsf{A}} = |\varphi_i\rangle\langle\varphi_i|$. The conditional states (317) take the form

$$\rho_{\mathsf{B}|i} = \sum_{k=1}^{r} p_{k|i} |\phi_{ki}\rangle\langle\phi_{ki}| \quad \text{with} \quad p_{k|i} = \frac{p_k \eta_{i|k}}{\eta_i} \quad \text{and} \quad \sqrt{\eta_{i|k}}|\phi_{ki}\rangle = \langle\varphi_i|k\rangle \in \mathcal{H}_{\mathsf{B}}, \tag{343}$$

where $\eta_{i|k} = \|\langle\varphi_i|k\rangle\|^2$ is the probability of outcome i given the state $|k\rangle$ and $p_{k|i}$ is the "a posteriori" probability that the state is $|k\rangle$ given the measurement outcome i (Bayes rules). Since $\{|\phi_{ki}\rangle, p_{k|i}\}$ is a pure state decomposition of $\rho_{\mathsf{B}|i}$, the formula (118) yields

$$\sum_i \eta_i S(\rho_{\mathsf{B}|i}) \leq \sum_i \eta_i H(\{p_{k|i}\}). \tag{344}$$

The right-hand side is the classical conditional entropy given the measurement outcomes, see (110). By the non-negativity of the classical mutual information, it is bounded from above by the Shannon entropy $H(\{p_k\}) = -\sum_k p_k \ln p_k = S(\rho)$. Hence $\delta_{\mathsf{A}}^{\text{v.N.}}(\rho) \leq S(\rho_{\mathsf{A}})$ by (319). But $S(\rho_{\mathsf{A}}) \leq \ln n_{\mathsf{A}}$, thus we have proven (341).

Let us assume that $\delta_{\mathsf{A}}(\rho) = S(\rho_{\mathsf{A}})$. We know from Sec. VI A that a necessary and sufficient condition for (344) to be an equality is that $\{|\phi_{ki}\rangle, p_{k|i}\}$ be a spectral decomposition of $\rho_{\mathsf{B}|i}$, for any i. Setting $D_{kl} = \text{tr}_{\mathsf{B}}(|k\rangle\langle l|)$ as in the proof of Proposition 9.E.1, one gets $\sqrt{\eta_{i|l}\eta_{i|k}}\langle\phi_{li}|\phi_{ki}\rangle = \langle\varphi_i|D_{kl}|\varphi_i\rangle = 0$ if $k \neq l$ and $p_{k|i}p_{l|i} > 0$. Since $\delta_{\mathsf{A}}(\rho) = S(\rho_{\mathsf{A}})$, (344) holds with equality for any orthonormal basis $\{|\varphi_i\rangle\}$ and thus $D_{kl} = 0$ for such k and l. In addition, the conditional entropy in the right-hand side of (344) is equal to its upper bound $H(\{p_k\}) = S(\rho)$. This can happen only if $p_{k|i} = p_k$, i.e., $\eta_{i|k} = \langle\varphi_i|D_{kk}|\varphi_i\rangle = \eta_i$, for all i and k (indeed, the mutual information vanishes for independent random variables only). Hence $\delta_{\mathsf{A}}(\rho) = S(\rho_{\mathsf{A}})$ if and only if D_{kk} is independent of k and $D_{kl} = 0$ when $k \neq l$ and $p_k p_l > 0$. Suppose now that $\delta_{\mathsf{A}}(\rho) = \ln n_{\mathsf{A}}$. Then $\delta_{\mathsf{A}}(\rho) = S(\rho_{\mathsf{A}}) = \ln n_{\mathsf{A}}$ and the foregoing conditions on D_{kl} are fulfilled. In addition, $\rho_{\mathsf{A}} = \sum p_k D_{kk} = 1/n_{\mathsf{A}}$, whence $D_{kk} = 1/n_{\mathsf{A}}$ for all k with $p_k > 0$. One concludes that the eigenvectors $|k\rangle$ are as in Proposition 9.E.1 by following the same steps as in the proof of this proposition. □

Note that when $n_{\mathsf{A}} > n_{\mathsf{B}}$, $\delta_{\mathsf{A}}(\rho)$ is strictly smaller than $\ln n_{\mathsf{A}}$ for any $\rho \in \mathcal{E}(\mathcal{H}_{\mathsf{AB}})$. In fact, in that case $\text{rank}(D_{kk}) \leq n_{\mathsf{B}} < n_{\mathsf{A}}$ by the Schmidt decomposition (9), and the necessary condition $D_{kk} = 1/n_{\mathsf{A}}$ for having $\delta_{\mathsf{A}}(\rho) = \ln n_{\mathsf{A}}$ cannot be fulfilled.

3. Monotonicity when disregarding a part of the measured subsystem

We close this review of the properties of the discord by a simple remark concerning tripartite systems ABC. If such a system is in the state ρ_{ABC}, it is easy to show that

$$J_{\mathsf{B}|\mathsf{AC}}(\rho_{\mathsf{ABC}}) \geq J_{\mathsf{B}|\mathsf{A}}(\rho_{\mathsf{AB}}) \,. \tag{345}$$

This means that if B is coupled to both A and C, the gain of information on B from joint measurements on A and C is larger than the gain of information by measuring A only and ignoring C, as this sounds reasonable. A similar bound exists for the total correlations: by (316) and the monotonicity of the relative entropy (or, equivalently, the strong subadditivity of S),

$$I_{\mathsf{AC:B}}(\rho_{\mathsf{ABC}}) \geq I_{\mathsf{A:B}}(\rho_{\mathsf{AB}})\,. \tag{346}$$

Remark 10.C.3. The Holevo bound (113) can be derived by using the monotonicity of the quantum mutual information under operations acting on one subsystem (Sec. XA) and the property (346).

Sketch of the proof.[117] Given an ensemble $\{\rho_i, \eta_i\}_{i=1}^m$ of states on $\mathcal{H}_\mathsf{A}$ and a family $\{A_j\}_{j=1}^p$ of Kraus operators describing the measurement on A, consider the state $\rho_{\mathsf{ARP}} = \sum_i \eta_i \rho_i \otimes |\nu_i\rangle\langle\nu_i| \otimes |0\rangle\langle 0|$ on $\mathcal{H}_{\mathsf{ARP}}$, where R and P are auxiliary systems with orthonormal bases $\{|\nu_i\rangle\}_{i=1}^m$ and $\{|j\rangle\}_{j=0}^{p-1}$. These systems represent a register of the state preparation and a pointer for the measurement, respectively. Let $\mathcal{M}_{\mathsf{AP}}$ be the quantum operation on $\mathcal{B}(\mathcal{H}_{\mathsf{AP}})$ with Kraus operators $A_j \otimes U_j$, U_j being the unitary on $\mathcal{H}_\mathsf{P}$ defined by $U_j|l\rangle = |l + j\rangle$ for any $l = 0, \ldots, p-1$ (the addition is modulo p). It is an easy exercise to show that the Holevo bound (113) is equivalent to $I_{\mathsf{R:P}}([\mathcal{M}_{\mathsf{AP}} \otimes 1(\rho_{\mathsf{ARP}})]_{\mathsf{RP}}) \leq I_{\mathsf{AP:R}}(\rho_{\mathsf{ARP}})$.

D. Monogamy relation

Consider a tripartite system ABC in a pure state $|\Psi_{\mathsf{ABC}}\rangle$. If B and C are entangled, is there a limit on the amount of entanglement B can have with A? In other words, can entanglement be freely shared between different subsystems? A negative answer to the last question has been highlighted in Ref. 44, where it is shown that when A, B, and C are qubits, the sum $C(\rho_{\mathsf{AB}})^2 + C(\rho_{\mathsf{BC}})^2$ of the square concurrences is smaller or equal to $4\det(\rho_\mathsf{B})$. It is instructive to consider the limiting case where B and C are maximally entangled. Then, if one also assumes that $rn_\mathsf{B} \leq n_\mathsf{C} < (r+1)n_\mathsf{B}$ with $1 \leq r \leq n_\mathsf{A}$, A and B cannot be entangled and even have vanishing discords $\delta_\mathsf{A}(\rho_{\mathsf{AB}}) = \delta_\mathsf{B}(\rho_{\mathsf{AB}}) = 0$. In fact, the state of BC being maximally entangled, one has $\rho_{\mathsf{BC}} = \sum_k p_k |k\rangle\langle k|$ for some orthogonal maximally entangled states $|k\rangle$ satisfying $D_{kl} = \mathrm{tr}_\mathsf{C}(|k\rangle\langle l|) = n_\mathsf{B}^{-1}\delta_{kl}$ (see Proposition 9.E.1). Hence, the pure state of ABC is $|\Psi_{\mathsf{ABC}}\rangle = \sum_k \sqrt{p_k}|\alpha_k\rangle|k\rangle$ with $\{|\alpha_k\rangle\}$ an orthonormal family of $\mathcal{H}_\mathsf{A}$ (Sec. II C). Consequently, $\rho_{\mathsf{AB}} = (\sum_k p_k |\alpha_k\rangle\langle\alpha_k|) \otimes (1/n_\mathsf{B})$ is a product state and thus a classical state.

The proposition below exhibits an astonishing bound, called the monogamy relation, between the entanglement of formation of ρ_{BC} and the POVM-discord of ρ_{AB} measuring A.

Proposition 10.D.1. (Koashi and Winter[94]) *Let* ABC *be a tripartite system in the state* ρ_{ABC}. *Let* $\rho_{\mathsf{AB}} = \mathrm{tr}_\mathsf{C}(\rho_{\mathsf{ABC}})$ *and* $\rho_{\mathsf{BC}} = \mathrm{tr}_\mathsf{A}(\rho_{\mathsf{ABC}})$ *denote the reduced states of the bipartite systems* AB *and* BC, *respectively. Then*

$$E_{\mathrm{EoF}}(\rho_{\mathsf{BC}}) \leq S(\rho_\mathsf{B}) - J_{\mathsf{B|A}}(\rho_{\mathsf{AB}}) = \delta_\mathsf{A}(\rho_{\mathsf{AB}}) + S(\rho_{\mathsf{AB}}) - S(\rho_\mathsf{A}). \tag{347}$$

Moreover, the inequality is an equality if ρ_{ABC} *is a pure state*.

The inequality (347) tells us that the more classically correlated are A and B, the less B can be entangled to a third system C. If $n_\mathsf{B} \leq n_\mathsf{C}$ and B and C are maximally entangled, i.e., $E_{\mathrm{EoF}}(\rho_{\mathsf{BC}}) = \ln(n_\mathsf{B})$, then this inequality entails $J_{\mathsf{B|A}}(\rho_{\mathsf{AB}}) = 0$ (since $S(\rho_\mathsf{B}) \leq \ln(n_\mathsf{B})$). Thus A and B are not correlated classically, in agreement with the above statement that ρ_{AB} is a product state.

The entropy difference $S_{\mathsf{B|A}}(\rho_{\mathsf{AB}}) = S(\rho_{\mathsf{AB}}) - S(\rho_\mathsf{A})$ in the right-hand side of (347) is called the *conditional von Neumann entropy*. It is known that $S_{\mathsf{B|A}}(\rho_{\mathsf{AB}}) \geq 0$ if ρ_{AB} is separable.[79,36] Thanks to the subadditivity of S one has $-S(\rho_\mathsf{B}) \leq S_{\mathsf{B|A}}(\rho_{\mathsf{AB}}) \leq S(\rho_\mathsf{B})$ (the first inequality is obtained by considering a purification of ρ_{AB} on $\mathcal{H}_{\mathsf{ABC}}$ and using the subadditivity for ρ_{BC} together with the identities $S(\rho_{\mathsf{BC}}) = S(\rho_\mathsf{A})$ and $S(\rho_\mathsf{C}) = S(\rho_{\mathsf{AB}})$). The quantity $-S_{\mathsf{B|A}}(\rho_{\mathsf{AB}})$ is the coherent information introduced by Schumacher and Nielsen in the context of the quantum channel capacity.[142]

Two consequences of the claim that (347) is an equality for tripartite systems ABC in pure states deserve further comments. First, one easily deduces from this claim and the identity (117) that[58]

$$E_{\mathrm{EoF}}(\rho_{\mathsf{AB}}) + E_{\mathrm{EoF}}(\rho_{\mathsf{BC}}) = \delta_{\mathsf{A}}(\rho_{\mathsf{AB}}) + \delta_{\mathsf{C}}(\rho_{\mathsf{BC}}). \tag{348}$$

Hence the sum of all entanglement of formations describing the bipartite entanglement shared by B is equal to the sum of the corresponding quantum discords with measurements on the other subsystems. Second, if B is a qubit and ρ_{AB} is of rank two, then ρ_{AB} admits a purification $|\Psi_{\mathsf{ABC}}\rangle$ on $\mathcal{H}_{\mathsf{AB}} \otimes \mathbb{C}^2$ (see (12)) and the entanglement of formation of the two-qubit state ρ_{BC} can be computed with the help of the Wootters formula (309). One may in this way determine $\delta_{\mathsf{A}}(\rho_{\mathsf{AB}})$ via (347).

Proof. We first assume that ABC is in a pure state $|\Psi_{\mathsf{ABC}}\rangle$. Let $\{M^{\mathrm{opt}}_{\mathsf{A},i}\}$ be an optimal measurement on A maximizing the gain of information on B, that is, such that $J_{\mathsf{B}|\mathsf{A}}(\rho_{\mathsf{AB}}) = S(\rho_{\mathsf{B}}) - \sum_i \eta_i^{\mathrm{opt}} S(\rho^{\mathrm{opt}}_{\mathsf{B}|i})$, where η_i^{opt} and $\rho^{\mathrm{opt}}_{\mathsf{B}|i}$ are the outcome probabilities and conditional states of B for this measurement. Without loss of generality one may assume that $M^{\mathrm{opt}}_{\mathsf{A},i} = |\widetilde{\mu}_i^{\mathrm{opt}}\rangle\langle\widetilde{\mu}_i^{\mathrm{opt}}|$ are of rank one (see the discussion after (318)). Since $\rho_{\mathsf{AB}} = \mathrm{tr}_{\mathsf{C}}(|\Psi_{\mathsf{ABC}}\rangle\langle\Psi_{\mathsf{ABC}}|)$, one has $\eta_i^{\mathrm{opt}} = \mathrm{tr}(\rho_{\mathsf{AB}} M^{\mathrm{opt}}_{\mathsf{A},i} \otimes 1) = \|\langle\widetilde{\mu}_i^{\mathrm{opt}}|\Psi_{\mathsf{ABC}}\rangle\|^2$. Moreover, the post-measurement conditional state of BC is the pure state

$$|\Psi_{\mathsf{BC}|i}\rangle = (\eta_i^{\mathrm{opt}})^{-\frac{1}{2}} \langle\widetilde{\mu}_i^{\mathrm{opt}}|\Psi_{\mathsf{ABC}}\rangle \tag{349}$$

and the conditional state of B is $\rho^{\mathrm{opt}}_{\mathsf{B}|i} = \mathrm{tr}_{\mathsf{C}}(|\Psi_{\mathsf{BC}|i}\rangle\langle\Psi_{\mathsf{BC}|i}|)$. The ensemble $\{|\Psi_{\mathsf{BC}|i}\rangle, \eta_i^{\mathrm{opt}}\}$ gives a pure state decomposition of ρ_{BC}. Actually, let us consider the post-measurement state of ABC in the absence of readout, $\rho'_{\mathsf{ABC}} = \mathcal{M}^{\mathrm{opt}}_{\mathsf{A}} \otimes 1(|\Psi_{\mathsf{ABC}}\rangle\langle\Psi_{\mathsf{ABC}}|)$. The measurement being performed on A, it does not change the state of BC, i.e.,

$$\rho_{\mathsf{BC}} = \rho'_{\mathsf{BC}} = \sum_i \eta_i^{\mathrm{opt}} |\Psi_{\mathsf{BC}|i}\rangle\langle\Psi_{\mathsf{BC}|i}|. \tag{350}$$

From the definition (303) of the entanglement of formation one has

$$E_{\mathrm{EoF}}(\rho_{\mathsf{BC}}) \le \sum_i \eta_i^{\mathrm{opt}} S(\rho^{\mathrm{opt}}_{\mathsf{B}|i}) = S(\rho_{\mathsf{B}}) - J_{\mathsf{B}|\mathsf{A}}(\rho_{\mathsf{AB}}). \tag{351}$$

Conversely, let $\{|\Psi_{\mathsf{BC},i}\rangle, \eta_i\}$ be a pure state decomposition of ρ_{BC} which achieves the minimum in the definition of the entanglement of formation. Let us show that there exists a generalized measurement $\{M_i^{\mathsf{A}}\}$ on A such that η_i is the probability of outcome i and $|\Psi_{\mathsf{BC},i}\rangle$ the corresponding conditional state of BC, i.e.,

$$\mathrm{tr}_{\mathsf{A}}(M_i^{\mathsf{A}} \otimes 1 |\Psi_{\mathsf{ABC}}\rangle\langle\Psi_{\mathsf{ABC}}|) = \eta_i |\Psi_{\mathsf{BC},i}\rangle\langle\Psi_{\mathsf{BC},i}| \ . \tag{352}$$

In fact, let us observe that $|\Psi'_{\mathsf{ABCE}}\rangle = \sum_i \sqrt{\eta_i}|\Psi_{\mathsf{BC},i}\rangle|\phi_i\rangle$ is a purification of ρ_{BC} on $\mathcal{H}_{\mathsf{ABCE}}$ for some ancilla E, where $\{|\phi_i\rangle\}$ is an orthonormal family of $\mathcal{H}_{\mathsf{AE}}$. Given an arbitrary state $|\epsilon_0\rangle \in \mathcal{H}_{\mathsf{E}}$, $|\Psi_{\mathsf{ABC}}\rangle|\epsilon_0\rangle$ is also a purification of ρ_{BC} on the same space. As a result, there is a unitary U_{AE} on $\mathcal{H}_{\mathsf{AE}}$ such that $|\Psi'_{\mathsf{ABCE}}\rangle = 1 \otimes U_{\mathsf{AE}}|\Psi_{\mathsf{ABC}}\rangle|\epsilon_0\rangle$ (see Sec. II C). Define

$$M_i^{\mathsf{A}} = \langle\epsilon_0|U_{\mathsf{AE}}^*|\phi_i\rangle\langle\phi_i|U_{\mathsf{AE}}|\epsilon_0\rangle \tag{353}$$

(note the analogy with (41)). Then (352) is satisfied. Let $\rho_{\mathsf{B}|i} = \mathrm{tr}_{\mathsf{C}}(|\Psi_{\mathsf{BC},i}\rangle\langle\Psi_{\mathsf{BC},i}|)$ be the post-measurement states of B, so that $E_{\mathrm{EoF}}(|\Psi_{\mathsf{BC},i}\rangle) = S(\rho_{\mathsf{B}|i})$. Since by assumption $E_{\mathrm{EoF}}(\rho_{\mathsf{BC}}) = \sum_i \eta_i E_{\mathrm{EoF}}(|\Psi_{\mathsf{BC},i}\rangle)$, one infers from the definition (318) of the classical correlations that

$$J_{\mathsf{B}|\mathsf{A}}(\rho_{\mathsf{AB}}) \ge S(\rho_{\mathsf{B}}) - \sum_i \eta_i S(\rho_{\mathsf{B}|i}) = S(\rho_{\mathsf{B}}) - E_{\mathrm{EoF}}(\rho_{\mathsf{BC}}). \tag{354}$$

Together with (351) this proves that

$$E_{\mathrm{EoF}}(\rho_{\mathsf{BC}}) = S(\rho_{\mathsf{B}}) - J_{\mathsf{B}|\mathsf{A}}(\rho_{\mathsf{AB}}). \tag{355}$$

Let us now turn to the case of a tripartite system ABC in a mixed state ρ_{ABC}. Consider a purification $|\Psi_{\mathsf{ABCE}}\rangle$ of ρ_{ABC} in the Hilbert space $\mathcal{H}_{\mathsf{ABC}} \otimes \mathcal{H}_{\mathsf{E}}$. Thanks to (345) one then has $J_{\mathsf{B}|\mathsf{A}}(\rho_{\mathsf{AB}}) \le J_{\mathsf{B}|\mathsf{AE}}(\rho_{\mathsf{ABE}})$. The inequality (347) then follows by applying (355) with $\mathsf{A} \to \mathsf{AE}$. □

XI. DISTANCE AND ENTROPIC MEASURES OF QUANTUM CORRELATIONS

In this section we study the measures of entanglement and quantum correlations based on the Bures distance and the relative entropies. First, we introduce in Sec. XI A the geometric measure of entanglement, defined as the minimal square distance between the state ρ and a separable state, as well as similar measures obtained by replacing the square distance by relative entropies. We define analogously in Sec. XI B the geometric discord as the minimal square distance between ρ and an A-classical state. We show there that this discord is related to a quantum state discrimination task and determine the closest A-classical states to ρ in terms of the corresponding optimal measurements.

A. Geometric and relative-entropy measures of entanglement

1. Definition and main properties

From a geometrical point of view, it is natural to quantify the amount of entanglement in a state ρ of a bipartite system AB by the distance $d(\rho, \mathcal{S}_{\mathsf{AB}})$ of ρ to the subset $\mathcal{S}_{\mathsf{AB}} \subset \mathcal{E}(\mathcal{H}_{\mathsf{AB}})$ of separable states (see Fig. 3). As it will become clear below, in order to obtain an entanglement monotone measure the distance d must be contractive. Choosing the Bures distance, it is easy to verify that

$$E_{\mathrm{Bu}}(\rho) = d_{\mathrm{B}}(\rho, \mathcal{S}_{\mathsf{AB}})^2 = \min_{\sigma_{\mathrm{sep}} \in \mathcal{S}_{\mathsf{AB}}} \left\{ d_B(\rho, \sigma_{\mathrm{sep}})^2 \right\} \tag{356}$$

satisfies all the axioms of an entanglement measure in Definition 9.C.1. Actually, the axiom (i) holds because d_{B} is a distance on $\mathcal{E}(\mathcal{H}_{\mathsf{AB}})$. The convexity property (ii) is a consequence of the convexity of $\mathcal{S}_{\mathsf{AB}}$ and the joint convexity of the square Bures distance (Corollary 7.B.3). (This justifies the square in our definition (356).) Finally, the monotonicity (iii) is shown in the following way. Let $\sigma_\rho \in \mathcal{S}_{\mathsf{AB}}$ be a closest separable state to ρ, i.e., $E_{\mathrm{Bu}}(\rho) = d_{\mathrm{B}}(\rho, \sigma_\rho)^2$. Let us recall from Sec. IX C that any LOCC is a separable quantum operation and can be written as $\mathcal{M}(\rho) = \sum_i A_i \otimes B_i \rho A_i^* \otimes B_i^*$. Furthermore, one has $\mathcal{M}(\mathcal{S}_{\mathsf{AB}}) \subset \mathcal{S}_{\mathsf{AB}}$. One can then use the contractivity of d_{B} to obtain

$$E_{\mathrm{Bu}}(\rho) \geq d_{\mathrm{B}}(\mathcal{M}(\rho), \mathcal{M}(\sigma_\rho))^2 \geq E_{\mathrm{Bu}}(\mathcal{M}(\rho)). \tag{357}$$

This shows that E_{Bu} is monotonous with respect to separable operations and, in particular, to LOCCs. The entanglement measure E_{Bu} has been first introduced by Vedral and Plenio.[160] Another measure was considered in Refs. 159 and 160 by replacing the square distance in (356) by the relative entropy $S(\rho||\sigma_{\mathrm{sep}})$. More generally, we can define

$$E_\alpha(\rho) = \min_{\sigma_{\mathrm{sep}} \in \mathcal{S}_{\mathsf{AB}}} \left\{ S_\alpha(\rho||\sigma_{\mathrm{sep}}) \right\}, \tag{358}$$

where S_α is the quantum relative Rényi entropy (Sec. VI C). For $1/2 \leq \alpha \leq 1$, this defines an entanglement measure by the same arguments as above, because S_α is jointly convex and contractive (see Theorem 6.C.1; the property (i) in this theorem ensures that $E_\alpha(\rho) \geq 0$ with equality if and only if $\rho \in \mathcal{S}_{\mathsf{AB}}$). One establishes the following result by invoking the fact that S_α is non-decreasing in α (Proposition 6.C.4) and by using (175) and the relation (149) between $S_{1/2}(\rho||\sigma)$ and the fidelity $F(\rho, \sigma)$.

Corollary 11.A.1. $\{E_\alpha\}_{1/2 \leq \alpha \leq 1}$ *constitutes a non-decreasing family of entanglement measures and*

$$E_{\frac{1}{2}}(\rho) = -2 \ln\left(1 - \frac{E_{\mathrm{Bu}}(\rho)}{2}\right) \leq E_\alpha(\rho), \quad \frac{1}{2} \leq \alpha \leq 1 . \tag{359}$$

The measure E_1 associated to the relative entropy (122) is less geometrical than E_{Bu} (it is not associated to a distance) but has the following interesting property.

Proposition 11.A.2. (Vedral and Plenio[160]) *The entanglement measure E_1 coincides with the entanglement of formation E_{EoF} for pure states, and for mixed states $\rho \in \mathcal{E}(\mathcal{H}_{\mathsf{AB}})$ it is bounded from above by E_{EoF},*

$$E_1(\rho) \leq E_{\mathrm{EoF}}(\rho). \tag{360}$$

Proof. We refer the reader to Ref. 160 for a detailed proof of the first statement. It is based on the observation that for a pure state with Schmidt decomposition $|\Psi\rangle = \sum_i \sqrt{\mu_i}|\alpha_i\rangle|\beta_i\rangle$, the minimum in (358) is achieved when $\sigma_{\rm sep}$ is the classical state

$$\sigma_* = \sum_{i=1}^{n} \mu_i |\alpha_i\rangle\langle\alpha_i| \otimes |\beta_i\rangle\langle\beta_i|. \tag{361}$$

Since $S(\rho_\Psi||\sigma_*) = -\langle\Psi|\ln\sigma_*|\Psi\rangle = -\sum_i \mu_i \ln\mu_i$, the equality $E_1(|\Psi\rangle) = E_{\rm EoF}(|\Psi\rangle)$ follows once one has proven that $S(\rho_\Psi||\sigma_{\rm sep}) \geq S(\rho_\Psi||\sigma_*)$ for all $\sigma_{\rm sep} \in \mathcal{S}_{\mathsf{AB}}$. This is done in Ref. 160 by showing that for any $\sigma_{\rm sep} \in \mathcal{S}_{\mathsf{AB}}$,

$$\frac{{\rm d}f_\Psi(t,\sigma_{\rm sep})}{{\rm d}t}\Big|_{t=0} = 1 - \int_0^\infty {\rm d}t\ {\rm tr}\big((\sigma_*+t)^{-1}\rho_\Psi(\sigma_*+t)^{-1}\sigma_{\rm sep}\big)\ \geq\ 0 \tag{362}$$

with $f_\Psi(t,\sigma) = S(\rho_\Psi||(1-t)\sigma_* + t\sigma)$. Indeed, assume that $S(\rho_\Psi||\sigma_{\rm sep}) < S(\rho_\Psi||\sigma_*)$ for some $\sigma_{\rm sep} \in \mathcal{S}_{\mathsf{AB}}$. By taking advantage of the right convexity of the relative entropy, one then finds for any $t \in (0,1]$

$$\frac{f_\Psi(t,\sigma_{\rm sep}) - f_\Psi(0,\sigma_{\rm sep})}{t} \leq -S(\rho_\Psi||\sigma_*) + S(\rho_\Psi||\sigma_{\rm sep}) < 0, \tag{363}$$

in contradiction with (362). Note that it suffices to prove the non-negativity in (362) for the pure product states $\sigma_{\rm sep} = |\phi\otimes\chi\rangle\langle\phi\otimes\chi|$, because of the linearity in $\sigma_{\rm sep}$ of the trace in the right-hand side.

The second statement in the proposition is a consequence of the first one and of the convexity of E_1. Actually, if $\{|\Psi_i\rangle, \eta_i\}$ is a pure state decomposition of ρ minimizing the average entanglement, then

$$E_{\rm EoF}(\rho) = \sum_i \eta_i E_{\rm EoF}(|\Psi_i\rangle) = \sum_i \eta_i E_1(|\Psi_i\rangle) \geq E_1\Big(\sum_i \eta_i |\Psi_i\rangle\langle\Psi_i|\Big) = E_1(\rho). \tag{364}$$

□

Note that the inequality (360) can be strict. Examples of two-qubit states ρ for which $E_1(\rho) < E_{\rm EoF}(\rho)$ are given in Ref. 159. Thanks to (359) and (360), one can place an upper bound on $E_{\rm Bu}(\rho)$ by a function of the entanglement of formation $E_{\rm EoF}$. Such a bound does not seem to be known in the literature, but it is not optimal for pure states as a consequence of the next proposition.

Remark 11.A.3. As shown in Ref. 160, E_1 fulfills the stronger monotonicity condition (iii″) of Sec. IX D 2.

2. Relation between the geometric measure of entanglement and convex roof constructions

Let $F(\rho,\mathcal{S}_{\mathsf{AB}})$ denote the maximal fidelity between ρ and a separable state,

$$F(\rho,\mathcal{S}_{\mathsf{AB}}) = \max_{\sigma_{\rm sep}\in\mathcal{S}_{\mathsf{AB}}} \big\{F(\rho,\sigma_{\rm sep})\big\}. \tag{365}$$

Proposition 11.A.4. (Streltsov, Kampermann, and Bruß[152]) *The geometric measure of entanglement is given for pure states by*

$$E_{\rm Bu}(|\Psi\rangle) = 2 - 2\sqrt{F(|\Psi\rangle,\mathcal{S}_{\mathsf{AB}})} = 2(1-\sqrt{\mu_{\max}}), \tag{366}$$

where $\mu_{\max} = \max\{\mu_i\}$ is the largest Schmidt coefficient of $|\Psi\rangle$. For mixed states, $F(\rho,\mathcal{S}_{\mathsf{AB}})$ is obtained via a maximization over the pure state decompositions of ρ,

$$F(\rho,\mathcal{S}_{\mathsf{AB}}) = \max_{\{|\Psi_i\rangle,\eta_i\}} \Big\{\sum_i \eta_i F(|\Psi_i\rangle,\mathcal{S}_{\mathsf{AB}})\Big\}. \tag{367}$$

The nice relation (367) is intimately related to Uhlmann's theorem (Sec. VII B) and to the convexity of $\mathcal{S}_{\mathsf{AB}}$. Note that the relative-entropy measure E_1 does not fulfill a similar property

(compare with Proposition 11.A.2). Even though E_{Bu} is not a convex roof, it is a simple function of another entanglement measure E_{G} defined via a convex-roof construction like in (303) and from its expression for pure states[144,164]

$$E_{\mathrm{G}}(|\Psi\rangle) = 1 - \max_{|\Phi\rangle\in\mathcal{S}_{\mathsf{AB}}} \left\{|\langle\Phi|\Psi\rangle|^2\right\}. \tag{368}$$

Actually, we will see that a pure state always admits a pure product state as closest separable state, hence the maximum in (368) coincides with $F(|\Psi\rangle, \mathcal{S}_{\mathsf{AB}})$ and $E_{\mathrm{G}}(\rho) = 1 - F(\rho, \mathcal{S}_{\mathsf{AB}})$ by the proposition above. According to (366), $E_{\mathrm{G}}(|\Psi\rangle) = 1 - \mu_{\max}$ is of the form (289) with $f_{\mathrm{G}}(\rho_{\mathsf{A}}) = 1 - \|\rho_{\mathsf{A}}\|$ satisfying all hypothesis of Proposition 9.C.2. Therefore, by a similar reasoning as in the proof of Proposition 9.D.3, E_{G} is an entanglement measure which fulfills the strong monotonicity property (iii″). In contrast, $E_{\mathrm{Bu}}(|\Psi\rangle) = f_{\mathrm{Bu}}([\rho_\Psi]_{\mathsf{A}}) = 2(1 - \sqrt{\|[\rho_\Psi]_{\mathsf{A}}\|})$ but f_{Bu} is not concave, whence Proposition 9.C.2 indicates that E_{Bu} does not fulfill (iii′). We should not be bothered too much about that, the two measures E_{Bu} and E_{G} being equivalent (that is, they define the same order of entanglement) and simply related to each other.

Proof. For a pure state $\rho_\Psi = |\Psi\rangle\langle\Psi|$, the fidelity reads $F(\rho_\Psi, \sigma_{\mathrm{sep}}) = \langle\Psi|\sigma_{\mathrm{sep}}|\Psi\rangle$. Writing the decomposition of separable states into pure product states, $\sigma_{\mathrm{sep}} = \sum_i \xi_i |\varphi_i \otimes \chi_i\rangle\langle\varphi_i \otimes \chi_i|$, we get

$$F(\rho_\Psi, \mathcal{S}_{\mathsf{AB}}) = \max_{\{|\varphi_i\rangle, |\chi_i\rangle, \xi_i\}} \left\{\sum_i \xi_i |\langle\varphi_i \otimes \chi_i|\Psi\rangle|^2\right\} = \max_{\|\varphi\|=\|\chi\|=1} \left\{|\langle\varphi\otimes\chi|\Psi\rangle|^2\right\}, \tag{369}$$

where we have used $\sum_i \xi_i = 1$. For any normalized vectors $|\varphi\rangle \in \mathcal{H}_A$ and $|\chi\rangle \in \mathcal{H}_B$, one derives from the Schmidt decomposition (9) and the Cauchy-Schwarz inequality that

$$\begin{aligned} |\langle\varphi\otimes\chi|\Psi\rangle| &\le \sum_{j=1}^n \sqrt{\mu_j} \left|\langle\varphi|\alpha_j\rangle\langle\chi|\beta_j\rangle\right| \le \sqrt{\mu_{\max}} \sum_{j=1}^n \left|\langle\varphi|\alpha_j\rangle\langle\chi|\beta_j\rangle\right| \\ &\le \sqrt{\mu_{\max}} \left(\sum_{j=1}^n |\langle\varphi|\alpha_j\rangle|^2\right)^{1/2} \left(\sum_{j=1}^n |\langle\chi|\beta_j\rangle|^2\right)^{1/2} \le \sqrt{\mu_{\max}}. \end{aligned} \tag{370}$$

All bounds are saturated for $|\varphi\rangle = |\alpha_{j_{\max}}\rangle$ and $|\chi\rangle = |\beta_{j_{\max}}\rangle$, where $j_{\max}$ is the index for which μ_j is maximum. Thus $F(\rho_\Psi, \mathcal{S}_{\mathsf{AB}}) = \mu_{j_{\max}} = \mu_{\max}$ and the formula (366) is proven. It is of interest to note that the pure product state $|\alpha_{j_{\max}}\rangle|\beta_{j_{\max}}\rangle$ is a closest separable state to $|\Psi\rangle$ (a characterization of all these closest separable states will be given in Proposition 11.B.2 below).

We now proceed to show (367). Consider a fixed separable state $\sigma_{\mathrm{sep}} = \sum_{i=1}^p \xi_i |\Phi_i\rangle\langle\Phi_i|$ with $|\Phi_i\rangle \in \mathcal{S}_{\mathsf{AB}}$ and $\xi_i \ge 0$. Without loss of generality one may assume $p = (n_{\mathsf{A}} n_{\mathsf{B}})^2 + 1$ (see the discussion after Definition 2.D.1). Let $\{|f_i\rangle\}_{i=1}^p$ be an orthonormal basis of an ancilla space $\mathcal{K}$ and $|\Phi\rangle = \sum_i \sqrt{\xi_i}|\Phi_i\rangle|f_i\rangle$ be a purification of σ_{sep} on $\mathcal{H}\otimes\mathcal{K}$. Thanks to Theorem 7.B.2, $F(\rho, \sigma_{\mathrm{sep}})$ is the maximum over all purifications $|\Psi\rangle$ of ρ on $\mathcal{H}\otimes\mathcal{K}$ of the transition probability $|\langle\Psi|\Phi\rangle|^2$. Writing $|\Psi\rangle$ in the form (15) and using the one-to-one correspondence between pure state decompositions and purifications (see Sec. II C), one can equivalently maximize over all pure state decompositions $\{|\Psi_i\rangle, \eta_i\}$ of ρ. Moreover, the maximization of $F(\rho, \sigma_{\mathrm{sep}})$ over the separable states σ_{sep} leads to a maximization over the pure state ensembles $\{|\Phi_i\rangle, \xi_i\}$ in $\mathcal{S}_{\mathsf{AB}}$. This yields

$$F(\rho, \mathcal{S}_{\mathsf{AB}}) = \max_{\{|\Phi_i\rangle, \xi_i\}} \max_{\{|\Psi_i\rangle, \eta_i\}} \left\{\left|\sum_{i=1}^p \sqrt{\eta_i \xi_i}\langle\Psi_i|\Phi_i\rangle\right|^2\right\}. \tag{371}$$

But, using once more the Cauchy-Schwarz inequality and $\sum_i \xi_i = 1$, one has

$$\max_{\{|\Phi_i\rangle, \xi_i\}} \left\{\left|\sum_{i=1}^p \sqrt{\eta_i \xi_i}\langle\Psi_i|\Phi_i\rangle\right|^2\right\} = \sum_{i=1}^p \eta_i \max_{|\Phi\rangle\in\mathcal{S}_{\mathsf{AB}}} \left\{|\langle\Psi_i|\Phi\rangle|^2\right\}. \tag{372}$$

It has been argued above that the maximal fidelity between $|\Psi_i\rangle$ and a separable state is attained for pure product states, thus $F(|\Psi_i\rangle, \mathcal{S}_{\mathsf{AB}}) = \max_{|\Phi\rangle\in\mathcal{S}_{\mathsf{AB}}} |\langle\Psi_i|\Phi\rangle|^2$. Substituting this expression into (372) and (371), we arrive at the required relation (367). □

According to (366), $E_{\mathrm{Bu}}(|\Psi\rangle) = 0$ if and only if $|\Psi\rangle$ is a product state, in agreement with the fact that by definition separable pure states are product states. Another consequence of (366) and of the bound $\mu_{\max} \geq 1/n$ (which follows from $\sum_i \mu_i = 1$) is $F(|\Psi\rangle, \mathcal{S}_{\mathsf{AB}}) \geq 1/n$, with $n = \min\{n_{\mathsf{A}}, n_{\mathsf{B}}\}$. Furthermore, $F(|\Psi\rangle, \mathcal{S}_{\mathsf{AB}}) = 1/n$ if and only if $|\Psi\rangle$ is maximally entangled (Sec. II D). One deduces from (367) that

$$E_{\mathrm{Bu}}(\rho) \leq 2 - \frac{2}{\sqrt{n}}. \tag{373}$$

By the same arguments as in the proof of Proposition 9.E.1, this bound is saturated if and only if ρ has maximal entanglement of formation $E_{\mathrm{EoF}}(\rho) = \ln n$. This means that E_{Bu} and E_{EoF} capture the same maximally entangled states.

3. Geometric measure of entanglement for two qubits

In the case of two qubits, a closed formula for $E_{\mathrm{Bu}}(\rho)$ can be obtained with the help of Proposition 11.A.4 and of Wootters's result on the concurrence (Sec. IX D 3). It reads[152]

$$E_{\mathrm{Bu}}(\rho) = 2 - \sqrt{2}\big(1 + \sqrt{1 - C(\rho)^2}\big)^{\frac{1}{2}} \tag{374}$$

with $C(\rho)$ given by (311). Actually, for pure states one finds by comparing $C(|\Psi\rangle) = 2\sqrt{\mu_0\mu_1}$ and (366) that $F(|\Psi\rangle, \mathcal{S}_{\mathsf{AB}}) = g(C(|\Psi\rangle))$ with $g(C) = (1 + \sqrt{1 - C^2})/2$. As g is decreasing and concave, (305) and (367) yield $F(\rho, \mathcal{S}_{\mathsf{AB}}) \leq g(C(\rho))$. But it is shown in Ref. 169 that there is an optimal pure state decomposition $\{|\Psi_i\rangle, \eta_i\}$ of ρ such that $C(\rho) = C(|\Psi_i\rangle)$ for any i. Thus

$$g\big(C(\rho)\big) \geq F(\rho, \mathcal{S}_{\mathsf{AB}}) \geq \sum_i \eta_i F(|\Psi_i\rangle, \mathcal{S}_{\mathsf{AB}}) = \sum_i \eta_i g\big(C(|\Psi_i\rangle)\big) = g\big(C(\rho)\big), \tag{375}$$

which justifies (374).

B. Geometric quantum discord

1. Discord-like measures of quantum correlations

In the same spirit as for the geometric measure of entanglement, one defines the geometric quantum discord as

$$D_{\mathsf{A}}(\rho) = d_{\mathrm{B}}(\rho, \mathcal{C}_{\mathsf{A}})^2 = 2(1 - \sqrt{F(\rho, \mathcal{C}_{\mathsf{A}})}), \quad F(\rho, \mathcal{C}_{\mathsf{A}}) = \max_{\sigma_{\mathsf{A-cl}}\in\mathcal{C}_{\mathsf{A}}} \big\{F(\rho, \sigma_{\mathsf{A-cl}})\big\}, \tag{376}$$

where $\mathcal{C}_{\mathsf{A}}$ is the (non-convex) set of A-classical states (see Definition 10.B.3). One can introduce similarly the relative-entropy discords

$$D_{\mathsf{A}}^{(\alpha)}(\rho) = \min_{\sigma_{\mathsf{A-cl}}\in\mathcal{C}_{\mathsf{A}}} \big\{S_\alpha(\rho||\sigma_{\mathsf{A-cl}})\big\}. \tag{377}$$

As in Corollary 11.A.1 one has $D_{\mathsf{A}}^{(1/2)}(\rho) = -2\ln(1 - D_{\mathsf{A}}(\rho)/2) \leq D_{\mathsf{A}}^{(\alpha)}(\rho)$ for any $\alpha \in [1/2, 1]$.

An analog of the geometric discord D_{A} based on the Hilbert-Schmidt distance d_2 has been first introduced by Dakić, Vedral, and Brukner.[46] We hope to have convinced the reader in Sec. VII that the Bures distance is a more natural choice in quantum information. We will see that the discord (376) shares many of the properties of the quantum discord δ_{A} of Sec. X, while its analog with the d_2-distance has unpleasant features. In particular, like δ_{A} the Bures geometric discord is invariant under conjugations by local unitaries and contractive with respect to quantum operations $\mathcal{M}_{\mathsf{B}}$ on B. For indeed, the set of A-classical states is invariant under such transformations (see (326)), whence

$$D_{\mathsf{A}}(U_{\mathsf{A}} \otimes U_{\mathsf{B}}\, \rho\, U_{\mathsf{A}}^* \otimes U_{\mathsf{B}}^*) = D_{\mathsf{A}}(\rho), \quad D_{\mathsf{A}}(1 \otimes \mathcal{M}_{\mathsf{B}}(\rho)) \leq D_{\mathsf{A}}(\rho) \tag{378}$$

by unitary invariance and contractivity of d_{B}. These properties also hold for $D_{\mathsf{A}}^{(\alpha)}$, $1/2 \leq \alpha \leq 1$, because the relative Rényi entropy is also contractive (Theorem 6.C.1). This should be contrasted with the non-monotonicity with respect to operations on B of the Hilbert-Schmidt geometric discord, which is due to the lack of monotonicity of d_2 (Sec. VII A). An explicit counter-example is given in Ref. 131. We now precise the axioms on discord-like correlation measures.

Definition 11.B.1. *A measure of quantum correlations of a bipartite system* AB *with respect to subsystem* A *is a function* $D_{\mathsf{A}} : \mathcal{E}(\mathcal{H}_{\mathsf{AB}}) \to [0, \infty)$ *satisfying*

(i) $D_{\mathsf{A}}(\rho) = 0$ *if and only if* ρ *is* A*-classical*;
(ii) D_{A} *is invariant under local unitary transformations and contractive under quantum operations on* B, *that is, (378) holds true*;
(iii) D_{A} *coincides with an entanglement measure for pure states*.

This definition is at the time of writing of this article believed to capture all relevant physical requirements for quantifying the amount of quantum correlations in AB given that one can access to subsystem A only.[134] The axioms (i-iii) are in particular satisfied by the quantum discord δ_{A} (Propositions 10.B.1 and 10.C.1). This is also true for the geometric discord D_{A}. Actually, we have just shown above that D_{A} satisfies (ii), and (i) is trivial. Since the closest separable state to a pure state is a pure product state, which is A-classical, D_{A} coincides with the geometric measure of entanglement E_{Bu} for pure states (see (381) below). Hence D_{A} is a measure of quantum correlations. Similarly, the relative-entropy based discord $D_{\mathsf{A}}^{(1)}$ is a measure of quantum correlations. The property (iii) follows in this case from the fact that if ρ_Ψ is a pure state then a separable state σ_{sep} minimizing $S(\rho_\Psi||\sigma_{\mathrm{sep}})$ is the classical state given by (361) (see the proof of Proposition 11.A.2), so that $D_{\mathsf{A}}^{(1)}(\rho_\Psi)$ coincides with the entanglement measure $E_1(\rho_\Psi)$ defined in (358). It is an open problem to show that $D_{\mathsf{A}}^{(\alpha)}$ satisfies (iii) when $\alpha \neq 1/2, 1$.

The B-discords D_{B} and $D_{\mathsf{B}}^{(\alpha)}$ are defined by exchanging A and B in (376) and (377). As for the quantum discord of Sec. X, in general $D_{\mathsf{A}} \neq D_{\mathsf{B}}$. Symmetric measures of quantum correlations are obtained by considering the square distance to the set of classical states $\mathcal{C}_{\mathsf{AB}} = \mathcal{C}_{\mathsf{A}} \cap \mathcal{C}_{\mathsf{B}}$,

$$D_{\mathsf{AB}}(\rho) = 2\Big(1 - \max_{\sigma_{\mathrm{clas}} \in \mathcal{C}_{\mathsf{AB}}} \big\{\sqrt{F(\rho, \sigma_{\mathrm{clas}})}\big\}\Big), \quad D_{\mathsf{AB}}^{(\alpha)}(\rho) = \min_{\sigma_{\mathrm{clas}} \in \mathcal{C}_{\mathsf{AB}}} \big\{S(\rho||\sigma_{\mathrm{clas}})\big\}\,. \tag{379}$$

Let us mention that a similar symmetric information-based discord can be defined by modifying the maximization procedure in (325) so as to involve projectors $\pi_i^{\mathsf{A}} \otimes \pi_i^{\mathsf{B}}$ (or generalized measurement operators $M_i^{\mathsf{A}} \otimes M_i^{\mathsf{B}}$), instead of $M_i^{\mathsf{A}} \otimes 1$. It is called the *measurement-induced disturbance*.[103] The relative-entropy symmetric discord $D_{\mathsf{AB}}^{(1)}$ has been studied in Ref. 109, together with other quantities characterizing quantum and classical correlations. We will not elaborate further here on the numerous discord-like measures defined in the literature and their operational interpretations (see, e.g., Ref. 110).

We emphasize that since $\mathcal{C}_{\mathsf{AB}} \subset \mathcal{C}_{\mathsf{A}} \subset \mathcal{S}_{\mathsf{AB}}$ (see Fig. 3), the geometric measures are ordered as

$$E_{\mathrm{Bu}}(\rho) \leq D_{\mathsf{A}}(\rho) \leq D_{\mathsf{AB}}(\rho). \tag{380}$$

This ordering is a nice feature of the geometrical approach. It also holds for the relative-entropy measures. In contrast, depending on ρ the entanglement of formation $E_{\mathrm{EoF}}(\rho)$ can be larger or smaller than the quantum discord $\delta_{\mathsf{A}}(\rho)$.

Before going on to general results, let us say few words about explicit calculations of the discords. In the special case of two-qubit states ρ with maximally mixed marginals $\rho_{\mathsf{A}} = \rho_{\mathsf{B}} = 1/2$, the relative-entropy measure $D_{\mathsf{AB}}^{(1)}(\rho)$ coincides with the usual discord $\delta_{\mathsf{A}}^{\mathrm{v.N.}}(\rho)$.[109,108] For the same states, a closed formula for $D_{\mathsf{A}}(\rho)$ has been found in Refs. 1 and 150 and the closest A-classical states to ρ have been determined explicitly (this is done in Ref. 150 with the help of Corollary 11.B.6 below). The Hilbert-Schmidt geometric discord is much easier to calculate. A simple formula for arbitrary 2-qubit states is derived in Ref. 46 and has been later on extended to higher dimensions. The geometric discord defined with the trace distance d_1 has been determined recently for certain families of two-qubit states (the so-called X-states, containing in particular

the states with maximally mixed marginals, and the B-classical states).[43,115] Note that since d_1 is contractive, this geometric discord fulfills the axiom (ii) of Definition 11.B.1.

2. Geometric discord for pure states

We now proceed to determine the geometric discord D_{A} for pure states. It has been seen in the proof of Proposition 11.A.4 that the family of closest separable states of a pure state $|\Psi\rangle$ contains a pure product state, which is a classical state. By inspection of (366) and (380), one gets

$$D_{\mathsf{A}}(|\Psi\rangle) = D_{\mathsf{B}}(|\Psi\rangle) = D_{\mathsf{AB}}(|\Psi\rangle) = E_{\mathrm{Bu}}(|\Psi\rangle) = 2(1-\sqrt{\mu_{\max}}). \tag{381}$$

One deduces from the bound $\mu_{\max} \geq 1/n$ (which follows from $\sum_{i=1}^n \mu_i = 1$) that

$$D_{\mathsf{A}}(|\Psi\rangle) \leq 2\Big(1-\frac{1}{\sqrt{n}}\Big), \quad n = \min\{n_{\mathsf{A}}, n_{\mathsf{B}}\}. \tag{382}$$

This bound is saturated when $\mu_i = 1/n$ for any i, that is, for the maximally entangled states. We will see below that this statement is still true for mixed states provided that $n_{\mathsf{A}} \leq n_{\mathsf{B}}$.

The identities (381) are analogous to the equality between the entanglement of formation E_{EoF} and the discord δ_{A} for pure states (Sec. X A). As said before, they reflect the existence of a pure product state which is closer or at the same distance from $|\Psi\rangle$ than any other separable state. It is of interest to find all the closest A-classical states to $|\Psi\rangle$. This is done in the next proposition.

Proposition 11.B.2. (Spehner and Orszag[149]) *Let $\rho_\Psi = |\Psi\rangle\langle\Psi|$ be a pure state of* AB *with largest Schmidt coefficient $\mu_{\max}$. If $\mu_{\max}$ is non-degenerate, then the closest* A*-classical (respectively classical, separable) state to ρ_Ψ for the Bures distance is unique. It is given by the pure product state $|\alpha_{\max}\rangle|\beta_{\max}\rangle$, where $|\alpha_{\max}\rangle$ and $|\beta_{\max}\rangle$ are eigenvectors with eigenvalue $\mu_{\max}$ of $[\rho_\Psi]_{\mathsf{A}}$ and $[\rho_\Psi]_{\mathsf{B}}$, respectively. If $\mu_{\max}$ is r-fold degenerate, say $\mu_{\max} = \mu_1 = \ldots = \mu_r > \mu_{r+1}, \ldots, \mu_n$, then infinitely many* A*-classical (respectively classical, separable) states σ minimize $d_{\mathsf{B}}(\rho_\Psi, \sigma)$. These closest states are convex combinations of the pure product states $|\varphi_l\rangle|\chi_l\rangle$ with*

$$|\varphi_l\rangle = \sum_{i=1}^r u_{il}|\alpha_i\rangle, \quad |\chi_l\rangle = \sum_{i=1}^r \overline{u_{il}}|\beta_i\rangle, \quad l = 1, \ldots, r\ , \tag{383}$$

where $\{|\alpha_i\rangle\}_{i=1}^r$ and $\{|\beta_i\rangle\}_{i=1}^r$ are orthonormal families of Schmidt vectors associated to $\mu_{\max}$ in the Schmidt decomposition (9), and $(u_{il})_{i,l=1}^r$ is an arbitrary $r \times r$ unitary matrix.

It should be noticed that when $\mu_{\max}$ is degenerate, the vectors (383) provide together with $|\alpha_i\rangle$, $|\beta_i\rangle$, $i = r+1, \ldots, n$, a Schmidt decomposition of $|\Psi\rangle$ (in that case this decomposition is not unique, see Sec. II B). Conversely, disregarding the degeneracies of the other eigenvalues $\mu_i < \mu_{\max}$, all Schmidt decompositions of $|\Psi\rangle$ are of this form for some unitary matrix $(u_{il})_{i,l=1}^r$. Thus, the existence of an infinite family of closest A-classical states to $|\Psi\rangle$ is related to the non-uniqueness of the Schmidt vectors associated to $\mu_{\max}$, and this family contains the products $|\varphi_l\rangle|\chi_l\rangle$ of these vectors and convex combinations thereof. This shows in particular that the maximally entangled pure states are the pure states with the largest family of closest states (this family forms a (n^2+n-2) real-parameter sub-manifold of $\mathcal{E}(\mathcal{H}_{\mathsf{AB}})$.)

Proof. An arbitrary A-classical state σ can be decomposed as $\sigma = \sum_{ij} q_{ij}|\varphi_i\rangle\langle\varphi_i| \otimes |\chi_{j|i}\rangle\langle\chi_{j|i}|$. In much the same way as in the proof of Proposition 11.A.4, $F(|\Psi\rangle, \mathcal{C}_{\mathsf{A}}) = \mu_{\max}$ and the closest A-classical states to ρ fulfill

$$|\langle\varphi_i \otimes \chi_{j|i}|\Psi\rangle|^2 = \max_{\|\varphi\|=\|\chi\|=1} \big\{|\langle\varphi\otimes\chi|\Psi\rangle|^2\big\} = \mu_{\max} \quad \text{when } q_{ij} > 0. \tag{384}$$

We have thus to determine all $|\varphi\rangle \in \mathcal{H}_{\mathsf{A}}$ and $|\chi\rangle \in \mathcal{H}_{\mathsf{B}}$ such that $|\langle\varphi\otimes\chi|\Psi\rangle|^2 = \mu_{\max}$. This occurs if all inequalities in (370) are equalities. Let us first assume that $\mu_1 = \mu_{\max} > \mu_2, \ldots, \mu_n$. After a close look to (370) one immediately finds that $|\langle\varphi\otimes\chi|\Psi\rangle|^2 = \mu_{\max}$ if and only if $|\varphi\rangle = |\alpha_1\rangle$ and $|\chi\rangle = |\beta_1\rangle$ up to irrelevant phase factors. Hence (384) is satisfied for a single pair (i,j). Therefore, all the q_{ij} vanish except one and the closest A-classical state to $|\Psi\rangle$ is the pure product state $|\alpha_1\rangle|\beta_1\rangle$.

We now proceed to the degenerate case $\mu_1 = \ldots = \mu_r = \mu_{\max} > \mu_{r+1}, \ldots, \mu_n$. Let us establish the necessary and sufficient conditions for the inequalities in (370) to be equalities. For the first inequality, the condition is $\arg(\langle\varphi|\alpha_j\rangle\langle\chi|\beta_j\rangle) = \theta$ with θ independent of j. For the second one, the condition is that $|\varphi\rangle$ belongs to $V_{\max} = \mathrm{span}\{|\alpha_j\rangle\}_{j=1}^r$ or $|\chi\rangle$ belongs to $W_{\max} = \mathrm{span}\{|\beta_j\rangle\}_{j=1}^r$. The Cauchy-Schwarz inequality in (370) is saturated if and only if $|\langle\varphi|\alpha_j\rangle| = \lambda|\langle\chi|\beta_j\rangle|$ for all j, with $\lambda \geq 0$. Finally, the last inequality holds with equality if and only if $|\varphi\rangle \in \mathrm{span}\{|\alpha_j\rangle\}_{j=1}^n$ and $|\chi\rangle \in \mathrm{span}\{|\beta_j\rangle\}_{j=1}^n$. Putting all conditions together, we obtain $|\varphi\rangle \in V_{\max}$, $|\chi\rangle \in W_{\max}$, and $\langle\chi|\beta_j\rangle = e^{\mathrm{i}\theta}\langle\alpha_j|\varphi\rangle$ for $j = 1, \ldots, r$. Therefore, from any orthonormal family $\{|\varphi_l\rangle\}_{l=1}^r$ of $V_{\max}$ one can construct r orthogonal vectors $|\varphi_l \otimes \chi_l\rangle$ satisfying $|\langle\varphi_l \otimes \chi_l|\Psi\rangle|^2 = \mu_{\max}$ for all $l = 1, \ldots, r$, with $\langle\chi_l|\beta_j\rangle = \langle\alpha_j|\varphi_l\rangle$. The probabilities $\{q_{ij}\}$ are then given by $q_{ij} = q_i$ if $i = j \leq r$ and zero otherwise, $\{q_l\}_{l=1}^r$ being an arbitrary probability distribution. The corresponding A-classical states σ maximizing the fidelity $F(\rho_\Psi, \sigma)$ are the classical states

$$\sigma = \sum_{l=1}^{r} q_l |\alpha_l \otimes \beta_l\rangle\langle\alpha_l \otimes \beta_l| \,. \tag{385}$$

□

3. Geometric discord for mixed states and quantum state discrimination

As for all other measures of entanglement and quantum correlations, determining $D_{\mathsf{A}}(\rho)$ is harder for mixed states than for pure states. Interestingly, this problem is related to an ambiguous quantum state discrimination task.

Proposition 11.B.3. (Spehner and Orszag[149]) *For any state ρ of the bipartite system AB, the maximal fidelity between ρ and an A-classical state reads*

$$F(\rho, \mathcal{C}_{\mathsf{A}}) = \max_{\{|\varphi_i\rangle\}} \left\{ P_{\mathrm{S,a}}^{\text{opt v.N.}}(\{\rho_i, \eta_i\}) \right\} = \max_{\{|\varphi_i\rangle\}} \max_{\{\Pi_i\}} \left\{ \sum_{i=1}^{n_{\mathsf{A}}} \eta_i \,\mathrm{tr}(\Pi_i \rho_i) \right\}, \tag{386}$$

where the maxima are over all orthonormal bases $\{|\varphi_i\rangle\}_{i=1}^{n_{\mathsf{A}}}$ of $\mathcal{H}_{\mathsf{A}}$ and all von Neumann measurements given by orthonormal families $\{\Pi_i\}_{i=1}^{n_{\mathsf{A}}}$ of projectors of $\mathcal{H}_{\mathsf{AB}}$ with rank n_{B}. Here, $P_{\mathrm{S,a}}^{\text{opt v.N.}}(\{\rho_i, \eta_i\})$ is the maximal success probability in discriminating ambiguously by such measurements the states ρ_i with probabilities η_i defined by

$$\eta_i = \langle\varphi_i|\rho_{\mathsf{A}}|\varphi_i\rangle, \quad \rho_i = \eta_i^{-1}\sqrt{\rho}|\varphi_i\rangle\langle\varphi_i| \otimes 1\sqrt{\rho}, \quad i = 1, \ldots, n_{\mathsf{A}} \tag{387}$$

(if $\eta_i = 0$ then ρ_i is not defined but does not contribute to the sum in (386)). Furthermore, the closest A-classical states to ρ are given by

$$\sigma_\rho = \frac{1}{F(\rho, \mathcal{C}_{\mathsf{A}})} \sum_{i=1}^{n_{\mathsf{A}}} |\varphi_i^{\text{opt}}\rangle\langle\varphi_i^{\text{opt}}| \otimes \langle\varphi_i^{\text{opt}}|\sqrt{\rho}\,\Pi_i^{\text{opt}}\sqrt{\rho}|\varphi_i^{\text{opt}}\rangle \,, \tag{388}$$

where $\{|\varphi_i^{\text{opt}}\rangle\}$ and $\{\Pi_i^{\text{opt}}\}$ are any orthonormal basis of $\mathcal{H}_{\mathsf{A}}$ and von Neumann measurement maximizing the right-hand side of (386).

The ρ_i are quantum states if $\eta_i > 0$ because $\rho_i \geq 0$ and η_i is chosen such that $\mathrm{tr}(\rho_i) = 1$. Moreover, $\{\eta_i\}_{i=1}^{n_{\mathsf{A}}}$ is a probability distribution (since $\eta_i \geq 0$ and $\sum_i \eta_i = \mathrm{tr}(\rho) = 1$) and the ensemble $\{\rho_i, \eta_i\}_{i=1}^{n_{\mathsf{A}}}$ is a convex decomposition of ρ, i.e., $\rho = \sum_i \eta_i \rho_i$.

Corollary 11.B.4. *If ρ is invertible then one can substitute $P_{\mathrm{S,a}}^{\text{opt v.N.}}(\{\rho_i, \eta_i\})$ in (386) by the maximal success probability $P_{\mathrm{S,a}}^{\text{opt}}(\{\rho_i, \eta_i\})$ over all POVMs, given by (64).*

Proof. This is a simple consequence of Corollary 5.D.3. Actually, if $\rho > 0$ then the states ρ_i defined in (387) are linearly independent, thus the optimal measurement to discriminate them is a von Neumann measurement with projectors of rank $r_i = \mathrm{rank}(\rho_i)$. The linear independence can be justified as follows. Let us first notice that ρ_i has rank $r_i = n_{\mathsf{B}}$ (for indeed, it has the same rank as

$\eta_i \rho^{-1/2} \rho_i = |\varphi_i\rangle\langle\varphi_i| \otimes 1\sqrt{\rho}$). A necessary and sufficient condition for $|\xi_{ij}\rangle$ to be an eigenvector of ρ_i with eigenvalue $\lambda_{ij} > 0$ is $|\xi_{ij}\rangle = (\lambda_{ij}\eta_i)^{-1}\sqrt{\rho}|\varphi_i\rangle \otimes |\zeta_{ij}\rangle$, where $|\zeta_{ij}\rangle \in \mathcal{H}_{\mathsf{B}}$ is an eigenvector of $R_i = \langle\varphi_i|\rho|\varphi_i\rangle$ with eigenvalue $\lambda_{ij}\ \eta_i > 0$. For any i, the Hermitian invertible matrix R_i admits an orthonormal eigenbasis $\{|\zeta_{ij}\rangle\}_{j=1}^{n_{\mathsf{B}}}$. Thanks to the invertibility of $\sqrt{\rho}$, $\{|\xi_{ij}\rangle\}_{i=1,\ldots,n_{\mathsf{A}}}^{j=1,\ldots,n_{\mathsf{B}}}$ is a basis of $\mathcal{H}_{\mathsf{AB}}$ and thus the states ρ_i are linearly independent and span $\mathcal{H}_{\mathsf{AB}}$. □

Before going into the proof of the proposition, let us discuss the state discrimination problems when ρ is pure or A-classical. Of course, the values of $D_{\mathsf{A}}(\rho)$ are already known in these cases, being given by (381) and by $D_{\mathsf{A}}(\rho) = 0$, respectively, but it is instructive to recover that from Proposition 11.B.3. If $\rho = \rho_\Psi$ is pure then all states ρ_i with $\eta_i > 0$ are identical and equal to ρ_Ψ, so that $P_{\mathrm{S,a}}^{\text{opt v.N.}} = \max_{\{\Pi_i\}}\{\sum_i \eta_i \langle\Psi|\Pi_i|\Psi\rangle\} = \eta_{\max}$. One gets $F(\rho_\Psi, \mathcal{C}_{\mathsf{A}}) = \mu_{\max}$ by optimization over the basis $\{|\varphi_i\rangle\}$. If ρ is an A-classical state, i.e., if it can be decomposed as in (326), then the optimal basis $\{|\varphi_i^{\text{opt}}\rangle\}$ coincides with the basis appearing in this decomposition. With this choice one obtains $\eta_i = q_i$ and $\rho_i = |\varphi_i\rangle\langle\varphi_i| \otimes \sigma_{\mathsf{B}|i}$ for all i such that $q_i > 0$. The states ρ_i are orthogonal and can thus be perfectly discriminated by von Neumann measurements. This yields $F(\rho, \mathcal{C}_{\mathsf{A}}) = 1$ and $D_{\mathsf{A}}(\rho) = 0$ as it should be. Reciprocally, if $F(\rho, \mathcal{C}_{\mathsf{A}}) = 1$ then $P_{\mathrm{S,a}}^{\text{opt v.N.}}(\{\rho_i, \eta_i\}) = 1$ for some basis $\{|\varphi_i\rangle\}$ and the corresponding ρ_i must be orthogonal (Sec. V). Hence one can find an orthonormal family $\{\Pi_i\}$ of projectors with rank n_{B} such that $\rho_i = \Pi_i\rho_i\Pi_i$ for any i with $\eta_i > 0$. It is an easy exercise to show that this implies that $\Pi_i = |\varphi_i\rangle\langle\varphi_i| \otimes 1$ if $\rho|_{\Pi_i\mathcal{H}}$ is invertible. Thus $\rho = \sum_i \eta_i\rho_i$ is A-classical, in agreement with the fact (following directly from the definition) that $D_{\mathsf{A}}(\rho) = 0$ if and only if ρ is A-classical.

The above discussion provides a clear interpretation of the result of Proposition 11.B.3: the states ρ with non-zero discord are characterized by ensembles $\{\rho_i, \eta_i\}$ of non-orthogonal states, which thereby are not perfectly distinguishable, for any orthonormal basis $\{|\varphi_i\rangle\}$ of $\mathcal{H}_{\mathsf{A}}$. The less distinguishable are the ρ_i's, the most distant is ρ from the set of zero-discord states.

We will establish Proposition 11.B.3 by relying on the slightly more general statement summarized in the following lemma.

Lemma 11.B.5. *For a fixed family $\{\sigma_{\mathsf{A}|i}\}_{i=1}^n$ of states $\sigma_{\mathsf{A}|i} \in \mathcal{E}(\mathcal{H}_{\mathsf{A}})$ having orthogonal supports and spanning $\mathcal{H}_{\mathsf{A}}$, with $1 \le n \le n_{\mathsf{A}}$, let us define*

$$\mathcal{C}_{\mathsf{A}}(\{\sigma_{\mathsf{A}|i}\}) = \Big\{\sigma = \sum_{i=1}^n q_i \sigma_{\mathsf{A}|i} \otimes \sigma_{\mathsf{B}|i}\ ;\ \{q_i, \sigma_{\mathsf{B}|i}\}_{i=1}^n \text{ is a state ensemble on } \mathcal{H}_{\mathsf{B}}\Big\}. \tag{389}$$

Then

$$F\big(\rho, \mathcal{C}_{\mathsf{A}}(\{\sigma_{\mathsf{A}|i}\})\big) = \max_{\sigma \in \mathcal{C}_{\mathsf{A}}(\{\sigma_{\mathsf{A}|i}\})}\big\{F(\rho, \sigma)\big\} = \max_U\Big\{\sum_{i=1}^n \|W_i(U)\|_2^2\Big\}, \tag{390}$$

where the last maximum is over all unitaries U on $\mathcal{H}_{\mathsf{AB}}$ and

$$W_i(U) = \mathrm{tr}_{\mathsf{A}}\big(\sqrt{\sigma_{\mathsf{A}|i}} \otimes 1\ \sqrt{\rho}\, U\big). \tag{391}$$

Moreover, there exists a unitary U_{opt} achieving the maximum in (390) which is such that $W_i(U_{\text{opt}}) \ge 0$. The states σ_{opt} satisfying $F(\rho, \sigma_{\text{opt}}) = F(\rho, \mathcal{C}_{\mathsf{A}}(\{\sigma_{\mathsf{A}|i}\}))$ are given in terms of this unitary by

$$\sigma_{\text{opt}} = \frac{1}{F(\rho, \mathcal{C}_{\mathsf{A}}(\{\sigma_{\mathsf{A}|i}\}))} \sum_{i=1}^n \sigma_{\mathsf{A}|i} \otimes W_i(U_{\text{opt}})^2\ . \tag{392}$$

Proof. Using the spectral decompositions of the states $\sigma_{\mathsf{B}|i}$, any $\sigma \in \mathcal{C}_{\mathsf{A}}(\{\sigma_{\mathsf{A}|i}\})$ can be written as

$$\sigma = \sum_{i=1}^n \sum_{j=1}^{n_{\mathsf{B}}} q_{ij} \sigma_{\mathsf{A}|i} \otimes |\chi_{j|i}\rangle\langle\chi_{j|i}| \quad \text{with} \quad q_{ij} \ge 0\ ,\ \sum_{ij} q_{ij} = 1\ , \tag{393}$$

where $\{|\chi_{j|i}\rangle\}_{j=1}^{n_{\mathsf{B}}}$ is an orthonormal basis of $\mathcal{H}_{\mathsf{B}}$ for any i (compare with (333)). By assumption, if $i \ne i'$ then $\mathrm{ran}\,\sigma_{\mathsf{A}|i} \perp \mathrm{ran}\,\sigma_{\mathsf{A}|i'}$, so that $\sqrt{\sigma} = \sum_{i,j} \sqrt{q_{ij}}\sqrt{\sigma_{\mathsf{A}|i}} \otimes |\chi_{j|i}\rangle\langle\chi_{j|i}|$. We start by evaluating

the trace norm in the definition (176) of the fidelity by means of the formula $\|O\|_1 = \max_U |\mathrm{tr}(UO)|$ to obtain

$$F\big(\rho, \mathcal{C}_{\mathsf{A}}(\{\sigma_{\mathsf{A}|i}\})\big) = \max_{\sigma\in\mathcal{C}_{\mathsf{A}}(\{\sigma_{\mathsf{A}|i}\})} \max_U \Big\{ \big|\mathrm{tr}(U^*\sqrt{\rho}\sqrt{\sigma})\big|^2 \Big\}$$
$$= \max_U \left\{ \max_{\{q_{ij}\},\{|\chi_{j|i}\rangle\}} \left| \sum_{i,j} \sqrt{q_{ij}} \langle \chi_{j|i}| W_i(U)^* |\chi_{j|i}\rangle \right|^2 \right\}. \tag{394}$$

The square modulus can be bounded by invoking twice the Cauchy-Schwarz inequality and $\sum_{ij} q_{ij} = 1$,

$$\left| \sum_{i,j} \sqrt{q_{ij}} \langle \chi_{j|i}| W_i(U)^* |\chi_{j|i}\rangle \right|^2 \le \sum_{i,j} \big| \langle \chi_{j|i}| W_i(U)^* |\chi_{j|i}\rangle \big|^2$$
$$\le \sum_{i,j} \big\| W_i(U)|\chi_{j|i}\rangle \big\|^2 = \sum_i \|W_i(U)\|_2^2. \tag{395}$$

The foregoing inequalities are equalities if the following conditions are satisfied:

(1) $W_i(U) = W_i(U)^* \ge 0$;
(2) $q_{ij} = \langle \chi_{j|i}| W_i(U)|\chi_{j|i}\rangle^2 / (\sum_{i,j} \langle \chi_{j|i}| W_i(U)|\chi_{j|i}\rangle^2)$;
(3) $\{|\chi_{j|i}\rangle\}_{j=1}^{n_{\mathsf{B}}}$ is an eigenbasis of $W_i(U)$ for any i.

Therefore, (390) holds true provided that there is a unitary U on $\mathcal{H}_{\mathsf{AB}}$ satisfying (1). For a given U, let us define $U_{\mathrm{opt}} = U \sum_i \pi_i^{\mathsf{A}} \otimes V_i^*$, where π_i^{A} is the projector onto $\mathrm{ran}\,\sigma_{\mathsf{A}|i}$ and V_i a unitary on $\mathcal{H}_{\mathsf{B}}$ such that $W_i(U) = |W_i(U)^*|V_i$ (polar decomposition). Then U_{opt} is unitary since by hypothesis $\pi_i^{\mathsf{A}}\pi_{i'}^{\mathsf{A}} = \delta_{ii'}\pi_i^{\mathsf{A}}$ and $\sum_i \pi_i^{\mathsf{A}} = 1$, and one readily shows that $W_i(U_{\mathrm{opt}}) = W_i(U)V_i^* \ge 0$. As $\sum_i \|W_i(U)\|_2^2 = \sum_i \|W_i(U_{\mathrm{opt}})\|_2^2$, the identity (390) follows from (394) and (395). From condition (3) one has $W_i(U_{\mathrm{opt}})|\chi_{j|i}^{\mathrm{opt}}\rangle = w_{ji}|\chi_{j|i}^{\mathrm{opt}}\rangle$ with $\sum_{i,j} w_{ji}^2 = F(\rho, \mathcal{C}_{\mathsf{A}}(\{\sigma_{\mathsf{A}|i}\}))$, see (395). Condition (2) entails

$$\sigma_{\mathsf{B}|i}^{\mathrm{opt}} = \sum_j q_{ij}^{\mathrm{opt}} |\chi_{j|i}^{\mathrm{opt}}\rangle\langle\chi_{j|i}^{\mathrm{opt}}| = \frac{W_i(U_{\mathrm{opt}})^2}{F(\rho, \mathcal{C}_{\mathsf{A}}(\{\sigma_{\mathsf{A}|i}\}))}\,, \tag{396}$$

which together with (393) leads to (392). □

Proof of Proposition 11.B.3. Let $\{|\varphi_i\rangle\}_{i=1}^{n_{\mathsf{A}}}$ be an orthonormal basis of $\mathcal{H}_{\mathsf{A}}$. Applying Lemma 11.B.5 with $\sigma_{\mathsf{A}|i} = |\varphi_i\rangle\langle\varphi_i|$ one gets

$$F\big(\rho, \mathcal{C}_{\mathsf{A}}(\{|\varphi_i\rangle\})\big) = \max_U \left\{ \sum_{i=1}^{n_{\mathsf{A}}} \mathrm{tr}\big[U|\varphi_i\rangle\langle\varphi_i| \otimes 1\, U^*\sqrt{\rho}\,|\varphi_i\rangle\langle\varphi_i| \otimes 1\,\sqrt{\rho}\big] \right\},$$
$$= \max_{\{\Pi_i\}} \left\{ \sum_{i=1}^{n_{\mathsf{A}}} \mathrm{tr}\big[\Pi_i\sqrt{\rho}|\varphi_i\rangle\langle\varphi_i| \otimes 1\,\sqrt{\rho}\big] \right\} = P_{\mathrm{S,a}}^{\mathrm{opt\,v.N.}}(\{\rho_i, \eta_i\})\,. \tag{397}$$

The last maximum is over all orthonormal families $\{\Pi_i\}_{i=1}^{n_{\mathsf{A}}}$ of projectors of rank n_{B} and $P_{\mathrm{S,a}}^{\mathrm{opt\,v.N.}}(\{\rho_i, \eta_i\})$ is given by (94). Since the fidelity $F(\rho, \mathcal{C}_{\mathsf{A}})$ is the maximum of $F(\rho, \mathcal{C}_{\mathsf{A}}(\{|\varphi_i\rangle\}))$ over all bases $\{|\varphi_i\rangle\}$, this leads to (386) and (388). □

4. The qubit case

It has been emphasized in Sec. V that the optimal success probability and measurement for discriminating ambiguously more than two states are not known explicitly in general. Nonetheless, if the subsystem A is a qubit, the ensemble $\{\rho_i, \eta_i\}$ in Proposition 11.B.3 contains only $n_{\mathsf{A}} = 2$ states and the optimal probability and measurement are easily determined. Following the steps yielding to

(68) we find

$$P_{\mathrm{S,a}}^{\mathrm{opt\,v.N.}}(\{\rho_i,\eta_i\}) = \frac{1}{2}\big(1-\mathrm{tr}\,\Lambda\big) + \sum_{l=1}^{n_{\mathsf{B}}}\lambda_l\ , \tag{398}$$

where $\lambda_1 \geq \cdots \geq \lambda_{n_{\mathsf{B}}}$ are the n_{B} largest eigenvalues of $\Lambda = \eta_0\rho_0 - \eta_1\rho_1$. The optimal von Neumann measurement is formed by the spectral projector Π_0^{opt} of Λ for these n_{B} eigenvalues and its complement $\Pi_1^{\mathrm{opt}} = 1 - \Pi_0^{\mathrm{opt}}$. For the states ρ_i associated to the orthonormal basis $\{|\varphi_i\rangle\}_{i=0}^1$ of $\mathbb{C}^2$ via formula (387), one has $\Lambda = \sqrt{\rho}\,(|\varphi_0\rangle\langle\varphi_0| - |\varphi_1\rangle\langle\varphi_1|)\otimes 1\,\sqrt{\rho}$. The operator inside the parenthesis in the last identity is equal to $\sigma_{\mathbf{u}} \equiv \sum_{m=1}^3 u_m\sigma_m$ for some unit vector $\mathbf{u}\in\mathbb{R}^3$ depending on $\{|\varphi_i\rangle\}$ (here σ_1, σ_2, and σ_3 are the Pauli matrices). Conversely, one can associate to any unit vector $\mathbf{u}\in\mathbb{R}^3$ the eigenbasis $\{|\varphi_i\rangle\}_{i=0}^1$ of $\sigma_{\mathbf{u}}$. According to Proposition 11.B.3, $F(\rho,\mathcal{C}_{\mathsf{A}})$ is obtained by maximizing the right-hand side of (398) over all Hermitian matrices

$$\Lambda(\mathbf{u}) = \sqrt{\rho}\,\sigma_{\mathbf{u}}\otimes 1\,\sqrt{\rho} \tag{399}$$

with $\mathbf{u}\in\mathbb{R}^3$, $|\mathbf{u}| = 1$. The following corollary of Proposition 11.B.3 is a refinement of a result in Ref. 150.

Corollary 11.B.6. *Let* A *be a qubit, i.e.,* $n_{\mathsf{A}} = 2$. *The fidelity between* ρ *and the set of* A*-classical states is given by*

$$F(\rho,\mathcal{C}_{\mathsf{A}}) = \frac{1}{2}\max_{\|\mathbf{u}\|=1}\big\{1+\|\Lambda(\mathbf{u})\|_1\big\}\ , \tag{400}$$

where $\Lambda(\mathbf{u})$ *is the* $2n_B \times 2n_B$ *matrix (399). The closest* A*-classical states to* ρ *are given by (388) where* Π_0^{opt} *is the spectral projector associated to the* n_{B} *largest eigenvalues of* $\Lambda(\mathbf{u}^{\mathrm{opt}})$ *and* $\mathbf{u}^{\mathrm{opt}}\in\mathbb{R}^3$ *is a unit vector achieving the maximum in (400).*

Proof. Let $\lambda_l(\mathbf{u})$ be the eigenvalues of $\Lambda(\mathbf{u})$ in non-increasing order. We claim that

$$-\frac{1}{2}\,\mathrm{tr}(\Lambda(\mathbf{u})) + \sum_{l=1}^{n_{\mathsf{B}}}\lambda_l(\mathbf{u}) = \frac{1}{2}\sum_{l=1}^{n_{\mathsf{B}}}\lambda_l(\mathbf{u}) - \frac{1}{2}\sum_{l=n_{\mathsf{B}}+1}^{2n_{\mathsf{B}}}\lambda_l(\mathbf{u}) = \frac{1}{2}\,\mathrm{tr}\,|\Lambda(\mathbf{u})|\ . \tag{401}$$

To prove this claim it suffices to show that $\Lambda(\mathbf{u})$ has at most n_{B} positive eigenvalues $\lambda_l(\mathbf{u}) > 0$ and at most n_{B} negative eigenvalues $\lambda_l(\mathbf{u}) < 0$, counting multiplicities. As $\ker\rho\subset\ker\Lambda(\mathbf{u})$ one may without loss of generality restrict $\Lambda(\mathbf{u})$ to the subspace $\Pi\mathcal{H}_{\mathsf{AB}}$, with Π the projector onto $\mathrm{ran}(\rho)$. A standard linear algebra argument shows that if S is a finite invertible matrix and Σ a self-adjoint matrix, then the number of positive (respectively, negative) eigenvalues of Σ is equal to the number of positive (respectively negative) eigenvalues of $S^*\Sigma S$. Let $P_\Sigma^\pm$ be the spectral projectors of $\Sigma = \Pi\,\sigma_{\mathbf{u}}\otimes 1\,\Pi$ on $\mathbb{R}_\pm\setminus\{0\}$. Since $\sqrt{\rho}:\Pi\mathcal{H}_{\mathsf{AB}}\to\Pi\mathcal{H}_{\mathsf{AB}}$ is invertible, in order to prove (401) it is thus enough to verify that $\mathrm{rank}(P_\Sigma^\pm)\leq n_{\mathsf{B}}$. This is evident if $\mathrm{rank}(\Pi)\leq n_{\mathsf{B}}$. If $\mathrm{rank}(\Pi) > n_{\mathsf{B}}$, then $\pm\langle\Psi|\sigma_{\mathbf{u}}\otimes 1|\Psi\rangle = \pm\langle\Psi|\Sigma|\Psi\rangle > 0$ for any $|\Psi\rangle\in P_\Sigma^\pm\mathcal{H}_{\mathsf{AB}}\subset\Pi\mathcal{H}_{\mathsf{AB}}$. This implies that $\mathrm{rank}(P_\Sigma^\pm)\leq\mathrm{rank}(P_{\sigma_{\mathbf{u}}\otimes 1}^\pm) = n_{\mathsf{B}}$, as otherwise one could find a non-vanishing vector $|\Psi\rangle\in P_\Sigma^\pm\mathcal{H}_{\mathsf{AB}}$ belonging to the n_{B}-dimensional eigenspace of $\sigma_{\mathbf{u}}\otimes 1$ with eigenvalue ∓ 1, in contraction with the foregoing inequality. This establishes (401). Then (400) follows from (398) and Proposition 11.B.3. □

5. States with the highest geometric discord

The geometric discord D_{A}, as the quantum discord δ_{A}, quantifies the degree of quantumness of a state. Let us recall from Sec. X C 2 that when the space dimensions of A and B are such that $n_{\mathsf{A}}\leq n_{\mathsf{B}}$, the "most quantum" states ρ having the highest discord $\delta_{\mathsf{A}}(\rho)$ are the maximally entangled states, i.e., the states with the highest entanglement of formation $E_{\mathrm{EoF}}(\rho) = \ln n_{\mathsf{A}}$. It is comforting that a similar result holds for the geometric discord.

Corollary 11.B.7. *If $n_A \leq n_B$, the highest value of $D_A(\rho)$ on $\mathcal{E}(\mathcal{H}_{AB})$ is equal to $2 - 2/\sqrt{n_A}$. The most distant states ρ from the set of A-classical states, which are such that $D_A(\rho) = 2 - 2/\sqrt{n_A}$, are the maximally entangled states given by Proposition 9.E.1.*

Comparing with the results of Sec. XI A 2, we see that when $n_A \leq n_B$ the most distant states from $\mathcal{C}_A$ are also the most distant from the set of separable states $\mathcal{S}_{AB}$. If $n_A \leq n_B < 2n_A$, these most distant states are maximally entangled pure states, as illustrated in Fig. 3.

Proof. This is again a corollary of Proposition 11.B.3. The success probability $P_{S,a}^{\text{opt v.N.}}$ is clearly larger or equal to the highest prior probability $\eta_{\max} = \max_i\{\eta_i\}$. (A receiver would obtain $P_{S,\,a} = \eta_{\max}$ by simply guessing that his state is $\rho_{i_{\max}}$, with $\eta_{i_{\max}} = \eta_{\max}$, whatever the measurement outcomes. A better strategy is of course to perform the von Neumann measurement $\{\Pi_i\}$ such that $\Pi_{i_{\max}}$ projects onto a n_B-dimensional subspace containing $\text{ran}(\rho_{i_{\max}})$. This range has a dimension $\text{rank}(\rho_{i_{\max}}) \leq n_B$ by a similar reasoning as in the proof of Corollary 11.B.4.) In view of Proposition 11.B.3 and $\eta_{\max} \geq 1/n_A$, we get

$$F(\rho, \mathcal{C}_A) \geq \frac{1}{n_A} \tag{402}$$

for any state ρ. When $n = n_A \leq n_B$ this bound is optimal, the value $1/n$ being achieved for the maximally entangled pure states (Sec. XI B 2). This proves the first statement. Let ρ be a state such that $F(\rho, \mathcal{C}_A) = 1/n$. According to (386) and since it has been argued above that $P_{S,a}^{\text{opt v.N.}} \geq \eta_{\max} \geq 1/n$, this implies that $P_{S,a}^{\text{opt v.N.}}(\{\rho_i, \eta_i\}) = 1/n$ whatever the orthonormal basis $\{|\varphi_i\rangle\}$. It is intuitively clear that this can happen only if the receiver gets a collection of identical states ρ_i with equal prior probabilities $\eta_i = 1/n$ (an explicit proof of this fact can be found in Ref. 149). From (387) and $\rho = \sum \eta_i \rho_i$ one obtains $\rho_A = 1/n$ and $\rho_i = \rho$ for any i and $\{|\varphi_i\rangle\}$. Plugging the spectral decomposition $\rho = \sum_k p_k |k\rangle\langle k|$ into (387), the second equality yields $D_{kl} = \text{tr}_B(|k\rangle\langle l|) = n^{-1}\delta_{kl}$ for all k and l such that $p_k p_l \neq 0$. One concludes that ρ has maximal entanglement of formation by following the same steps as in the proof of Proposition 9.E.1. □

One may wonder if Corollary 11.B.7 could also hold for $n_A > n_B$ (modulo the exchange $n_A \leftrightarrow n_B$), as what happens for the geometric measure of entanglement (see Sec. XI A 2). However, unlike $E_{Bu}(\rho)$ the geometric discord is not symmetric under the exchange of the two subsystems. The problem of determining its highest value and the corresponding "most quantum" states is still open for $n_A > n_B$. For such space dimensions the bound (402) is still correct but it is not optimal, that is, there are no states ρ with fidelity $F(\rho, \mathcal{C}_A) = 1/n_A$. Indeed, one can show as in the proof above that if $F(\rho, \mathcal{C}_A) = 1/n_A$ then the eigenvectors $|k\rangle$ of ρ with eigenvalues $p_k > 0$ have maximally mixed marginals $D_{kk} = (|k\rangle\langle k|)_A = 1/n_A$. But this is impossible since $\text{rank}(D_{kk}) \leq n_B$ by (9).

Remark 11.B.8. One can place a lower bound on $F(\rho, \mathcal{C}_A)$ for $n_A > n_B$ by invoking the inequality[149]

$$F(\rho, \mathcal{C}_A) \geq \frac{\|\rho\|}{n_B} + \frac{1 - \|\rho\|}{n_A}\frac{n_B - \delta_\rho}{n_B} \tag{403}$$

where $\delta_\rho = 0$ if $\text{rank}(\rho) \leq n_B$ and 1 otherwise.

Table I presents a comparison of the properties of the entanglement of formation, the quantum discord, and their geometrical analogs based on the Bures distance.

6. Geometric discord and least square measurements

The ensemble $\{\rho_i, \eta_i\}$ in the discrimination task associated to the geometric discord in Proposition 11.B.3 turns out to be related to the transpose operation of the von Neumann measurement in the basis $\{|\varphi_i\rangle\}$. In fact, let us denote by $\mathcal{M}_A$ the measurement on A with rank-one orthonormal projectors $\pi_i^A = |\varphi_i\rangle\langle\varphi_i|$. Let

$$\eta_i = \langle\varphi_i|\rho_A|\varphi_i\rangle, \quad \rho_{AB|i} = \eta_i^{-1}|\varphi_i\rangle\langle\varphi_i| \otimes \langle\varphi_i|\rho|\varphi_i\rangle \tag{404}$$

TABLE I. Summary of the definitions and properties of the entanglement of formation (Sec. IX), quantum discord (Sec. X), geometric measure of entanglement (Sec. XI A), and geometric discord (Sec. XI B). Here n_{A} and n_{B} are the space dimensions of the subsystems A and B, $n = \min\{n_{\mathsf{A}}, n_{\mathsf{B}}\}$, and μ_i are the Schmidt coefficients in (9).

	Entanglement of formation	Quantum discord	Geometric entanglement	Geometric discord
AB in a pure state	$E_{\mathrm{EoF}}(\lvert\Psi\rangle) = \delta_{\mathsf{A}}(\lvert\Psi\rangle) = H(\{\mu_i\})$		$E_{\mathrm{Bu}}(\lvert\Psi\rangle) = D_{\mathsf{A}}(\lvert\Psi\rangle) = 2(1-\sqrt{\mu_{\max}})$	
AB in a mixed state	$E_{\mathrm{EoF}}(\rho) = \min\{\sum_i \eta_i E_{\mathrm{EoF}}(\lvert\Psi_i\rangle)\}$ (convex roof)	$\delta_{\mathsf{A}}(\rho) = I_{\mathsf{A}:\mathsf{B}}(\rho) - \underbrace{\max\{I_{\mathsf{A}:\mathsf{B}}(\mathcal{M}_{\mathsf{A}}\otimes 1(\rho))\}}_{\text{classical correlations}}$	$E_{\mathrm{Bu}}(\rho) = 2(1-\underbrace{\max\{\sqrt{F(\rho,\sigma_{\mathrm{sep}})}\}}_{=\text{ convex roof}})$	$D_{\mathsf{A}}(\rho) = 2(1-\underbrace{\max\{\sqrt{F(\rho,\sigma_{\mathsf{A}-\mathrm{cl}})}\}}_{\text{= max. success proba. in state discrimination}})$
Vanishes iff	ρ is separable	ρ is A-classical	ρ is separable	ρ is A-classical
Maximal iff with max. value	ρ is max. entangled $\ln n$	EoF :true $\forall\, n_{\mathrm{A,B}}$; δ_{A} : true if $n_{\mathrm{A}} \le n_{\mathrm{B}}$	ρ is max. entangled $2(1-1/\sqrt{n})$	E_{Bu} : true $\forall\, n_{\mathrm{A,B}}$; D_{A} : true if $n_{\mathrm{A}} \le n_{\mathrm{B}}$
Local unit. invariance	✓	✓	✓	✓
Monotonicity w.r.t.	LOCCs	operations on B	LOCCs	operations on B
Convexity	✓	no	✓	no
Ordering	no		$E_{\mathrm{Bu}}(\rho) \le D_{\mathsf{A}}(\rho)$	
ABC in a pure state	$E_{\mathrm{EoF}}(\rho_{\mathsf{BC}}) = \delta_{\mathsf{A}}(\rho_{\mathsf{AB}}) + S(\rho_{\mathsf{AB}}) - S(\rho_{\mathsf{A}})$		?	

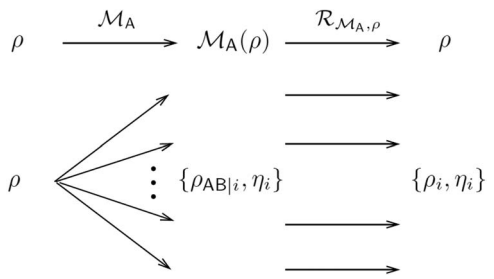

FIG. 4. State changes under the von Neumann measurement $\mathcal{M}_{\mathsf{A}}$ with rank-one projectors $\pi_i^{\mathsf{A}} = |\varphi_i\rangle\langle\varphi_i|$ followed by its transpose operation $\mathcal{R}_{\mathcal{M}_{\mathsf{A}},\rho}$. The upper line corresponds to a measurement without readout and the other lines to the different measurement outcomes.

be the corresponding probabilities and post-measurement conditional states when the initial state is ρ. The transpose operation of $\mathcal{M}_{\mathsf{A}}$ for ρ is (see (45))

$$\mathcal{R}_{\mathcal{M}_{\mathsf{A}},\rho}(\sigma) = \sum_{i=1}^{n_{\mathsf{A}}} \sqrt{\rho}|\varphi_i\rangle\langle\varphi_i| \otimes \langle\varphi_i|\rho|\varphi_i\rangle^{-\frac{1}{2}}\langle\varphi_i|\sigma|\varphi_i\rangle\langle\varphi_i|\rho|\varphi_i\rangle^{-\frac{1}{2}}\sqrt{\rho}\ . \tag{405}$$

We observe that

$$\rho_i = \mathcal{R}_{\mathcal{M}_{\mathsf{A}},\rho}(\rho_{\mathsf{AB}|i}), \quad i = 1, \ldots, n_{\mathsf{A}}. \tag{406}$$

Comparing (44) and (406), one expects from the discussion in Sec. IV C that the least square measurement $\{M_i^{\mathrm{lsm}}\}$ for the ensemble $\{\rho_i, \eta_i\}$ is associated to the transpose operation of $\mathcal{R}_{\mathcal{M}_{\mathsf{A}},\rho}$ for $\mathcal{M}_{\mathsf{A}}(\rho) = \sum_i \eta_i \rho_{\mathsf{AB}|i}$. But this two-fold transpose operation coincides with $\mathcal{M}_{\mathsf{A}}$, hence $\{M_i^{\mathrm{lsm}}\}$ is nothing but the von Neumann measurement on A in the basis $\{|\varphi_i\rangle\}$. This can be readily checked: since $\{\rho_i, \eta_i\}$ is a convex decomposition of ρ, (57) leads to

$$M_i^{\mathrm{lsm}} = \eta_i \rho^{-1/2}\rho_i\rho^{-1/2} = \pi_i^{\mathsf{A}} \otimes 1\ . \tag{407}$$

One can bound $P_{\mathrm{S,a}}^{\text{opt v.N.}}(\{\rho_i, \eta_i\})$ from below by the success probability obtained by discriminating the ρ_i with $\{M_i^{\mathrm{lsm}}\}$, and from above by the square root of this probability, see (83). By

Proposition 11.B.3, this yields

$$\max_{\{|\varphi_i\rangle\}}\left\{\sum_{i=1}^{n_{\mathsf{A}}}\mathrm{tr}_{\mathsf{B}}\big[\langle\varphi_i|\sqrt{\rho}|\varphi_i\rangle^2\big]\right\}\le F(\rho,\mathcal{C}_{\mathsf{A}})\le\max_{\{|\varphi_i\rangle\}}\left\{\sum_{i=1}^{n_{\mathsf{A}}}\mathrm{tr}_{\mathsf{B}}\big[\langle\varphi_i|\sqrt{\rho}|\varphi_i\rangle^2\big]\right\}^{\frac{1}{2}}. \tag{408}$$

The left- and right-hand sides become nearly equal when $F(\rho,\mathcal{C}_{\mathsf{A}})$ is almost one, that is, if ρ is close to $\mathcal{C}_{\mathsf{A}}$. Other inequalities on $F(\rho,\mathcal{C}_{\mathsf{A}})$ can be obtained in terms of the fidelities $F(\rho_i,\,\rho_j)$ with the help of Proposition 5.E.1.

The aforementioned observations can be summarized by Fig. 4, where outcomes.

ACKNOWLEDGMENTS

I am grateful to V. Eremeev, G. Ferrini, A. Joye, M. Orszag, and A. Smerzi for interesting discussions and to V. Jakšić for pointing out to me the works of Refs. 114 and 60. I acknowledge support from the ANR project no. ANR-13-JS01-0005-01.

APPENDIX A: OPERATOR MONOTONE AND OPERATOR CONVEX FUNCTIONS

We recall in this appendix some basic facts about operator monotone and operator convex functions. We refer the reader to the lecture notes[33] and the book[27] for more complete presentations of these notions.

We denote by $\mathcal{B}(\mathcal{H})_+$ the set of non-negative operators on $\mathcal{H}$, with $\dim(\mathcal{H})=n<\infty$. A function $f:\mathbb{R}_+\to\mathbb{R}$ is *operator convex* if for any $n\times n$ matrices $A,B\in\mathcal{B}(\mathcal{H})_+$ and any $0\le\eta\le1$, it holds $f((1-\eta)A+\eta B)\le(1-\eta)f(A)+\eta f(B)$. It is *strictly operator convex* if the inequality holds with equality if and only if $\eta\in0,1$ or $A=B$. It is *operator concave* if $-f$ is operator convex. It is *operator monotone-increasing* if for any $A,B\in\mathcal{B}(\mathcal{H})_+$, $A\le B\Rightarrow f(A)\le f(B)$, and *operator monotone-decreasing* if the reverse equality holds.

It is not hard to show (see, e.g., Ref. 33) that $f(x)=x^{-1}$ is operator monotone-decreasing and strictly operator convex. Clearly, this is then also true for $f(x)=(x+t)^{-1}$ for any $t\ge0$. According to the integral representation

$$A^{\alpha}=\frac{\sin(\alpha\pi)}{\pi}\int_0^{\infty}\mathrm{d}t\,t^{\alpha}\Big(\frac{1}{t}-\frac{1}{t+A}\Big), \tag{A1}$$

it follows that $f_\alpha(x)=x^\alpha$ is operator monotone-increasing and strictly operator concave for $0<\alpha<1$. Similarly, one shows that f_α is operator monotone-decreasing and operator convex for $\alpha\in[-1,0]$ and operator convex for $\alpha\in[1,2]$. However, for instance, the square function f_2 is not operator monotone and the cube function f_3 is not operator convex. One can establish that $g(x)=\ln x$ and $f(x)=x\ln x$ are operator concave and operator convex, respectively, thanks to the identities

$$\ln A=\lim_{\alpha\to0}\alpha^{-1}(A^{\alpha}-1),\qquad A\ln A=\lim_{\alpha\to1}\frac{A^{\alpha}-A}{\alpha-1}. \tag{A2}$$

Another example of monotone-increasing function is $f(x)=(x-1)/\ln x=\int_0^1\mathrm{d}\alpha\,x^{\alpha}$.

Operator monotonicity is much stronger than usual monotonicity of real functions. This is clear from Löwner's theorem, which states that if $f:(-1,1)\to\mathbb{R}$ is operator monotone and non-constant, then f admits the integral representation

$$f(x)=f(0)+f'(0)\int_{-1}^{1}\mathrm{d}\mu(t)\frac{x}{1-xt}, \tag{A3}$$

where μ is a probability measure on $[-1,1]$ (see Ref. 27, Corollary V.4.5). Furthermore, if $f:\mathbb{R}_+\to\mathbb{R}_+$ is continuous, then f is operator monotone if and only if it is operator concave (Ref. 27, Theorem V.2.5). The fact that concavity implies monotonicity is easily obtained by noting that if $0\le A\le B$, $C=B-A\ge0$, and $0\le\eta<1$, then $f(\eta B)\ge\eta f(A)+(1-\eta)f(\eta(1-\eta)^{-1}C)$ (by concavity). As $f(x)\ge0$ the second term in the right-hand side is non-negative and thus $f(\eta B)\ge\eta f(A)$. Letting $\eta\to1$ we get $f(B)\ge f(A)$. The converse implication can be shown by similar

arguments as those used to establish (A4) below and by invoking the fact that if (A4) is satisfied for any contraction C then f is operator convex (see Ref. 27 for more detail).

Another remarkable result valid for continuous functions $f : [0, a) \to \mathbb{R}$ is that f is operator convex and $f(0) \leq 0$ if and only if $g(x) = x^{-1} f(x)$ is operator monotone on $(0, a)$ (Ref. 27, Theorem V.2.9). Similarly, for functions $f : (-1, 1) \to \mathbb{R}$ of class C^2, if f is operator convex and $f(0) = 0$ then $g(x)$ is operator monotone (Ref. 27, Corollary V.3.11). An integral representation for non-linear operator convex functions f can be obtained with the help of the last property, by applying (A3) to $g(x)$.

If $f : \mathbb{R}_+ \to \mathbb{R}$ is operator convex and $f(0) \leq 0$, then

$$f(C^* A C) \leq C^* f(A) C \tag{A4}$$

for any contraction $C \in \mathcal{B}(\mathcal{H})$, $\|C\| \leq 1$, and any $A \in \mathcal{B}(\mathcal{H})_+$. This inequality can be shown as follows.[69] Let us consider the matrices

$$\widehat{A} = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}, \quad \widehat{U}_\pm = \begin{pmatrix} C & \pm D \\ E & \mp C^* \end{pmatrix} \tag{A5}$$

with $D = \sqrt{1 - CC^*}$ and $E = \sqrt{1 - C^* C}$ (the latter operators are well defined since $\|C\| \leq 1$). An explicit calculation shows that $\widehat{U}_\pm$ is unitary and

$$\begin{pmatrix} C^* A C & 0 \\ 0 & DAD \end{pmatrix} = \frac{1}{2} \sum_{\epsilon = \pm} \widehat{U}_\epsilon^* \widehat{A}\, \widehat{U}_\epsilon \ . \tag{A6}$$

If f is operator convex and $f(0) \leq 0$, then

$$\begin{aligned} \begin{pmatrix} f(C^* A C) & 0 \\ 0 & f(DAD) \end{pmatrix} &= f \begin{pmatrix} C^* A C & 0 \\ 0 & DAD \end{pmatrix} \\ &\leq \frac{1}{2} \sum_{\epsilon = \pm} f(\widehat{U}_\epsilon^* \widehat{A}\, \widehat{U}_\epsilon) \\ &\leq \frac{1}{2} \sum_{\epsilon = \pm} \widehat{U}_\epsilon^* \begin{pmatrix} f(A) & 0 \\ 0 & 0 \end{pmatrix} \widehat{U}_\epsilon = \begin{pmatrix} C^* f(A) C & 0 \\ 0 & D f(A) D \end{pmatrix} . \end{aligned} \tag{A7}$$

This implies in particular the bound (A4). Conversely, it is shown in Ref. 69 that if this bound is satisfied for any orthogonal projection C and any $A \in \mathcal{B}(\mathcal{H})_+$, then f is operator convex and $f(0) \leq 0$.

Let $\mathcal{M}$ be a quantum operation on $\mathcal{B}(\mathcal{H})$ and $f : \mathbb{R}_+ \to \mathbb{R}$ be operator convex. Then the following Jensen-type inequality holds:[51]

$$f(\mathcal{M}^*(A)) \leq \mathcal{M}^*(f(A)), \quad A \in \mathcal{B}(\mathcal{H})_+ . \tag{A8}$$

A simple justification of this inequality is as follows. Since $\mathcal{M}^*(c\, 1) = c\, 1$ for any constant $c \in \mathbb{R}$, one may assume without loss of generality that $f(0) = 0$. Let $A \in \mathcal{B}(\mathcal{H})_+$. According to Stinespring's theorem (Sec. III) one can find a unitary operator U on an enlarged space $\mathcal{H} \otimes \mathcal{H}_\mathsf{E}$ and a vector $|\epsilon_0\rangle \in \mathcal{H}_\mathsf{E}$ such that $\mathcal{M}^*(A) = \langle \epsilon_0 | U^* A \otimes 1 U | \epsilon_0 \rangle$. Let us set $P_0 = |\epsilon_0\rangle\langle\epsilon_0|$. Applying (A4) with $C = 1 \otimes P_0$, one gets

$$\begin{aligned} f(\mathcal{M}^*(A)) \otimes P_0 &= f(1 \otimes P_0\, U^* A \otimes 1 U\, 1 \otimes P_0) \\ &\leq 1 \otimes P_0 f(U^* A \otimes 1 U) 1 \otimes P_0 = \mathcal{M}^*(f(A)) \otimes P_0 \ . \end{aligned} \tag{A9}$$

APPENDIX B: TRACE INEQUALITIES

In this appendix some inequalities involving the $\|\cdot\|_p$-norms are stated or derived.

1. Let us first recall the triangle and "inverse triangle" inequalities: for any matrices A and B one has

$$\|A+B\|_p \begin{cases} \le \|A\|_p + \|B\|_p & \text{if } p \ge 1 \\ \ge \|A\|_p + \|B\|_p & \text{if } 0 < p < 1. \end{cases} \tag{B1}$$

This shows that the map $A \mapsto \|A\|_p$ defined by (2) is a norm for $p \ge 1$, but this is not the case for $p < 1$. One deduces the bound

$$\mathrm{tr}[\sqrt{|A|^2 + |B|^2}] \le \mathrm{tr}\,|A| + \mathrm{tr}\,|B| \tag{B2}$$

by applying (B1) for $p = 1$ to the matrices

$$\widehat{A} = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}, \quad \widehat{B} = \begin{pmatrix} 0 & 0 \\ B & 0 \end{pmatrix}.$$

2. Another standard result is the Lieb-Thirring inequality.[100] We quote here without proof a generalization of this inequality derived by Araki.[12] Let $k > 0$ and A and B be non-negative operators. If $\alpha \ge 1$ then

$$\left\| B^{\frac{1}{2}} A B^{\frac{1}{2}} \right\|_{\alpha k}^{\alpha} \le \left\| B^{\frac{\alpha}{2}} A^{\alpha} B^{\frac{\alpha}{2}} \right\|_k . \tag{B3}$$

Taking $\alpha \to \alpha^{-1}$ and $k \to k/\alpha$, one can deduce that the reverse inequality holds true if $0 \le \alpha \le 1$.

3. Next, let us show that for any square matrices A, B, C, and D of the same size, the following bound generalizing the Cauchy-Schwarz inequality $\|AB\|_1 \le \|A\|_2 \|B\|_2$ holds true[112]

$$\|AB + CD\|_1^2 \le \left(\|A\|_2^2 + \|D\|_2^2\right)\left(\|B\|_2^2 + \|C\|_2^2\right) . \tag{B4}$$

Actually, let us form the 2×2 block matrices

$$\widehat{E} = \begin{pmatrix} A^* & 0 \\ C^* & 0 \end{pmatrix}, \quad \widehat{F} = \begin{pmatrix} B & 0 \\ D & 0 \end{pmatrix}.$$

Then

$$\|AB + CD\|_1^2 = \left\|\widehat{E}^* \widehat{F}\right\|_1^2 \le \|\widehat{E}\|_2^2 \|\widehat{F}\|_2^2 = \left(\|A\|_2^2 + \|C\|_2^2\right)\left(\|B\|_2^2 + \|D\|_2^2\right) .$$

But $CD = UD^*C^*U$ with U unitary by the polar decomposition. Applying the above inequality with C and D replaced by UD^* and C^*U and using the unitary invariance of $\|\cdot\|_2$, one gets the desired result (B4).

4. Let $B = (B_{ij})_{i,j=1}^m$ be a non-negative $m \times m$ operator-valued matrix, whose entries B_{ij} are given by $p_i \times p_j$ matrices. Denote by $A = \sqrt{B} = (A_{ij})_{i,j=1}^m$ the square root of B. Then for any $j = 1, \ldots, m$, one has[19]

$$\sum_{i, i \ne j} \left\| A_{ij} \right\|_2^2 \le \frac{1}{2} \sum_{i, i \ne j} \left\| B_{ij} \right\|_1 . \tag{B5}$$

Let us first establish (B5) for $m = 2$. Thanks to the singular value decomposition and the unitary invariance of the $\|\cdot\|_p$-norms, we may assume without loss of generality that A_{12} is a diagonal $p_1 \times p_2$ matrix, i.e., $A_{12} = \sum_{k=1}^p \sqrt{\nu_k} |k\rangle\langle k|$ with $p = \min\{p_1, p_2\}$. By a standard argument, the non-negativity of A implies

$$|\langle \varphi_1 | A_{12} | \varphi_2 \rangle|^2 \le \langle \varphi_1 | A_{11} | \varphi_1 \rangle \langle \varphi_2 | A_{22} | \varphi_2 \rangle$$

for any vectors $|\varphi_1\rangle \in \mathbb{C}^{p_1}$ and $|\varphi_2\rangle \in \mathbb{C}^{p_2}$. Using this bound and the relation $B_{12} = A_{11}A_{12} + A_{12}A_{22}$, we find

$$\|A_{12}\|_2^2 = \sum_{k=1}^p \nu_k \le \sum_{k=1}^p \sqrt{\nu_k \langle k | A_{11} | k \rangle \langle k | A_{22} | k \rangle} \le \frac{1}{2} \sum_{k=1}^p \sqrt{\nu_k} \left(\langle k | A_{11} | k \rangle + \langle k | A_{22} | k \rangle \right) = \frac{1}{2} \|B_{12}\|_1 .$$

Consider now the general case $m \geq 2$. The idea is to write B as a 2×2 block matrix such that the upper left and lower right blocks are the $(m-1) \times (m-1)$ matrix $(B_{ij})_{i,j=1}^{m-1}$ and the single entry B_{mm}, respectively, whereas the upper right (lower left) block forms a column (line) vector with entries B_{im} (B_{mi}). A similar block decomposition can be made for A. Applying the foregoing result for $m = 2$, one gets

$$\sum_{i,i\neq m} \|A_{im}\|_2^2 = \left\| \begin{pmatrix} A_{1m} \\ \vdots \\ A_{(m-1)m} \end{pmatrix} \right\|_2^2 \leq \frac{1}{2} \left\| \begin{pmatrix} B_{1m} \\ \vdots \\ B_{(m-1)m} \end{pmatrix} \right\|_1 = \frac{1}{2} \left\| \sqrt{\sum_{i,i\neq m} |B_{im}|^2} \right\|_1 \leq \frac{1}{2} \sum_{i,i\neq m} \|B_{im}\|_1 ,$$

where we have used (B2) in the last bound. This proves (B5) for $j = m$. By an appropriate unitary conjugation, one deduces that the bound holds for any j.

5. The following trace inequality plays a central role in the derivation of the quantum Chernoff bound:[7] for any positive square matrices $A > 0$ and $B > 0$ and any $0 \leq s \leq 1$,

$$\frac{1}{2}\Big(\mathrm{tr}(A) + \mathrm{tr}(B) - \mathrm{tr}\,|A - B|\Big) \leq \mathrm{tr}(A^{1-s}B^s) . \tag{B6}$$

This inequality was first shown in Ref. 7, but the proof in this reference is not very transparent. We present here a much simpler proof due to Ozawa, which has been first reported in Ref. 87. Denoting by $O_\pm = (|O| \pm O)/2 \geq 0$ the positive and negative parts of O, one may express $\mathrm{tr}\,|A - B|$ as $2\,\mathrm{tr}(A - B)_+ - \mathrm{tr}(A) + \mathrm{tr}(B)$. Thus (B6) is equivalent to

$$\mathrm{tr}\Big((A^s - B^s)A^{1-s}\Big) \leq \mathrm{tr}(A - B)_+ .$$

Since $f(x) = x^s$ is operator monotone (see Appendix A) and $A \leq A + (A - B)_- = B + (A - B)_+$, one has $A^s \leq (B + (A - B)_+)^s$. Hence

$$\begin{aligned} \mathrm{tr}\Big((A^s - B^s)A^{1-s}\Big) &\leq \mathrm{tr}\Big(\big[(B + (A - B)_+)^s - B^s\big]A^{1-s}\Big) \\ &\leq \mathrm{tr}\Big(\big[(B + (A - B)_+)^s - B^s\big](B + (A - B)_+)^{1-s}\Big), \end{aligned}$$

where the second inequality relies on the similar bound $B^s \leq (B + (A - B)_+)^s$. By rearranging the product in the last trace and using the latter bound with $s \leftrightarrow (1 - s)$, one gets

$$\mathrm{tr}\Big((A^s - B^s)A^{1-s}\Big) \leq \mathrm{tr}(B) + \mathrm{tr}(A - B)_+ - \mathrm{tr}\Big(B^s(B + (A - B)_+)^{1-s}\Big) \leq \mathrm{tr}(A - B)_+.$$

This concludes the justification of (B6).

[1] B. Aaronson, R. L. Franco, and G. Adesso, "Comparative investigation of the freezing phenomena for quantum correlations under nondissipative decoherence," Phys. Rev. A **88**, 012120 (2013).
[2] A. E. Allahverdyan, R. Balian, and T. M. Nieuwenhuizen, "Quantum measurement as a driven phase transition: An exactly solvable model," Phys. Rev. A **64**, 032108 (2001).
[3] A. E. Allahverdyan, R. Balian, and T. M. Nieuwenhuizen, "Curie-Weiss model of the quantum measurement process," Europhys. Lett. **61**, 452–458 (2003).
[4] A. E. Allahverdyan, R. Balian, and T. M. Nieuwenhuizen, "Understanding quantum measurement from the solution of dynamical models," Phys. Rep. **525**, 1–166 (2013).
[5] P. M. Alberti, "A note on the transition-probability over C^*-algebras," Lett. Math. Phys. **7**, 25–32 (1983).
[6] M. Ali, A. R. P. Rau, and G. Alber, "Quantum discord for two-qubit X states," Phys. Rev. A **81**, 042105 (2010).
[7] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. I. Masanes, A. Acin, and F. Verstraete, "Discriminating States: The Quantum Chernoff Bound," Phys. Rev. Lett. **98**, 160501 (2007).
[8] T. Ando, "Convexity of certain maps on positive definite matrices and applications to Hadamard products," Lin. Alg. Appl. **26**, 203–241 (1979).
[9] H. Araki, "A remark on Bures distance function for normal states," Publ. RIMS Kyoto Univ. **6**, 477–482 (1970).
[10] H. Araki, "Relative entropy for states of von Neumann algebras," Publ. RIMS Kyoto Univ. **11**, 809–833 (1976).
[11] H. Araki and T. Masudai, "Positive cones and L_p-spaces for von Neumann algebras," Publ. RIMS Kyoto Univ. **18**, 339–411 (1982).
[12] H. Araki, "On an Inequality of Lieb and Thirring," Lett. Math. Phys. **19**, 167–170 (1990).
[13] R. Balian, Y. Alhassid, and H. Reinhardt, "Dissipation in many-body systems: a geometric approach based on information theory," Phys. Rep. **131**, 1 (1986).
[14] R. Balian, *From Microphysics to Macrophysics: Methods and Applications of Statistical Physics* (Springer, 2007), Vol. 1.

[15] M. Ban, K. Kurokawa, R. Momose, and O. Hirota, "Optimum measurements for discrimination among symmetric quantum states and parameter estimation," Int. J. Theor. Phys. **36**, 1269–1288 (1997).
[16] S. M. Barnett, "Minimum error discrimination between multiply symmetric states," Phys. Rev. A **64**, 030303 (2001).
[17] S. M. Barnett, A. Chefles, and I. Jex, "Comparison of two unknown pure quantum states," Phys. Lett. A **307**, 189–195 (2003).
[18] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, "Noncommuting mixed states cannot be broadcast," Phys. Rev. Lett. **76**, 2818 (1996).
[19] H. Barnum and E. Knill, "Reversing quantum dynamics with near-optimal quantum and classical fidelity," J. Math. Phys. **43**, 2097–2106 (2002).
[20] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, 2006).
[21] C. H. Bennett, "Quantum cryptography using any two nonorthogonal states," Phys. Rev. Lett. **68**, 3121 (1992).
[22] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, "Concentrating partial entanglement by local operations," Phys. Rev. A **53**, 2046 (1996).
[23] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, "Mixed-state entanglement and quantum error correction," Phys. Rev. A **54**, 3824 (1996).
[24] C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters, "Quantum nonlocality without entanglement," Phys. Rev. A **59**, 1070 (1999).
[25] J. A. Bergou, "Discrimination of quantum states," J. Mod. Opt. **57**(3), 160–180 (2010).
[26] J. A. Bergou, U. Herzog, and M. Hillery, "Discrimination of quantum states," in *Quantum State Estimation*, Lecture Notes in Physics Vol. 649, edited by M. Paris and J. Rehacek (Springer, Berlin, 2004), pp. 417–465.
[27] R. Bhatia, *Matrix Analysis* (Springer, 1991).
[28] *The Physics of Quantum Information: Quantum Cryptography, Quantum Teleportation, Quantum Computation*, edited by D. Bouwmeester, A. Ekert, and A. Zeilinger (Springer, 2000).
[29] O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics* (Springer, Berlin, 1997), Vols. 1 and 2.
[30] S. L. Braunstein and C. M. Caves, "Statistical distance and the geometry of quantum states," Phys. Rev. Lett. **72**, 3439–3443 (1994).
[31] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, 2002).
[32] D. Bures, "An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite w^*-algebras," Trans. Am. Math. Soc. **135**, 199–212 (1969).
[33] E. A. Carlen, "Trace inequalities and quantum entropy: An introductory course," in *Entropy and the quantum*, Contemporary Mathematics, edited by R. Sims and D. Ueltschi, Vol. 529 (American Mathematical Society, Providence, RI, 2010).
[34] D. Cavalcanti, L. Aolita, S. Boixo, K. Modi, M. Piani, and A. Winter, "Operational interpretations of quantum discord," Phys. Rev. A **83**, 032324 (2011).
[35] N. N. Cencov, *Statistical Decision Rules and Optimal Interferences*, Translations of Mathematical Monographs, Vol. 53 (American Mathematical Society, Providence, 1982).
[36] N. J. Cerf and C. Adami, "Negative entropy and information in quantum mechanics," Phys. Rev. Lett. **79**, 5194–5197 (1997).
[37] A. Chefles, "Unambiguous discrimination between linearly independent quantum states," Phys. Lett. A **239**, 339–347 (1998).
[38] A. Chefles, "Quantum state discrimination," Contemp. Phys. **41**, 401–424 (2000).
[39] K. Chen and L.-A. Wu, "A matrix realignment method for recognizing entanglement," Quantum Inf. Comput. **3**, 193–202 (2003).
[40] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," Ann. Math. Stat. **23**, 493–507 (1952).
[41] C.-L. Chou and L. Y. Hsu, "Minimal-error discrimination between symmetric mixed quantum states," Phys. Rev. A **68**, 042305 (2003).
[42] R. B. M. Clarke, V. M. Kendon, A. Chefles, S. M. Barnett, E. Riis, and M. Sasaki, "Experimental realization of optimal detection strategies for overcomplete states," Phys. Rev. A **64**, 012303 (2001).
[43] F. Ciccarello, T. Tufarelli, and V. Giovannetti, "Towards computability of trace distance discord," New J. Phys. **16**, 013038 (2014).
[44] V. Coffman, J. Kundu, and W. K. Wootters, "Distributed entanglement," Phys. Rev. A **61**, 052306 (2000).
[45] G. M. D'Ariano, P. Lo Presti, and P. Perinotti, "Classical randomness in quantum measurements," J. Phys. A: Math. Gen. **38**, 5979–5991 (2005).
[46] B. Dakić, V. Vedral, and C. Brukner, "Necessary and sufficient condition for nonzero quantum discord," Phys. Rev. Lett. **105**, 190502 (2010).
[47] B. Dakić *et al.*, "Quantum discord as resource for remote state preparation," Nat. Phys. **8**, 666–670 (2012).
[48] A. Datta, S. T. Flammia, and C. M. Caves, "Entanglement and the power of one qubit," Phys. Rev. A **72**, 042316 (2005).
[49] A. Datta, A. Shaji, and C. M. Caves, "Quantum discord and the power of one qubit," Phys. Rev. Lett. **100**, 050502 (2008).
[50] E. B. Davies, "Information and quantum measurement," IEEE Trans. Inf. Theory **24**, 596–599 (1978).
[51] C. Davis, "A Schwarz inequality for convex operator functions," Proc. Am. Math. Soc. **8**, 42–44 (1957).
[52] D. Dieks, "Overlap and distinguishability of quantum states," Phys. Lett. A **126**, 303–306 (1988).
[53] R. Durrett, *Probability: Theory and Examples*, 2nd ed. (Duxbury Press, USA, 1996).

[54] Y. C. Eldar and G. D. Forney, Jr., "On quantum detection and the square-root measurement," IEEE Trans. Inf. Theory **47**, 858–872 (2001).
[55] Y. C. Eldar, A. Megretski, and G. C. Verghese, "Designing optimal quantum detectors via semidefinite programming," IEEE Trans. Inf. Theory **49**, 1007–1012 (2003).
[56] Y. C. Eldar, "von Neumann measurement is optimal for detecting linearly independent mixed quantum states," Phys. Rev. A **68**, 052303 (2003).
[57] Y. C. Eldar, A. Megretski, and G. C. Verghese, "Optimal detection of symmetric mixed quantum states," IEEE Trans. Inf. Theory **50**, 1198 (2004).
[58] F. F. Fanchini, M. F. Cornelio, M. C. de Oliveira, and A. O. Caldeira, "Conservation law for distributed entanglement of formation and quantum discord," Phys. Rev. A **84**, 012313 (2011).
[59] Y. Feng, R. Duan, and M. Ying, "Unambiguous discrimination between mixed quantum states," Phys. Rev. A **70**, 012308 (2004).
[60] R. L. Frank and E. H. Lieb, "Monotonicity of a relative Rényi entropy," J. Math. Phys. **54**, 122201 (2013).
[61] A. Fujiwara, "Quantum channel identification problem," Phys. Rev. A **63**, 042304 (2001).
[62] F. Galve, G. L. Giorgi, and Z. Zambrini, "Orthogonal measurements are almost sufficient for quantum discord of two qubits," Eur. Phys. Lett. **96**, 40005 (2011).
[63] V. Giovannetti, S. Lloyd, and L. Maccone, "Quantum metrology," Phys. Rev. Lett. **96**, 010401 (2006).
[64] D. Giulini *et al.*, *Decoherence and the Appearance of a Classical World in Quantum Theory* (Springer, 1996).
[65] C. Gross, T. Zibold, E. Nicklas, J. Estève, and M. K. Oberthaler, "Nonlinear atom interferometer surpasses classical precision limit," Nature **464**, 1165 (2010).
[66] M. Gu, H. M. Chrzanowski, S. M. Assad, T. Symul, K. Modi, T. C. Ralph, V. Vedral, and P. Koy Lam, "Observing the operational significance of discord consumption," Nat. Phys. **8**, 671 (2012).
[67] O. Gühne and G. Tóth, "Entanglement detection," Phys. Rep. **474**, 1 (2009).
[68] S. Hamieh, R. Kobes, and H. Zaraket, "Positive-operator-valued measure optimization of classical correlations," Phys. Rev. A **70**, 052325 (2004).
[69] F. Hansen and G. K. Pedersen, "Jensen's inequality for operator and Löwner's theorem," Math. Ann. **258**, 229–241 (1982).
[70] S. Haroche and J.-M. Raimond, *Exploring the Quantum: Atoms, Cavities and Photons* (Oxford University Press, 2006).
[71] P. Hausladen and W. K. Wootters, "A "pretty good" measurement for distinguishing quantum states," J. Mod. Opt. **41**, 2385–2390 (1994).
[72] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, "Classical information capacity of a quantum channel," Phys. Rev. A **54**, 1869–1876 (1996).
[73] P. Hayden, R. Jozsa, D. Petz, and A. Winter, "Structure of states which satisfy strong subadditivity of quantum entropy with equality," Commun. Math. Phys. **246**, 359–374 (2004).
[74] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
[75] L. Henderson and V. Vedral, "Classical, quantum and total correlations," J. Phys. A: Math. Gen. **34**, 6899–6905 (2001).
[76] F. Hiai and D. Petz, "The proper formula for relative entropy and its asymptotics in quantum probability," Commun. Math. Phys. **143**, 99–114 (1991).
[77] A. S. Holevo, "Statistical decisions in quantum theory," J. Multivar. Anal. **3**, 337–394 (1973).
[78] A. S. Holevo, "On asymptotically optimal hypothesis testing in quantum statistics," Theory Probab. Appl. **23**, 411–415 (1979).
[79] R. Horodecki and M. Horodecki, "Information-theoretic aspects of inseparability of mixed states," Phys. Rev. A **54**, 1838–1843 (1996).
[80] M. Horodecki, P. Horodecki, and R. Horodecki, "Separability of mixed states: necessary and sufficient conditions," Phys. Lett. A **223**, 1–8 (1996).
[81] P. Horodecki, "Separability criterion and inseparable mixed states with positive partial transposition," Phys. Lett. A **232**, 333–339 (1997).
[82] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, "Quantum entanglement," Rev. Mod. Phys. **81**, 865–942 (2009).
[83] S. Huang, "Square-root measurement for pure states," Phys. Rev. A **72**, 022324 (2005).
[84] M. Hübner, "Explicit computation of the Bures distance for density matrices," Phys. Lett. A **163**, 239–242 (1992).
[85] I. D. Ivanovic, "How to differentiate between non-orthogonal states," Phys. Lett. A **123**, 257–259 (1987).
[86] G. Jaeger and A. Shimony, "Optimal distinction between two non-orthogonal quantum states," Phys. Lett. A **197**, 83–87 (1995).
[87] V. Jakšić, Y. Ogata, C.-A. Pillet, and R. Seiringer, "Quantum hypothesis testing and non-equilibrium statistical mechanics," Rev. Math. Phys. **24**(6), 1230002 (2012).
[88] V. Jakšić and C.-A. Pillet, "Entropic functionals in quantum statistical mechanics," in *Proceedings of XVIIth International Congress of Mathematical Physics, Aalborg, 2012* (World Scientific, Singapore, 2013), pp. 336–343.
[89] V. Jakšić, C.-A. Pillet, and M. Westrich, "Entropic fluctuations of quantum dynamical semigroups," J. Stat. Phys. **154**, 153–187 (2014).
[90] D. Jonathan and M. B. Plenio, "Entanglement-assisted local manipulation of pure quantum states," Phys. Rev. Lett. **83**, 3566–3569 (1999).
[91] R. Jozsa, "Fidelity for mixed quantum states," J. Mod. Opt. **41**, 2315–2323 (1994).
[92] R. Jozsa and N. Linden, "On the role of entanglement in quantum-computational speed-up," Proc. R. Soc. London, Ser. A **459**, 2011–2032 (2003).
[93] E. Knill and R. Laflamme, "Power of one bit of quantum information," Phys. Rev. Lett. **81**, 5672 (1998).
[94] M. Koashi and A. Winter, "Monogamy of quantum entanglement and other correlations," Phys. Rev. A **69**, 022309 (2004).

[95] K. Kraus, "General state changes in quantum theory," Ann. Phys. **64**, 311–335 (1971).
[96] F. Kubo and T. Ando, "Means of positive linear operators," Math. Ann. **246**, 205–224 (1980).
[97] B. P. Lanyon, M. Barbieri, M. P. Almeida, and A. G. White, "Experimental quantum computing without entanglement," Phys. Rev. Lett. **101**, 200501 (2008).
[98] E. H. Lieb, "Convex trace functions and the Wigner-Yanase-Dyson conjecture," Adv. Math. **11**, 267–288 (1973).
[99] E. H. Lieb and M. B. Ruskai, "Proof of the strong subadditivity of quantum-mechanical entropy," J. Math. Phys. **14**, 1938–1941 (1973).
[100] E. H. Lieb and W. Thirring, "Inequalities for the Moments of the Eigenvalues of the Schrödinger Hamiltonian and their Relation to Sobolev Inequalities," in Studies in Mathematical Physics: Essays in Honor of Valentine Bargman, edited by E. H. Lieb, B. Simon, A. S. Wightman (Princeton University Press, Princeton, 1976), pp. 269–297.
[101] G. Lindblad, "Expectations and entropy inequalities for finite quantum systems," Commun. Math. Phys. **39**, 111–119 (1974).
[102] G. Lindblad, "Completely positive maps and entropy inequalities," Commun. Math. Phys. **40**, 147–151 (1975).
[103] S. Luo, "Using measurement-induced disturbance to characterize correlations as classical or quantum," Phys. Rev. A **77**, 022301 (2008).
[104] S. Luo, "Quantum discord for two-qubit systems," Phys. Rev. A **77**, 042303 (2008).
[105] S. Luo and S. Fu, "Geometric measure of quantum discord," Phys. Rev. A **82**, 034302 (2010).
[106] V. Madhok and A. Datta, "Role of quantum discord in quantum communication," e-print arXiv:1107.0994 [quant-ph].
[107] V. Madhok and A. Datta, "Interpreting quantum discord through quantum state merging," Phys. Rev. A **83**, 032323 (2011).
[108] L. Mazzola, J. Piilo, and S. Maniscalco, "Sudden transition between classical and quantum decoherence," Phys. Rev. Lett. **104**, 200401 (2010).
[109] K. Modi, T. Parerek, W. Son, V. Vedral, and M. Williamson, "Unified view of quantum and classical correlations," Phys. Rev. Lett. **104**, 080501 (2010).
[110] K. Modi, A. Brodutch, H. Cable, T. Paterek, and V. Vedral, "The classical-quantum boundary for correlations: Discord and related measures," Rev. Mod. Phys. **84**, 1655–1707 (2012).
[111] M. Mohseni, A. M. Steinberg, and J. A. Bergou, "Optical realization of optimal unambiguous discrimination for pure and mixed quantum states," Phys. Rev. Lett. **93**, 200403 (2004).
[112] A. Montanaro, "A lower bound on the probability of error in quantum state discrimination," in *Proceedings of IEEE Information Theory Workshop ITW'08* (IEEE, Piscataway, NJ, 2008), p. 378.
[113] E. A. Morozova and N. N. Chentsov, "Markov invariant geometry on state manifolds (in Russian)," Itogi Nauki i Tekhniki **36**, 69–102 (1990).
[114] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, "On quantum Rényi entropies: a new generalization and some properties," J. Math. Phys. **54**, 122203 (2013).
[115] T. Nakano, M. Piani, and G. Adesso, "Negativity of quantumness and its interpretations," Phys. Rev. A **88**, 012117 (2013).
[116] M. A. Nielsen, "Conditions for a class of entanglement transformations," Phys. Rev. Lett. **83**, 436–439 (1999).
[117] N. A. Nielsen and I. L. Chuang, *Quantum Computation and Information* (Cambridge University Press, 2000).
[118] M. Nussbaum and A. Szkola, "The Chernoff lower bound for symmetric quantum hypothesis testing," *The Annals of Statistics* (Institute of Mathematical Statistics, 2009), Vol. 37, pp. 1040–1057.
[119] M. Ohya and D. Petz, *Quantum Entropy and its Use* (Springer-Verlag, Berlin/Heidelberg, 1993).
[120] H. Ollivier and W. H. Zurek, "Quantum discord: A measure of the quantumness of correlations," Phys. Rev. Lett. **88**, 017901 (2001).
[121] T. Ogawa and H. Nagaoka, "Strong converse and Stein's lemma in quantum hypothesis testing," IEEE Trans. Inf. Theory **46**(7), 2428–2433 (2000).
[122] M. Ozawa, "Entanglement measures and the Hilbert-Schmidt distance," Phys. Lett. A **268**, 158–160 (2000).
[123] G. Passante, O. Moussa, D. A. Trottier, and R. Laflamme, "Experimental detection of nonclassical correlations in mixed-state quantum computation," Phys. Rev. A **84**, 044302 (2011).
[124] A. Peres, "How to differentiate between non-orthogonal states," Phys. Lett. A **128**, 19 (1988).
[125] A. Peres, "Separability criterion for density matrices," Phys. Rev. Lett. **77**, 1413–1415 (1996).
[126] A. Peres, "Neumark's theorem and quantum inseparability," Found. Phys. **20**, 1441–1453 (1990).
[127] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer Academic Publisher, 1995).
[128] D. Petz, "Monotone metrics on matrix spaces," Lin. Alg. Appl. **244**, 81–96 (1996).
[129] D. Petz, "Monotonicity of quantum relative entropy revisited," Rev. Math. Phys. **15**, 79–91 (2003).
[130] L. Pezzé and A. Smerzi, "Entanglement, nonlinear dynamics, and the Heisenberg limit," Phys. Rev. Lett. **102**, 100401 (2009).
[131] M. Piani, "Problem with geometric discord," Phys. Rev. A **86**, 034101 (2012).
[132] D. Qiu and L. Li, "Minimum-error discrimination of quantum states: Bounds and comparison," Phys. Rev. A **81**, 042329 (2010).
[133] F. Riedel, P. Böhi, Y. Li, T. W. Hänsch, A. Sinatra, and P. Treutlein, "Atom-chip-based generation of entanglement for quantum metrology," Nature **464**, 1170 (2010).
[134] W. Roga, S. M. Giampaolo, and F. Illuminati, "Discord of response," e-print arXiv:1401.8243 [quant-ph].
[135] T. Rudolph, R. W. Spekkens, and P. S. Turner, "Unambiguous discrimination of mixed states," Phys. Rev. A **68**, 010301 (2003).
[136] M. B. Ruskai, "Beyond strong subadditivity: improved bounds on the contraction of the generalized relative entropy," Rev. Math. Phys. **6**(5a), 1147–1161 (1994).
[137] A. Sanpera, R. Tarrach, and G. Vidal, "Quantum inseparability as local pseudomixture," Phys. Rev. A **58**, 826–830 (1998).

[138] M. Sasaki, K. Kato, M. Izutsu, and O. Hirota, "Quantum channels showing superadditivity in classical capacity," Phys. Rev. A **58**, 146–158 (1998).
[139] A. Sawicki, A. Huckleberry, and M. Kuś, "Symplectic geometry of entanglement," Commun. Math. Phys. **305**, 441–468 (2011).
[140] M. H. Schleier-Smith, I. D. Leroux, and V. Vuletić, "States of an ensemble of two-level atoms with reduced quantum uncertainty," Phys. Rev. Lett. **104**, 073604 (2010).
[141] B. W. Schumacher, "Sending entanglement through noisy quantum channels," Phys. Rev. A **54**, 2614–2628 (1996).
[142] B. W. Schumacher and M. A. Nielsen, "Quantum data processing and error correction," Phys. Rev. A **54**, 2629 (1996).
[143] C. E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423, 623–656 (1948).
[144] A. Shimony, "Degree of entanglement," Ann. N. Y. Acad. Sci. **755**, 675–679 (1995).
[145] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," SIAM J. Comput. **26**, 1484–1509 (1997).
[146] H.-J. Sommers and K. Życzkowski, "Bures volume of the set of mixed quantum states," J. Phys. A: Math. Gen. **36**, 10083–10100 (2003).
[147] D. Spehner and F. Haake, "Decoherence bypass of macroscopic superpositions in quantum measurement," J. Phys. A: Math. Theor. **41**, 072002 (2008).
[148] D. Spehner and F. Haake, "Quantum measurements without macroscopic superpositions," Phys. Rev. A **77**, 052114 (2008).
[149] D. Spehner and M. Orszag, "Geometric quantum discord with Bures distance," New J. Phys. **15**, 103001 (2013).
[150] D. Spehner and M. Orszag, "Geometric quantum discord with Bures distance: the qubit case," J. Phys. A: Math. Theor. **47**, 035302 (2014).
[151] W. F. Stinespring, "Positive functions on C^*-algebras," Proc. Am. Soc. **6**, 211–216 (1955).
[152] A. Streltsov, H. Kampermann, and D. Bruß, "Linking a distance measure of entanglement to its convex roof," New J. Phys. **12**, 123004 (2010).
[153] A. Uhlmann, "Endlich-dimensionale Dichtematrizen II," Wiss. Z. Karl-Marx-Univ. Leipzig, Math.-Nat R. **22**, 139–177 (1973).
[154] A. Uhlmann, "The "transition probability" in the state space of a *-algebra," Rep. Math. Phys. **9**, 273–279 (1976).
[155] A. Uhlmann, "Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory," Commun. Math. Phys. **54**, 21–32 (1977).
[156] A. Uhlmann, "Parallel transport and "quantum holonomy" along density operators," Rep. Math. Phys. **24**, 229–240 (1986).
[157] A. Uhlmann, "Geometric phases and related structures," Rep. Math. Phys. **36**, 461–481 (1995).
[158] U. Umegaki, "Conditional expectations in an operator algebra IV (entropy and information)," Kodai Math. Sem. Rep. **14**, 59–85 (1962).
[159] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, "Quantifying entanglement," Phys. Rev. Lett. **78**, 2275–2279 (1997).
[160] V. Vedral and M. B. Plenio, "Entanglement measures and purifications procedures," Phys. Rev. A **57**, 1619–1633 (1998).
[161] G. Vidal, "Entanglement monotones," J. Mod. Opt. **47**, 355–376 (2000).
[162] S. Vogelsberger, "Dynamique des systèmes quantiques ouverts: décohérence et perte d'intrication (in French)," Ph.D. thesis (University Joseph Fourier, Grenoble, 2012).
[163] W. Wasilewski, K. Jensen, H. Krauter, J. J. Renema, M. V. Balabas, and E. S. Polzik, "Quantum noise limited and entanglement-assisted magnetometry," Phys. Rev. Lett. **104**, 133601 (2010).
[164] T. C. Wei and P. M. Goldbart, "Geometric measure of entanglement and applications to bipartite and multipartite quantum states," Phys. Rev. A **68**, 042307 (2003).
[165] R. F. Werner, "Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model," Phys. Rev. A **40**, 4277–4281 (1989).
[166] M. M. Wilde, A. Winter, and D. Yang, "Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Renyi relative entropy," e-print arXiv:1306.1586 [quant-ph].
[167] M. M. Wolf, "Quantum channels and operations guided tour (2002)," see http://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf.
[168] W. K. Wootters, "Statistical distance and Hilbert space," Phys. Rev. D **23**, 357–362 (1981).
[169] W. K. Wootters, "Entanglement of formation of an arbitrary state of two qubits," Phys. Rev. Lett. **80**, 2245 (1998).
[170] S. L. Woronowicz, "Positive maps of low dimensional matrix algebras," Rep. Math. Phys. **10**, 165–183 (1976).
[171] H. P. Yuen, R. S. Kennedy, and M. Lax, "Optimum testing of multiple hypotheses in quantum detection theory," IEEE Trans. Inf. Theory **21**, 125–134 (1975).
[172] B. Yurke, S. L. McCall, and J. R. Klauder, "SU(2) and SU(1,1) interferometers," Phys. Rev. A **33**, 4033–4054 (1986).