Entropy distance: New quantum phenomena

Stephan Weis and Andreas Knauf

Related Articles

Additional information on J. Math. Phys.

Entropy distance: New quantum phenomena

Stephan Weis[1,a)] and Andreas Knauf[2,b)]

[1]*Max Planck Institute for Mathematics in the Sciences, Inselstr. 22,
D-04103 Leipzig, Germany*
[2]*Department of Mathematics, Friedrich-Alexander-University Erlangen-Nuremberg,
Cauerstr. 11, D-91058 Erlangen, Germany*

We study a curve of Gibbsian families of complex 3×3-matrices and point out new features, absent in commutative finite-dimensional algebras: a discontinuous maximum-entropy inference, a discontinuous entropy distance, and non-exposed faces of the mean value set. We analyze these problems from various aspects including convex geometry, topology, and information geometry. This research is motivated by a theory of infomax principles, where we contribute by computing first order optimality conditions of the entropy distance. © *2012 American Institute of Physics.*
[http://dx.doi.org/10.1063/1.4757652]

I. INTRODUCTION

The aim of the introduction is a discussion of the maximum-entropy inference under linear constraints, in two aspects: The problem of its discontinuity and its connection to infomax principles, asking for maximization of the entropy distance from an exponential family. Section I B gives an overview of the article.

A. Maximum-entropy inference and infomax principles

The *maximum-entropy principle*, dating back to Clausius and Boltzmann, became the information theoretic justification of the thermodynamic formalism in the work of Jaynes.[21] A comprehensive account of this development is included, e.g., in Ref. 12, which supports a Bayesian point of view of physics. Under a more restrictive frequentist point of view, let us add that the *Shannon entropy* and *von Neumann entropy*, quantifying the uncertainty in the maximum-entropy principle, are also characterized operationally by coding theorems: These entropies quantify the amount of physical resources needed on the average to encode the result of a repeatable experiment on the long run. This was first proved for independent identically distributed random variables in Ref. 39 and for pure quantum states in Ref. 38, see also Ref. 11 for generalizations.

We have discovered in three-level quantum systems a problem that can arise for non-commutative observables: The real analytic maximum-entropy inference under linear constraints has no continuous extension. An example is given in Remark 22; this phenomenon does not appear in commutative algebras of finite dimensions. The surprise of a discontinuity shows that we have currently no quantitative understanding of the maximum-entropy inference at all. Other branches of quantum statistics are much further developed, e.g., in hypothesis testing[4,29] it is known which observations have the fastest asymptotic exponential decay of error probability. Optimal tests are known in state tomography with respect to natural measures of efficiency[33,46] and bounds on the number of observations have been found for important special classes of quantum systems.[9,18]

What do we mean by a discontinuous maximum-entropy inference? We use a fixed set of *observables* $a_1, \ldots, a_k$, i.e. self-adjoint matrices in the algebra $\mathcal{A} = \mathrm{Mat}(N, \mathbb{C})$, and denote by $\mathcal{A}_{\mathrm{sa}}$

a)sweis@mis.mpg.de.
b)knauf@mi.uni-erlangen.de.

the real vector space of self-adjoint matrices. We assume a quantum system is described by a *density matrix* ρ, also called *state*, i.e. $\rho \in \text{Mat}(N, \mathbb{C})$ (N-level system), $\rho \succeq 0$ (positive semi-definite) and $\text{tr}(\rho) = 1$ (normalized). We denote by $\mathcal{S}(\mathcal{A})$ the set of density matrices, called *state space*. We assume a generic quantum systems where the density matrix ρ is invertible.

The von Neumann measurements (see Ref. 32) of $a_r = \sum_{\lambda \in \text{spec}(a_r)} \lambda P_{r, \lambda}$ yield eigenvalue λ with probability $\text{tr}(\rho P_{r, \lambda})$.

- If n copies of ρ are available for measurement (in form of the n-fold tensor product $\rho \otimes \cdots \otimes \rho \in \mathcal{A}_{\text{sa}}^{\otimes n}$), then n measurements of a_r give us eigenvalues $\lambda_1, \ldots, \lambda_n \in \text{spec}(a_r)$ such that the mean

$$\overline{a_r}(n) \ := \ \tfrac{1}{n}(\lambda_1 + \ldots + \lambda_n) \tag{1}$$

  converges to the true mean $\text{tr}(\rho a_r)$ by the strong law of large numbers.
- If nk copies of ρ are available, the measured values $m_1, \ldots, m_k$ of the k random variables $\overline{a_1}(n), \ldots, \overline{a_k}(n)$ define an affine subspace

$$\left\{ \sigma \in \mathcal{A}_{\text{sa}} \mid \big(\text{tr}(\sigma a_1), \ldots, \text{tr}(\sigma a_k)\big) \ = \ \big(m_1, \ldots, m_k\big) \right\}. \tag{2}$$

  We assume that this subspace intersects the state space $\mathcal{S}(\mathcal{A})$, since by large deviation theory (e.g., Chap. I.3 of Ref. 15) the probability of a distance larger than a given $\varepsilon > 0$ from ρ decays exponentially in n.

The *maximum-entropy inference* associates to the measured values $(m_1, \ldots, m_k)$ the unique density matrix $\widehat{\rho}(n)$ in the set of states σ satisfying (2) which maximizes the *von Neumann entropy*

$$S(\sigma) \ := \ -\text{tr}(\sigma \log(\sigma)). \tag{3}$$

The maximum-entropy inference is well-defined for all mean value tuples $(\text{tr}(\sigma a_1), \ldots, \text{tr}(\sigma a_k))$ where σ is a density matrix, since the von Neumann entropy is a strictly concave function on the state space $\mathcal{S}(\mathcal{A})$.[41] Later we work with a projection of $\mathcal{S}(\mathcal{A})$, called *mean value set*, which is affinely isomorphic to the set of mean value tuples. When restricted to the set of mean value tuples of invertible density matrices σ, the maximum-entropy inference is a real analytic mapping, see, e.g., Ref. 45. The image, called *Gibbsian family* (of density matrices), consists of all matrices of the form

$$\exp(a_0 + \lambda_1 a_1 + \cdots + \lambda_k a_k)/\text{tr}(\exp(a_0 + \lambda_1 a_1 + \cdots + \lambda_k a_k))$$

for real $\lambda_1, \ldots, \lambda_k$ and $a_0 = 0$. In general, if $a_r \in \mathcal{A}_{\text{sa}}$, this manifold of density matrices is called *exponential family*.

A natural measure of efficiency is the asymptotical variance of the maximum-entropy inference. Here we refer to the variance of the inference applied to the random variables $\overline{a_r}(n)$, $r = 1, \ldots, k$, for large n. Its asymptotic order $\mathcal{O}(1/n)$ has appeared in the context of model selection.[34] However, we discuss in Remark 22 a Gibbsian family where the real analytic maximum-entropy inference defined on the interior of the mean value set has no continuous extension to the full mean value set. While the variance of the random variables $\overline{a_r}(n)$ and of the tuple $(\overline{a_1}(n), \ldots, \overline{a_k}(n))$ is of order $\mathcal{O}(1/n)$, the statement is not obvious for the maximum-entropy inference $\widehat{\rho}(n)$. Indeed, the constant in the variance estimate $\mathcal{O}(1/n)$ of $\widehat{\rho}(n)$ can be arbitrarily large. For the non-generic choice of a singular density matrix ρ the limit $\lim_{n \to \infty} \widehat{\rho}(n)$ needs not even be a state of maximum von Neumann entropy and we ask for its distribution of limits.

Maximum-entropy inference is closely connected to the entropy distance from an exponential family. The *relative entropy* between states $\rho, \sigma \in \mathcal{S}(\mathcal{A})$ is $S(\rho, \sigma) := +\infty$ unless the image of σ contains that of ρ and then (using the natural logarithm)

$$S(\rho, \sigma) \ := \ \text{tr}\, \rho \big(\ln(\rho) - \ln(\sigma) \big). \tag{4}$$

The distance-like properties of $S(\rho, \sigma) \geq 0$ and of $S(\rho, \sigma) = 0 \Leftrightarrow \rho = \sigma$ hold.[41] However, the relative entropy is not a metric. For $\mathcal{E} \subset \mathcal{S}(\mathcal{A})$

$$d_{\mathcal{E}} \ : \quad \mathcal{S}(\mathcal{A}) \to \mathbb{R}, \quad \rho \mapsto \inf_{\sigma \in \mathcal{E}} S(\rho, \sigma) \tag{5}$$

is called *entropy distance* of ρ from $\mathcal{E}$. If $\mathcal{E}$ contains invertible density matrices, then $d_{\mathcal{E}}$ is bounded on $\mathcal{S}(\mathcal{A})$.

Entropy distance has several motivations. Under arbitrary constraints, maximizing the von Neumann entropy is the same as minimizing the relative entropy distance $d_{\{1/\text{tr}(1)\}}$ from the tracial state. In Sec. II we recall that for linear constraints the latter is equivalent to the unconstrained minimization of the relative entropy in its second argument from the corresponding Gibbsian family. This minimum is the entropy distance defined in (5). It also plays an important role in another field. *Infomax principles* support the hypothesis that natural systems tend to maximize structured correlations. This, in the work, see Ref. 5, is formalized as deviation from an exponential family $\mathcal{E}$, and is quantified by the entropy distance. An instructive example is the mutual information used in information theory:

Example 1 (Product States): The *mutual information* of a bipartite state ρ_{AB} is given by $S(\rho_{AB}, \rho_A \otimes \rho_B) \geq 0$ for the relative entropy S and for reduced states ρ_A resp. ρ_B on subsystem A resp. B. It is zero only when $\rho_{AB} = \rho_A \otimes \rho_B$. The relative entropy measures the distance of an arbitrary bipartite state from the Gibbsian family of all product states.

The mutual information of a quantum system measures the total correlation of a bipartite quantum system. Quantum correlations include *entanglement* which is an inherently quantum kind of correlation and its numerical measures are still being discussed. One approach is to use the entropy distance from the set of not entangled states, known as *relative entropy of entanglement*, see e.g. Refs. 19 and 40. Correlation measures having the form of the entropy distance from a Gibbsian family are used in statistical physics, image processing or in the theory of neural networks to just name a few, see e.g. Refs. 2, 7, 16, and 25. Maximizers of the entropy distance from an exponential family (of probability distributions) were studied, e.g., in Refs. 5, 6, 26, 27, and 35. In Sec. V we contribute to a non-commutative analog by computing first order optimality conditions for the entropy distance from a Gibbsian family.

B. Summary of our results

Rest of the paper will focus on observables in the algebra of Example 3. We study a curve of planes in a Grassmannian manifold of linear spaces that defines a curve of two-dimensional Gibbsian families of 3×3-density matrices. Unlike Gibbsian families in finite probability spaces, one of the families has a discontinuous entropy distance and its real analytic maximum-entropy inference does not extend continuously. We discuss several candidates of closures to extend Gibbsian families and we propose a convex geometric criterion to characterize discontinuities: Where non-exposed faces are born in a Grassmannian manifold of linear subspaces, families have a discontinuous inference. This conjecture is supported by the example of the Staffelberg family in Sec. IV B.

To compare classical and quantum physics, we consider **-subalgebras* $\mathcal{A}$ of $\text{Mat}(N, \mathbb{C})$. To allow low-dimensional examples we consider them *real*, i.e., $\mathcal{A}$ is a subring of $\text{Mat}(N, \mathbb{C})$, and an $\mathbb{R}$-module closed under conjugation $a \mapsto a^*$. However, it is not necessarily closed under complex scalar multiplication. The *state space* of $\mathcal{A}$ is the set $\mathcal{S} = \mathcal{S}(\mathcal{A}) = \{\rho \in \mathcal{A} \mid \rho \succeq 0, \text{tr}(\rho) = 1\}$ of density matrices. We denote $1 / 0$ resp. $1_N / 0_N$ the identity / zero in $\mathcal{A}$ resp. $\text{Mat}(N, \mathbb{C})$. We allow for $1 \neq 1_N$ which we need to study the *swallow family* in Sec. IV D and to prove an optimality condition in Sec. V, see also Remark 6. The real vector space of self-adjoint matrices $\mathcal{A}_{\text{sa}}$ is an Euclidean vector space for the *Hilbert-Schmidt* scalar product $\langle a, b \rangle = \text{tr}(ab)$, $a, b \in \mathcal{A}_{\text{sa}}$.

Remark 2: There are other natural definitions of the state space of a real *-subalgebra $\mathcal{A}$ of $\text{Mat}(N, \mathbb{C})$, e.g.,

1. the density matrices in $\mathcal{A}$ (like above),
2. the states on $\text{Mat}(N, \mathbb{C})$ restricted to $\mathcal{A}_{\text{sa}}$,
3. the positive linear functionals on $\mathcal{A}_{\text{sa}}$ that take the value 1 at the identity.

These definitions are mutually equivalent, assuming $\mathbb{1}_N \in \mathcal{A}$. The inclusions of 1, into 2, into 3. are trivial. The inclusion of 3. into 2. follows from the Riesz extension theorem and the inclusion of 2. into 1. follows from the fact that orthogonal projection from $\mathrm{Mat}(N, \mathbb{C})_{\mathrm{sa}}$ onto $\mathcal{A}_{\mathrm{sa}}$ takes density matrices to density matrices.

The following real *-subalgebra of the C*-algebra $\mathrm{Mat}(2, \mathbb{C}) \oplus \mathbb{C}$ is sufficiently rich for our purposes and it includes the curve of Gibbsian families. The state space of $\mathrm{Mat}(2, \mathbb{C}) \oplus \mathbb{C}$ has already been analyzed in Ref. 23 as the simplest example of a "hybrid" memory, and improved coding theorems were obtained, taking into account direct sums of full matrix algebras. Here we use this algebra for the sole purpose to have simple low-dimensional examples of state spaces.

Example 3: We consider the real *-subalgebra $\mathcal{B} \subset \mathrm{Mat}(2, \mathbb{C})$ spanned by $\mathbb{1}_2, \sigma_1, \sigma_2$, and $\mathrm{i}\sigma_3$ for the *Pauli σ-matrices* $\sigma_1 := \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $\sigma_2 := \left(\begin{smallmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{smallmatrix}\right)$, $\sigma_3 := \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. This algebra is isomorphic to $\mathrm{Mat}(2, \mathbb{R})$ by exchanging σ_2 and σ_3.

A real *-subalgebra $\mathcal{A} \subset \mathrm{Mat}(3, \mathbb{C})$ is defined by block diagonal matrices $\left(\begin{smallmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & * \end{smallmatrix}\right)$ with elements of $\mathcal{B}$ in the upper left corner and real numbers in the lower right corner. The state space of $\mathcal{B}$ is $\mathcal{S}(\mathcal{B}) = \mathrm{conv}\{\frac{1}{2}(\mathbb{1}_2 + \sin(\alpha)\sigma_1 + \cos(\alpha)\sigma_2) \mid \alpha \in \mathbb{R}\}$ where conv denotes convex hull. This disk is a section of the state space of $\mathrm{Mat}(2, \mathbb{C})$, known as *Bloch ball*. The state space of $\mathcal{A}$ is a three-dimensional cone based on $\mathcal{S}(\mathcal{B}) \oplus 0$ and with apex $0_2 \oplus 1$,

$$\mathcal{S}(\mathcal{A}) = \mathrm{conv}(0_2 \oplus 1, \rho(\alpha); \alpha \in \mathbb{R})$$

for

$$\rho(\alpha) := \tfrac{1}{2}(\mathbb{1}_2 + \sin(\alpha)\sigma_1 + \cos(\alpha)\sigma_2) \oplus 0.$$

It is the solid of revolution of an equilateral triangle.

It is well known that state spaces of commutative and non-commutative algebras have quite different geometries. Whereas in the commutative case we have a simplex (and thus every state is uniquely decomposed into pure states), in the non-commutative case such a decomposition is highly non-unique (think of the Bloch ball).

Still, from the point of view of convex geometry there is one common property of all these state spaces: all of their faces are exposed, that is, they can be described as the intersection of state space with a half space, see Ref. 1. *Non-exposed faces* are found, e.g., on the circumference of a stadium, at the four points where a half-circle meets a segment. See Sec. III for precise definitions. In the probabilistic setting of $\mathcal{A} \cong \mathbb{C}^N$, embedded as diagonal matrices, measurement of observables $f_1, \ldots, f_n$ leads to an orthogonal projection

$$\mathcal{S}(\mathcal{A}) \longrightarrow \mathbb{R}^n, \quad p \longmapsto \big(\mathbb{E}_p(f_1), \ldots, \mathbb{E}_p(f_n)\big)$$

of state space, based on expectation $\mathbb{E}_p$. The image, called *mean value set* or *convex support*[8] is no longer a simplex but still a polygon. So faces of a mean value set are exposed, too. The same applies to all exponential families and their natural projections, see Figure 1.

We exhibit here two main differences between exponential families in commutative and non-commutative algebras, at least in the curve of our example.

- First, we show in Sec. III that it is typical for a non-commutative algebra that mean value sets have non-exposed faces.
- Second, we show in Sec. IV B that the entropy distance from an exponential family can be discontinuous in exceptional cases.

In Figure 2 (left) we have drawn two-dimensional mean value sets of the state space $\mathcal{S}(\mathcal{A})$ from Example 3. A mean value set has non-exposed faces if it is the convex hull of a non-degenerate ellipse and of an exterior point. Mean value sets with non-exposed faces are bounded in the Grassmannian manifold by elliptical shapes that correspond to exponential families with discontinuous entropy distance (angles $\frac{1}{6}\pi, \frac{5}{6}\pi, \frac{7}{6}\pi$, and $\frac{11}{6}\pi$). It seems this boundary in the Grassmannian manifold is pivotal for discontinuity.
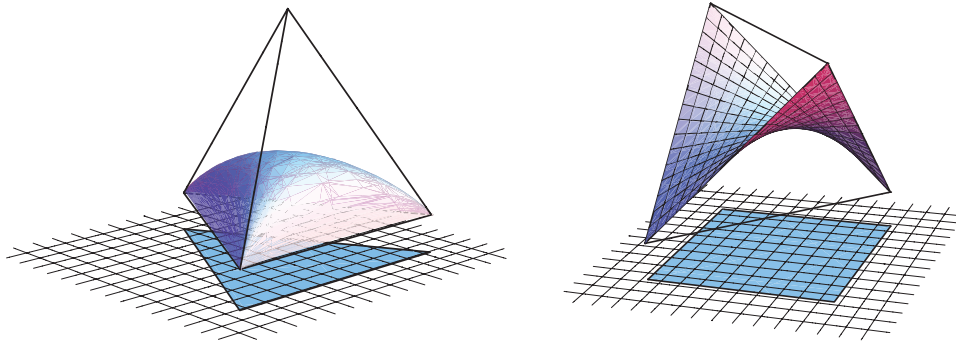
FIG. 1. Mean value sets for two probabilistic exponential families. Left: triangle; Right: square.

Sketching ideas to a classification of non-exposed faces of a mean value set, these non-exposed faces are related to singularities of a section of the state space with a plane parallel to the projection:[10, 44] The singularities have non-unique tangents and they are incident with at most one segment, see Figure 2. Projections and sections are related by convex duality. Here is much more to explore in terms of convex algebraic geometry, see, e.g., Ref. 28.

Already in 1963 ensembles of maximum chaos in $\mathcal{A} = \mathrm{Mat}(N, \mathbb{C})$ were studied in Ref. 45. However, non-exposed faces at a mean value set have attracted little attention in the literature. In particular Theorem I (e) in Ref. 45, concerning extreme points is wrong, it fails in all cases where non-exposed extreme points appear. An example is given in Remark 29 (a). We are convinced that non-exposed faces are important in the analysis of maximum-entropy inference and entropy distance. As we have seen in the beginning of this section the convex geometric notion of non-exposed face indicates discontinuity of the inference.

We will show later in this paper that the maximum-entropy inference does not extend continuously. So the question arises how a Gibbsian family $\mathcal{G}$ must be extended to a set of maximum-entropy density matrices under linear constraints. It is clear that the topological norm closure is too large. In the examples presented in Sec. IV we will prove that the *reverse information closure* or *rI-closure*

$$\mathrm{cl}^{\mathrm{rI}}(\mathcal{G}) \ := \ \{\rho \in \mathcal{S}(\mathcal{A}) \mid \inf_{\sigma \in \mathcal{G}} S(\rho, \sigma) = 0\} \tag{6}$$

gives the right answer. Its name is motivated from probability theory[13] and it consists of states that approximate $\mathcal{G}$ in relative entropy S. Information geometry[3] gives us two canonical choices
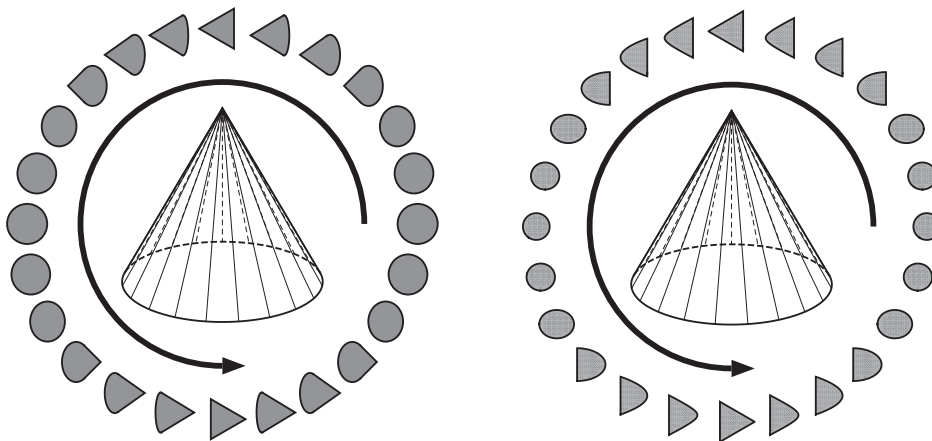

FIG. 2. The 3D cone is the state space of a non-commutative algebra. Left: Mean value sets (projections of the cone); Right: sections of the cone. Projections respectively sections are with respect to planes including the tracial state $\mathbb{1}/\mathrm{tr}(\mathbb{1})$, which is the centroid of the cone.

of geodesics on the manifold $\mathcal{G}$. These $(+1)$-geodesics and (-1)-geodesics will be defined in the Sec. II. They give rise to the $(+1)$-*closure*

$$\mathrm{cl}^{(+1)}(\mathcal{G}) \; := \; \mathcal{G} \; \cup \; \{ \text{ limit points of } (+1)\text{-geodesics in } \mathcal{G} \} \tag{7}$$

and the (-1)-*closure*

$$\mathrm{cl}^{(-1)}(\mathcal{G}) \; := \; \mathcal{G} \; \cup \; \{ \text{ limit points of } (-1)\text{-geodesics in } \mathcal{G} \} \, . \tag{8}$$

The inclusions of $\mathrm{cl}^{(+1)}(\mathcal{G})$ and $\mathrm{cl}^{(-1)}(\mathcal{G})$ into the norm closure $\overline{\mathcal{G}}$ are obvious. We show

$$\mathrm{cl}^{(+1)}(\mathcal{G}) \; \subset \; \mathrm{cl}^{\mathrm{rI}}(\mathcal{G}) \; \subset \; \overline{\mathcal{G}} \, , \tag{9}$$

where the second inclusion follows from the Pinsker-Csiszár inequality. We prove that the $(+1)$-closure is smaller than the set of maximum-entropy density matrices, and that the rI-closure and the (-1)-closure are possible candidates for the correct extension of $\mathcal{G}$.

II. TWO AFFINE CHARTS AND SOME REMARKS

We introduce two sorts of canonical geodesics on a Gibbsian family and we provide a geometric discussion of how the maximum-entropy inference relates to the entropy distance. We remark on the information geometric context of the geodesics, on quantum channels and on advantages of real *-subalgebras as opposed to C*-subalgebras.

In this section $\mathcal{A}$ denotes an arbitrary real *-subalgebra of $\mathrm{Mat}(N, \mathbb{C})$. The set of invertible states equals the relative interior of the state space

$$\mathrm{ri}\,\mathcal{S}(\mathcal{A}) \; = \; \{ \rho \in \mathcal{S}(\mathcal{A}) \mid \rho^{-1} \text{ exists in } \mathcal{A} \} \, ,$$

i.e., the interior of $\mathcal{S}(\mathcal{A})$ in its affine span $\mathcal{A}_1 := \{ a \in \mathcal{A}_{\mathrm{sa}} \mid \mathrm{tr}(a) = 1 \}$, see e.g. Proposition 2.9 in Ref. 43. The trace-normalized exponential is the real analytic mapping

$$\exp_1 \; : \quad \mathcal{A}_{\mathrm{sa}} \; \longrightarrow \; \mathrm{ri}\,\mathcal{S}(\mathcal{A}), \quad a \; \longmapsto \; \frac{e^a}{\mathrm{tr}(e^a)}$$

defined by functional calculus of self-adjoint matrices in $\mathcal{A}$. This is a diffeomorphism when restricted to traceless matrices. The real analytic inverse $\ln_0 : \mathrm{ri}(\mathcal{S}(\mathcal{A})) \to \mathcal{A}_0$ defined by

$$\ln_0 \; : \; \rho \mapsto \ln(\rho) - \mathbb{1}\,\mathrm{tr}(\ln(\rho))/\mathrm{tr}(\mathbb{1})$$

is the *canonical chart* of $\mathrm{ri}\,\mathcal{S}(\mathcal{A})$.

The image of a non-empty affine subspace of $\mathcal{A}_{\mathrm{sa}}$ under $\exp_1$ is an *exponential family* in $\mathcal{A}$. For an exponential family $\mathcal{E}$ we call $\ln_0 |_{\mathcal{E}}$ the *canonical chart* of $\mathcal{E}$. The affine space $\Theta := \ln_0(\mathcal{E})$ is the *canonical parameter space*, its translation vector space $V := \{ x - y \mid x, y \in \Theta \}$ is the *canonical tangent space* and the restriction of $\exp_1$ to Θ is the *canonical parametrization* of $\mathcal{E}$.

An exponential family is a Gibbsian family if $\Theta = V$, and for that case a different chart was introduced in Theorem 2 (b) in Ref. 45: If $\pi_V : \mathcal{A}_{\mathrm{sa}} \to V$ denotes orthogonal projection onto V, we define the *mean value set*

$$\mathbb{M}(V) \; = \; \mathbb{M}_{\mathcal{A}}(V) \; := \; \pi_V\big(\mathcal{S}(\mathcal{A})\big) \, . \tag{10}$$

The mean value set is affinely isomorphic to $\{(\langle \rho, v_1 \rangle, \ldots, \langle \rho, v_1 \rangle \mid \rho \in \mathcal{S}(\mathcal{A})\}$, if $v_1, \ldots, v_k$ is a basis of V, see e.g. Remark 1.1 in Ref. 43. The latter set was used in Ref. 45. It is not reasonable to choose a basis of V in our analysis, because vector spaces pVp for projections $p = p^2 = p^* \in \mathcal{A}$ will be used, see Remark 6, and multiplication with p can destroy linear independence.

The map $\pi_V \circ \exp_1 |_V : V \to \mathrm{ri}\big(\mathbb{M}(V)\big)$ is a real analytic diffeomorphism, its image is an open subset of V. The *mean value chart* for the Gibbsian family $\mathcal{E}$ is the bijection

$$\pi_V|_{\mathcal{E}} \; : \quad \mathcal{E} \; \longrightarrow \; \mathrm{ri}\big(\mathbb{M}(V)\big) \, . \tag{11}$$

The real analytic inverse $\pi_{\mathcal{E}} : \mathrm{ri}\big(\mathbb{M}(V)\big) \to \mathcal{E}$ shall be called *mean value parametrization*. Below we also write $\pi_{\mathcal{E}}$ for the map $\pi_{\mathcal{E}} \circ \pi_V$ defined on the *domain* $\mathrm{dom}\,\mathcal{E} := \mathcal{S}(\mathcal{A}) \cap (\mathcal{E} + V^{\perp})$, which was introduced in Ref. 5 (for probability distributions). In fact, the chart (11) was established in Ref. 45 for $\mathcal{A} = \mathrm{Mat}(N, \mathbb{C})$. Since V contains only traceless matrices, its is proved in Lemma 3.13 in

Ref. 43 that $\mathbb{M}_{\mathcal{A}}(V) = \mathbb{M}_{\mathrm{Mat}(N,\mathbb{C})}(V)$ holds for every C*-subalgebra $\mathcal{A} \subset \mathrm{Mat}(N, \mathbb{C})$ which contains V. Remark 6 extends this equality to real *-subalgebras $\mathcal{A}$ including V. So (11) holds for these algebras.

The two charts for a Gibbsian family $\mathcal{E}$ have open subsets of the canonical tangent space V as their images. Given that V is an affine space, two kinds of affine geodesics for $\mathcal{E}$ arise: Unparametrized $(+1)$-*geodesics* are the images of open segments in V under the canonical parametrization $\exp_1 : V \to \mathcal{E}$, and unparametrized (-1)-*geodesics* are the images of open segments in $\mathrm{ri}(\mathbb{M}(V))$ under the mean value parametrization $\pi_{\mathcal{E}} : \mathrm{ri}(\mathbb{M}(V)) \to \mathcal{E}$. We shall denote the *open segment* between $a, b \in \mathcal{A}_{\mathrm{sa}}$ by $]a, b[:= \{(1 - \lambda)a + \lambda b | 0 < \lambda < 1\}$ and the *closed segment* by $[a, b] := \{(1 - \lambda)a + \lambda b | 0 \leq \lambda \leq 1\}$. A more comprehensive introduction of $(\pm)$-geodesics is given in Sec. 7.2 in Ref. 3. The geodesics are part of a beautiful theory, called information geometry, about affine connections and Riemannian metrics on state spaces. See Remark 4 for some details.

The relative entropy suits exponential families very well. If ρ, σ and τ are states in $\mathcal{A}$ with σ and τ invertible, and if $\rho - \sigma \perp \ln(\tau) - \ln(\sigma)$, then

$$S(\rho, \sigma) + S(\sigma, \tau) = S(\rho, \tau) \tag{12}$$

holds, see e.g., Ref. 30. This is the *Pythagorean theorem* of the relative entropy. Clearly the Pythagorean theorem (12) holds if σ and τ belong to an exponential family $\mathcal{E}$ in $\mathcal{A}$ and if $\rho \in \mathcal{S}(\mathcal{A})$ satisfies $\rho - \sigma \perp V$. The *projection theorem* follows for $\rho \in \mathrm{dom}\, \mathcal{E}$:

$$\min_{\sigma \in \mathcal{E}} S(\rho, \sigma) = S(\rho, \pi_{\mathcal{E}}(\rho)), \tag{13}$$

the minimum being unique. See Remark 4 about the information geometry of these theorems.

The linearly constrained maximization of von Neumann entropy can be replaced by an unconstrained minimization of the relative entropy. As mentioned previously, for $V = \Theta$ the mean value parametrization

$$\pi_{\mathcal{E}} : \mathrm{ri}(\mathbb{M}(V)) \to \mathcal{E} \tag{14}$$

assigns to vectors $v \in \mathrm{ri}(\mathbb{M}(V))$ the unique state $\pi_{\mathcal{E}}(v)$ of maximum von Neumann entropy in the fiber $F(v) := (v + V^{\perp}) \cap \mathcal{S}(\mathcal{A})$. This is often proved using Lagrange multipliers or Klein's inequality, see e.g. Ref. 20.

A proof of (14) by information geometry opens a link to the entropy distance: Let $\tau = \mathbb{1}/\mathrm{tr}(\mathbb{1})$ be a reference state. Then $S(\rho, \tau) = -S(\rho) + \log(\mathrm{tr}(\mathbb{1}))$ for all $\rho \in \mathcal{S}(\mathcal{A})$, so maximizing the von Neumann entropy is equivalent to minimizing the relative entropy from τ, under arbitrary constraints (a different choice of τ corresponds to a biased inference[37]). For all $v \in \mathrm{ri}(\mathbb{M}(V))$ the fiber $F(v)$ is included in the domain $\mathrm{dom}\, \mathcal{E}$ of the Gibbsian family $\mathcal{E} = \exp_1(V)$ by the mean value chart (11). Since $\tau \in \mathcal{E}$, the Pythagorean theorem (12) shows for any state $\rho \in F(v)$

$$S(\rho, \pi_{\mathcal{E}}(\rho)) + S(\pi_{\mathcal{E}}(\rho), \tau) = S(\rho, \tau).$$

Minimizing $S(\cdot, \tau)$ over $\rho \in F(v)$ has the unique solution $\pi_{\mathcal{E}}(\rho)$. In addition, by the projection theorem (13), it is equivalent to the unconstrained minimization of $S(\rho, \cdot)$ on $\mathcal{E}$ (independent of the choice of $\rho \in F(v)$).

Pythagorean and projection theorems as well as the $(+1)$- and (-1)-geodesics are rooted in information geometry.

Remark 4: The exponential family $\mathrm{ri}(\mathcal{S}(\mathcal{A}))$ of invertible density matrices has the mean value chart $\rho \mapsto \rho - \mathbb{1}/\mathrm{tr}(\mathbb{1})$. Its tangent space at ρ is called the *(m)-representation* and equals $\mathcal{A}_0 := \{a \in \mathcal{A}_{\mathrm{sa}} | \mathrm{tr}(a) = 0\}$, see p. 148 in Ref. 3. According to Ref. 30, the BKM (Bogoliubov-Kubo-Mori) metric, a Riemannian metric on $\mathrm{ri}(\mathcal{S}(\mathcal{A}))$, can be defined for invertible density matrices ρ and tangent vectors A, B in the (m)-representation by

$$g(A, B)_{\rho} := \int_0^{\infty} \mathrm{tr}\left((t + \rho)^{-1} A (t + \rho)^{-1} B\right) \mathrm{d}t \,.$$

Although the BKM metric is a natural generalization of the Fisher metric to state spaces of non-commutative algebras, unlike the former it is not the only such monotone one, see e.g. Ref. 31.

The *(m)-connection* on the state space ri$(\mathcal{S}(\mathcal{A}))$, denoted $\nabla^{(m)}$, is defined through the parallel transport of translation on the affine hull $\mathcal{A}_1 = \{a \in \mathcal{A}_{sa} \mid \operatorname{tr}(a) = 1\}$ of the state space. If g is a Riemannian metric on the manifold ri$(\mathcal{S}(\mathcal{A}))$ then the *(e)-connection*, denoted $\nabla^{(e)}$, is defined by

$$X g(Y, Z) = g\big(\nabla_X^{(m)} Y, Z\big) + g\big(Y, \nabla_X^{(e)} Z\big)$$

for vector fields X, Y, Z on ri$(\mathcal{S}(\mathcal{A}))$. The connections $\nabla^{(m)}$ and $\nabla^{(e)}$ are said to be *dual* with respect to g. The (m)-connection is also called (-1)-connection, and when the BKM Riemannian metric g is used, then the dual (e)-connection is called $(+1)$-connection. The connections $\nabla^{(+1)}$ and $\nabla^{(-1)}$ give rise to the geodesics introduced above, see e.g. Secs. 7.2 and 7.3 in Ref. 3.

The state space of the (real) *-subalgebra $\mathcal{A}$ is trivially (-1)-autoparallel (i.e. totally geodesic) and it is $(+1)$-autoparallel as it is an exponential family. This shows that the $(\pm)$-connections restrict from the state space of Mat$(N, \mathbb{C})$ to ri$(\mathcal{S}(\mathcal{A}))$.

A Pythagorean theorem and a projection theorem are known in information geometry for *dually flat spaces*. The relative entropy is the *canonical divergence* of the dually flat space of invertible density matrices with respect to the BKM metric and the $(\pm)$-connections. Hence the Pythagorean theorem (12) arises from a more general theory, see, e.g., Sec. 3.4 in Ref. 3. The (-1)-geodesic through ρ and σ and the $(+1)$-geodesic through σ and τ meet at σ orthogonally with respect to the BKM metric.

The projection $\pi_V|_{\mathcal{S}(\mathcal{A})} : \mathcal{S}(\mathcal{A}) \to V$ can be seen as a quantum channel to a commutative algebra.

Remark 5: The mean value set $\mathbb{M}(V) = \pi_V(\mathcal{S}(\mathcal{A}))$ relates to a POVM quantum measurement. A *POVM* is defined as a finite sequence $F_1, \ldots, F_n$ of positive semidefinite matrices in $\mathcal{A}$, such that $F_1 + \cdots + F_n = \mathbb{1}$. The probability of outcome $i \in \{1, \ldots, n\}$ when measuring the quantum system $\rho \in \mathcal{S}(\mathcal{A})$ is $\mathbb{P}_\rho(i) := \operatorname{tr}(F_i \rho)$, see, e.g., Ref. 32. Given a POVM $F_1, \ldots, F_n$ in $\mathcal{A}$, a quantum channel

$$\mathcal{S}(\mathcal{A}) \to \mathcal{S}(\mathbb{C}^n), \quad \rho \mapsto (\mathbb{P}_\rho(1), \ldots, \mathbb{P}_\rho(n)) = (\operatorname{tr}(F_1 \rho), \ldots, \operatorname{tr}(F_n \rho))$$

is defined. If U is the real linear span of $F_1, \ldots, F_n$ and $\widetilde{U}$ is the orthogonal projection of U onto the space of traceless matrices $\mathcal{A}_0$, then the mean value sets $\mathbb{M}(U) = \pi_U(\mathcal{S}(\mathcal{A}))$ and $\mathbb{M}(\widetilde{U}) = \pi_{\widetilde{U}}(\mathcal{S}(\mathcal{A}))$ are affinely isomorphic to the image of the above channel $\mathcal{S}(\mathcal{A}) \to \mathcal{S}(\mathbb{C}^n)$. (For a proof see, e.g., Remark 1.1 in Ref. 43).

We would like to comment on (real) *-subalgebras.

Remark 6: As already mentioned earlier, *-subalgebras allow for low-dimensional examples. What makes *-subalgebra $\mathcal{A}$ of Mat$(N, \mathbb{C})$ eligible for our treatment is that all results in Ref. 43 are true for them (unfortunately the choice in that article was to argue with intersections of C*-subalgebras and real matrices Mat$(N, \mathbb{R})$). Some caution is needed, e.g. spectral projections of normal matrices need not be included in $\mathcal{A}$, as the matrix $i\sigma_3 \oplus 0$ in Example 3 shows. This error is present in Definition 2.5.3 of the above article. However, as only self-adjoint matrices are used, there is no problem arising.

An important feature of a *-subalgebra $\mathcal{A}$ of Mat$(N, \mathbb{C})$ is that spectral projections p of a self-adjoint matrix $a \in \mathcal{A}_{sa}$ can be written as $p = f(a)$ for a real polynomial f in one variable. This implies that

- if a is a self-adjoint matrix and g is a real valued function defined on the spectrum of a, then $g(a)$ belongs to $\mathcal{A}_{sa}$,
- the state space has codimension one in $\mathcal{A}_{sa}$, as the cone of positive semi-definite matrices has full dimension (decompose a self-adjoint matrix into a difference of two positive semi-definite matrices).

One superficial flaw of *-subalgebras (and of C*-subalgebras!) is that eigenvalues cannot be used directly, as the identity $\mathbb{1}$ of $\mathcal{A}$ may differ from the identity $\mathbb{1}_N$ in Mat$(N, \mathbb{C})$. On a closer inspection one realizes that this is exactly the flexibility we need, e.g., in Proposition 14 and Theorem 27 to analyze rI-closures. The $(+1)$-closure of an exponential family is formed by exponential families

of strictly smaller support, lying in *compressed algebras*

$$p\mathcal{A}p \;=\; \{pap \mid a \in \mathcal{A}\}$$

with identity $p = p^2 = p^* \in \mathcal{A}$. The algebra $p\mathcal{A}p$ as a *-subalgebra of $\mathrm{Mat}(N, \mathbb{C})$ may be treated in the same way as $\mathcal{A}$. The unorthodox use of spectral values within a finite-dimensional algebra was overlooked in Ref. 43, see the correction [Lin. Alg. Appl. **436** 1 p. xvi (2012)].

III. A CLASSICAL—QUANTUM METAMORPHOSIS

In the algebra $\mathcal{A}$ from Example 3 we study a curve of 2D mean value sets and we address the question whether they have non-exposed faces. The algebra $\mathcal{A}$ has the commutative *-subalgebra $\begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & * \end{pmatrix}$ of diagonal matrices, isomorphic to $\mathbb{R}^3$, and its left upper corner $\begin{pmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & 0 \end{pmatrix}$ is a non-commutative *-subalgebra.

The curve of mean value sets is rather a Grassmannian manifold of subspaces. More precisely, we consider 2D subspaces of the 4D space $\mathcal{A}_{\mathrm{sa}}$ of self-adjoint matrices and here we restrict to 2D subspaces of the 3D space of traceless matrices (since the state space is parallel to it). So by symmetry of the cone $\mathcal{S}(\mathcal{A})$ one real angular variable suffices to describe mean value sets. Thus we can consider a curve in the Grassmannian manifold. In Figure 2, left, mean value sets $\mathbb{M}(V)$ are drawn isometrically at equidistant $\frac{1}{12}\pi$ angles around a full circle.

Our example is minimal in two respects:

- Planar projections have minimal dimension to allow for non-exposed faces.
- The algebra $\mathcal{A}$ is (up to isomorphism) the smallest *-subalgebra $\mathcal{A}$ of $\mathrm{Mat}(N, \mathbb{C})$ allowing for a mean value set with non-exposed faces. If $\mathcal{A}$ has no *-subalgebra isomorphic to $\mathrm{Mat}(2, \mathbb{R})$ then, assuming $\mathbb{1} = \mathbb{1}_N$, then $\mathcal{A}$ is commutative (see Theorem 5.2 and 5.4 in Sec. 5 in Ref. 22). ($\mathbb{1} = \mathbb{1}_{\widetilde{N}}$ may be achieved by restricting a faithful representation of the C*-algebra $\mathbb{1}\,\mathrm{Mat}(N, \mathbb{C})\mathbb{1}$ onto a direct sum of full matrix algebras, see, e.g., Ref. 14.) Hence the state space $\mathcal{S}(\mathcal{A})$ is a simplex. Then all mean value sets are polytopes having no non-exposed faces. The algebra $\mathrm{Mat}(2, \mathbb{R}) \cong \mathrm{span}_{\mathbb{R}}\{\mathbb{1}_2, \sigma_1, \sigma_2, \mathrm{i}\sigma_3\}$ itself has a disk as state space, whose proper projections are a point or a segment, having no non-exposed faces.

We introduce precise definitions in convex geometry, see e.g. Refs. 36 and 17 for subsequent discussions e.g. in Lemma 13.

Definition 7: Let M be a compact and convex subset of a finite-dimensional Euclidean vector space $(\mathbb{E}, \langle \cdot, \cdot \rangle)$.

- A convex subset F of M is a *face* of M, if for all $x, y \in M$ and all $0 < \lambda < 1$ the inclusion of $(1 - \lambda)x + \lambda y \in F$ implies $x, y \in F$.
- A face of dimension zero is called *extreme point* and if it is not exposed, a *non-exposed point*. An extreme point of $\mathcal{S}(\mathcal{A})$ will be called *pure state*.
- If M is non-empty, then for non-zero $u \in \mathbb{E}$ the *supporting hyperplane* is defined by

$$H(M, u) \;:=\; \{x \in \mathbb{E} \mid \langle x, u \rangle = \max_{y \in M} \langle y, u \rangle\}.$$

- A face F of M is called *exposed* if F is the intersection of M with a supporting hyperplane

$$F(M, u) \;:=\; M \cap H(M, u). \tag{15}$$

$F = \emptyset$ and $F = M$ are exposed faces by definition.

The Grassmannian manifold $\mathbb{G}$ of real 2D subspaces of self-adjoint traceless matrices $\mathcal{A}_0 = \{a \in \mathcal{A}_{\mathrm{sa}} \mid \mathrm{tr}(a) = 0\}$ will be denoted

$$\mathbb{G} \;:=\; \{V \subset \mathcal{A}_0 \mid V \text{ is a real 2D subspace}\}.$$

We define the angle between a subspace $V \in \mathbb{G}$ and $z := (-\frac{1}{2}\mathbb{1}_2) \oplus 1$ (pointing along the axis of the cone),

$$\varphi \;=\; \varphi(V) \;:=\; \measuredangle(V, z). \tag{16}$$

The state space is $\mathcal{S}(\mathcal{A}) = \operatorname{conv}(\mathcal{S}(\mathcal{B}) \cup \{0_2 \oplus 1\})$ for the disk $\mathcal{S}(\mathcal{B})$ introduced in Example 3. The mean value set of $V \in \mathbb{G}$ is the convex hull of the ellipse $e := \pi_V(\mathcal{S}(\mathcal{B}))$ and of $x := \pi_V(0_2 \oplus 1)$,

$$\mathbb{M}(V) \;=\; \operatorname{conv}(e, x). \tag{17}$$

The problem of finding non-exposed faces at $\mathbb{M}(V)$ may be solved in $\mathbb{R}^3$ by studying projections of a symmetric 3D cone isometric to $\mathcal{S}(\mathcal{A})$. Explicit calculations with matrices are done in Example 1.2 in Ref. 43 by studying tangents to the elliptical boundary curve ∂e. For all subspaces $V \in \mathbb{G}$ the projection of V onto $\operatorname{span}_{\mathbb{R}}(\sigma_1, \sigma_2, \sigma_3) \oplus 0$ is a subspace of $\operatorname{span}_{\mathbb{R}}(\sigma_1, \sigma_2) \oplus 0$. Hence the state space $\mathcal{S}(\mathcal{A})$ equals the cone C in Ref. 43 and we have the following:

Lemma 8: Let $V \in \mathbb{G}$ be a 2D plane. If $\varphi = 0$, then ∂e is a segment (degenerate ellipse) and the mean value set $\mathbb{M}(V)$ is a triangle. If $0 < \varphi < \frac{\pi}{3}$, then ∂e is a non-degenerate ellipse, $x \notin e$ and the tangents from x to ∂e meet ∂e at two non-exposed points of $\mathbb{M}(V)$. If $\frac{\pi}{3} \leq \varphi \leq \frac{\pi}{2}$, then $\mathbb{M}(V) = e$ is bounded by a non-degenerate ellipse ∂e.

We see that non-exposed faces are typical in the following sense. A continuous curve γ: $[0, 1] \to \mathbb{G}$ induces a curve of mean value sets $\lambda \mapsto \mathbb{M}(\gamma(\lambda))$. By Lemma 8 a mean value set without non-exposed faces must be a triangle or an ellipse. If γ connects the classical mean value set of a triangle to an ellipse, then we have $\measuredangle(\gamma(0), z) = 0$ and $\measuredangle(\gamma(1), z) \in [\frac{\pi}{3}, \frac{\pi}{2}]$. Since the angle φ is continuous on $\mathbb{G}$, the curve γ must cross the range of angles $(0, \frac{\pi}{3})$ with mean value sets having non-exposed faces. This range corresponds to an open subset of the Grassmannian $\mathbb{G}$.

IV. CLOSURES OF EXPONENTIAL FAMILIES

The curve of 2D mean value sets $\mathbb{M}(V)$ in Sec. III shows that the angle of $\varphi = \varphi(V) = \frac{\pi}{3}$ divides mean value sets with non-exposed faces from others without non-exposed faces. In Sec. IV B we show that the Gibbsian family at $\varphi = \frac{\pi}{3}$, called Staffelberg family, has a discontinuous entropy distance. The analysis is based on more general results about $(+1)$-closures in Sec. IV A. In Sec. IV C we compute the (-1)-closure of the Staffelberg family. We will see in Sec. IV D that the $(+1)$-closure of a Gibbsian family, in general, is not a locus of maximum-entropy density matrices under linear constraints.

In the sequel we assume that $\mathcal{A}$ is a real *-subalgebra of $\operatorname{Mat}(N, \mathbb{C})$ and that $\mathcal{E}$ is an exponential family in $\mathcal{A}$ with canonical parameter space Θ and canonical tangent space $V = \operatorname{lin}(\Theta)$. In Sec. IV B through IV D we shall specialize to the algebra $\mathcal{A}$ defined in Example 3.

A. (+1)-closures of exponential families

In this section we compute the $(+1)$-closure $\operatorname{cl}^{(+1)}(\mathcal{E})$ defined in (7). We show that it is a union of exponential families. We also discuss aspects of the rI-closure $\operatorname{cl}^{rI}(\mathcal{E})$, defined in (6) and of the norm closure $\overline{\mathcal{E}}$. Among others, we show

$$\operatorname{cl}^{(+1)}(\mathcal{E}) \;\subset\; \operatorname{cl}^{rI}(\mathcal{E}) \;\subset\; \overline{\mathcal{E}}.$$

Strict inclusions are presented by example in Secs. IV D and IV B

In this section $\mathcal{A}$ denotes an arbitrary real *-subalgebra of $\operatorname{Mat}(N, \mathbb{C})$. In the analysis of $(+1)$- and rI-closures, subalgebras with various identities will appear, so spectral values shall be used in some statements, see also Remark 6. On the space $\mathcal{A}_{sa}$ of self-adjoint matrices we have the partial ordering defined by $a \preceq b$ if and only if $b - a \succeq 0$, i.e., $b - a$ is positive semi-definite. The set of projections $\{p \in \mathcal{A} \mid p^* = p^2 = p\}$ will be considered with this partial ordering. If $p \in \mathcal{A}$ is a

projection, then the *compressed algebra* by p is

$$p\mathcal{A}p := \{p\,ap \mid a \in \mathcal{A}\}. \tag{18}$$

The algebra $p\mathcal{A}p$ is a *-subalgebra of $\mathcal{A}$ with identity p. The spectral values of $a \in \mathcal{A}_{\mathrm{sa}}$ are the real numbers λ such that $a - \lambda\mathbb{1}$ is not invertible in $\mathcal{A}$. The sum of spectral projections of non-zero spectral values of a is the *support projection* $s(a)$; we notice

$$s(a) \in \mathcal{A}. \tag{19}$$

We denote by $\lambda^+(a)$ the maximal spectral value of a and by $p^+(a) \in \mathcal{A}$ the spectral projection of a corresponding to $\lambda^+(a)$, which we call the *maximal projection* of a. Notice in Remark 6 that eigenvalues cannot be used.

The *free energy*, defined for $a \in \mathcal{A}_{\mathrm{sa}}$ by $F(a) := \ln(\mathrm{tr}(e^a))$ is useful to discuss limits of $(+1)$-geodesics. Functions defined for projections $p \in \mathcal{A}$ by functional calculus on $(p\mathcal{A}p)_{\mathrm{sa}}$ will be decorated by a superscript p, e.g. $\ln^p(p) = 0$, while $\ln(p)$ is not defined if $p \neq \mathbb{1}$. The superscript $p = \mathbb{1}$ will often be omitted. For $a \in (p\mathcal{A}p)_{\mathrm{sa}}$ we notice $\exp^p(a) = p\exp(a)$, $\exp^p_1(a) = \frac{p\,e^a}{\mathrm{tr}(p\,e^a)}$ and $F^p(a) = \ln\mathrm{tr}(p\,e^a)$. We use the projection $\mathcal{A} \to p\mathcal{A}p$, $a \mapsto pap$ to define the exponential family in $p\mathcal{A}p$

$$\mathcal{E}^p := \left\{\exp^p_1(p\theta p) \mid \theta \in \Theta\right\}.$$

Lemma 9: Suppose $\theta, u \in \mathcal{A}_{\mathrm{sa}}$ and $p := p^+(u)$ is the maximal projection of u. We have

$$\lim_{t\to\infty} \exp_1(\theta + t\,u) = \exp^p_1(p\theta p) \tag{20}$$

and

$$\lim_{t\to\infty} \left(F(\theta + t\,u) - t\,\lambda^+(u)\right) = F^p(p\theta p). \tag{21}$$

Proof: If u has maximal spectral value $\lambda^+(u) = 0$ then by standard perturbation theory one proves

$$\lim_{t\to\infty} e^{\theta + t\,u} = p\,e^{p\theta p}. \tag{22}$$

Since $\exp_1(\theta + \alpha\mathbb{1}) = \exp_1(\theta)$ holds for $\alpha \in \mathbb{R}$ we have for arbitrary $u \in \mathcal{A}_{\mathrm{sa}}$

$$\lim_{t\to\infty} \exp_1(\theta + t\,u) = \lim_{t\to\infty} \exp_1(\theta + t\,(u - \lambda^+(u)\mathbb{1})) = \frac{p\,e^{p\theta p}}{\mathrm{tr}(p\,e^{p\theta p})}.$$

If u has maximal spectral value $\lambda^+(u) = 0$ then (22) and the continuity of the logarithm show $\lim_{t\to\infty} F(\theta + t\,u) = \ln\mathrm{tr}(p\,e^{p\theta p})$. We have $F(\theta + \alpha\mathbb{1}) = F(\theta) + \alpha$ for $\alpha \in \mathbb{R}$, hence for arbitrary $u \in \mathcal{A}_{\mathrm{sa}}$ the equality of

$$F(\theta + t\,u) - t\lambda^+(u) = F[\theta + t\,(u - \lambda^+(u))]$$

shows the second claim. $\square$

An immediate consequence of (20) is as follows:

Proposition 10: The $(+1)$-closure of $\mathcal{E}$ is $\mathrm{cl}^{(+1)}(\mathcal{E}) = \bigcup_p \mathcal{E}^p$ where the disjoint union extends over the maximal projections $p = p^+(v)$ of all vectors $v \in V$ (including $\mathbb{1} = p^+(0)$).

The first hurdle to tackle the rI-closure will be Lemma 13 which controls limits of the relative entropy of certain states ρ from states σ on $(+1)$-geodesics. This is remarkable since for $\mathcal{A} = \mathrm{Mat}(2, \mathbb{C})$

$$S_\rho(\sigma) := S(\rho, \sigma)$$

is not continuous on the set $\{\sigma \in \mathcal{S} \mid s(\sigma) \succeq s(\rho)\}$ with larger support projections (19). However, S_ρ is continuous throughout the simplex $\mathcal{S}$ for $\mathcal{A} \cong \mathbb{C}^N$.

Example 11 (Discontinuity of Relative Entropy): In the algebra $\mathcal{A} = \mathrm{Mat}(2, \mathbb{C})$ of a qubit we consider the pure state $\rho := \frac{1}{2}(\mathbb{1}_2 + \sigma_1)$.

For real $\alpha > 0$ let $s_\alpha \in [0, 1]$ such that $s_\alpha \overset{\alpha \to 0}{\to} 0$ and define the state

$$\sigma_\alpha := (1 - s_\alpha)\tfrac{1}{2}(\mathbb{1}_2 + \cos(\alpha)\sigma_1 + \sin(\alpha)\sigma_2) + s_\alpha \tfrac{1}{2}(\mathbb{1}_2 - \cos(\alpha)\sigma_1 - \sin(\alpha)\sigma_2).$$

Then $\sigma_\alpha \overset{\alpha \to 0}{\to} \rho$ as well as

$$S(\rho, \sigma_\alpha) = -\tfrac{1}{4}\alpha^2 \log(s_\alpha)(1 + o(1)) + o(1).$$

E.g. if we choose $c, \gamma > 0$ and put $s_\alpha := \exp(-c/\alpha^\gamma)$, then $s_\alpha \overset{\alpha \to 0}{\to} 0$ and

$$S(\rho, \sigma_\alpha) = \tfrac{c}{4}\alpha^{2-\gamma}(1 + o(1)) + o(1).$$

So any non-negative limit of $S(\rho, \sigma_\alpha)$ can be achieved for smooth paths converging to an arbitrary point ρ in the boundary of state space.

Using maximal spectral values λ^+ and maximal projections p^+ we summarize Proposition 2.9 in Ref. 43.

Lemma 12: If $u \in \mathcal{A}_{\mathrm{sa}}$ is a non-zero self-adjoint matrix, then the exposed face $F(\mathcal{S}(\mathcal{A}), u)$ consists of the states $\rho \in \mathcal{S}(\mathcal{A})$ such that $\langle \rho, u \rangle = \lambda^+(u)$ or, equivalently, $s(\rho) \preceq p^+(u)$.

The lemma says that the exposed face $F(\mathcal{S}(\mathcal{A}), u)$ is the state space of the compressed algebra $p\mathcal{A}p$ discussed in (18) for $p := p^+(u)$. Moreover, it follows that all faces of $\mathcal{S}(\mathcal{A})$ are exposed, see e.g. Sec. 2.3 in Ref. 43.

The derivative of the exponential function for $a, b \in \mathcal{A}_{\mathrm{sa}}$ is

$$\mathrm{D}|_a \exp(b) = \int_0^1 e^{ya}be^{(1-y)a}\mathrm{d}y.$$

It implies the derivative of the free energy F

$$\mathrm{D}|_a F(b) = \langle b, \exp_1(a) \rangle. \tag{23}$$

The derivative of the exponential for $\mathcal{A} = \mathrm{Mat}(N, \mathbb{C})$ is explained by power series expansion e.g. in Ref. 24 and may be generalized to any *-subalgebra $\mathcal{A}$ of $\mathrm{Mat}(N, \mathbb{C})$ by left- and right-multiplication with the identity $\mathbb{1}$ in $\mathcal{A}$.

Lemma 13: Suppose $\theta, u \in \mathcal{A}_{\mathrm{sa}}$ such that u is not a multiple of the identity $\mathbb{1}$ in $\mathcal{A}$ and let $p := p^+(u)$. If $\rho \in F(\mathcal{S}(\mathcal{A}), u)$, then $S_\rho(\exp_1(\theta + tu))$ is strictly monotone decreasing with $t \in \mathbb{R}$ and

$$S_\rho\big(\exp_1^p(p\theta p)\big) = \lim_{t \to \infty} S_\rho\big(\exp_1(\theta + tu)\big) = \inf_{t \in \mathbb{R}} S_\rho\big(\exp_1(\theta + tu)\big).$$

Proof: By definition (15) of an exposed face we have for $\tau \in \mathcal{S}(\mathcal{A})$ and for $\rho \in F(\mathcal{S}(\mathcal{A}), u)$ the inequality $\langle u, \tau - \rho \rangle \leq 0$. Since u is not proportional to $\mathbb{1}$, its maximal projection is $p := p^+(u)$ is not $\mathbb{1}$. If τ is invertible, then $s(\tau) = \mathbb{1}$ and it follows from Lemma 12 that $\tau \notin F(\mathcal{S}(\mathcal{A}), u)$. This implies the strict inequality $\langle u, \tau - \rho \rangle < 0$ to hold for all invertible states $\tau = \exp_1(\theta + tu)$ with $t \in \mathbb{R}$. Using (23) we have for all $t \in \mathbb{R}$

$$\tfrac{\partial}{\partial \lambda} S_\rho \circ \exp_1(\theta + tu) = \langle u, \exp_1(\theta + tu) - \rho \rangle < 0.$$

We conclude that $S_\rho \circ \exp_1(\theta + tu)$ is strictly monotone decreasing in t.

The limit of the $(+1)$-geodesic $g: t \mapsto \exp_1(\theta + tu)$ is calculated in (20),

$$\sigma := \lim_{t \to \infty} g(t) = \exp_1^p(p\theta p).$$

The states ρ and σ belong to the compressed algebra $p\mathcal{A}p$ defined in (18) and σ is invertible in $p\mathcal{A}p$. Then

$$
\begin{aligned}
-S(\rho, \sigma) - S(\rho) &= \mathrm{tr}\left(\rho \, \ln^p \circ \exp_1^p(p\,\theta p)\right) = \mathrm{tr}(\rho\,\theta) - F^p(p\,\theta p) \\
&= \lim_{t\to\infty}\left[\mathrm{tr}(\rho\,\theta) + t\,\lambda^+(u) - F(\theta + t\,u)\right] \\
&= \lim_{t\to\infty}\left[\mathrm{tr}\left(\rho\,(\theta + t\,u)\right) - F(\theta + t\,u)\right] \\
&= \lim_{t\to\infty}\mathrm{tr}\left(\rho\,\ln\circ\exp_1(\theta + t\,u)\right) = \lim_{t\to\infty}\left[-S(\rho, g(t)) - S(\rho)\right].
\end{aligned}
$$

We have used (21) in the third step. The result is $\lim_{t\to\infty} S_\rho \circ g(t) = S_\rho(\sigma)$. Since $S_\rho \circ g$ is monotone decreasing in λ we have $\inf_{t\in\mathbb{R}} S_\rho \circ g(t) = S_\rho(\sigma)$. $\qquad\square$

We show that $(+1)$-closures do not decrease the entropy distance, defined in (5), from exponential families.

Proposition 14: If $v \neq 0$ *belongs to the canonical tangent space* V *of the exponential family* $\mathcal{E}$ *and* ρ *to the exposed face* $F(\mathcal{S}(\mathcal{A}), v)$*, then* $\mathrm{d}_\mathcal{E}(\rho) = \mathrm{d}_{\mathcal{E}^{p^+(v)}}(\rho)$*. For arbitrary* $\rho \in \mathcal{S}(\mathcal{A})$ *we have* $\mathrm{d}_\mathcal{E}(\rho) = \inf\{S(\rho, \sigma) \mid \sigma \in \mathrm{cl}^{(+1)}(\mathcal{E})\}$*.*

Proof: We prove the first statement, let $p := p^+(v)$. If $p^+(v) = \mathbb{1}$, then there is nothing to prove. Otherwise we have by Lemma 13 and Lemma 9

$$
\begin{aligned}
\mathrm{d}_\mathcal{E}(\rho) &= \inf_{\sigma\in\mathcal{E}} S(\rho, \sigma) = \inf_{\theta\in\Theta}\inf_{t\in\mathbb{R}} S(\rho, \exp_1(\theta + tv)) \\
&= \inf_{\theta\in\Theta} S(\rho, \lim_{t\to\infty}\exp_1(\theta + tv)) = \inf_{\theta\in\Theta} S(\rho, \exp_1^p(p\theta p)) = \mathrm{d}_{\mathcal{E}^p}(\rho).
\end{aligned}
$$

For the second statement, let $\rho \in \mathcal{S}(\mathcal{A})$ be arbitrary. By Proposition 10 it suffices to show $\mathrm{d}_{\mathcal{E}^p}(\rho) \geq \mathrm{d}_\mathcal{E}(\rho)$ for all projections p of the form $p = p^+(v)$ where $v \in V$ is non-zero. If $\rho \notin F(\mathcal{S}(\mathcal{A}), v)$, then $s(\rho) \not\preceq p$ by Lemma 12. So for all $\sigma \in \mathcal{E}^p$ we have $S(\rho, \sigma) = \infty$. Otherwise, the equality $\mathrm{d}_{\mathcal{E}^p}(\rho) = \mathrm{d}_\mathcal{E}(\rho)$ follows from the first assertion above. $\qquad\square$

Corollary 15: We have $\mathrm{cl}^{(+1)}(\mathcal{E}) \subset \mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) \subset \overline{\mathcal{E}}$.

Proof: The first inclusion follows from Proposition 14: If $\rho \in \mathrm{cl}^{(+1)}(\mathcal{E})$, then $\mathrm{d}_\mathcal{E}(\rho) = \inf\{S(\rho, \sigma) \mid \sigma \in \mathrm{cl}^{(+1)}(\mathcal{E})\} = 0$ shows $\rho \in \mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$.

The second inclusion follows from the Pinsker-Csiszár inequality (see e.g. p. 40 in Ref. 32), which says that $\|\rho - \sigma\|_1^2 \leq \frac{1}{2}S(\rho, \sigma)$ holds for all states $\rho, \sigma \in \mathcal{S}(\mathcal{A})$ with the trace norm $\|a\|_1 := \mathrm{tr}(\sqrt{a^*a})$ for $a \in \mathcal{A}$. $\qquad\square$

Finally we prove an upper bound for the norm closure of a Gibbsian family.

Lemma 16: Let $\mathcal{E}$ be a Gibbsian family, i.e. $\Theta = V$. Then $\overline{\mathcal{E}} \subset \mathcal{E} \cup \bigcup_v F(\mathcal{S}(\mathcal{A}), v)$ where the union of exposed faces extends over all non-zero vectors $v \in V$.

Proof: We assume $\theta_i \subset \Theta$ and that $\rho_i := \exp_1(\theta_i) \in \mathcal{E}$ is a converging sequence with limit $\rho := \lim_{i\to\infty}\rho_i$. If $\pi_V(\rho) \in \mathrm{ri}(\mathbb{M}(V))$, then there is a neighborhood $U(\pi_V(\rho)) \subset \mathrm{ri}(\mathbb{M}(V))$ containing $\pi_V(\rho_i)$ for large i. Choosing this neighborhood sufficiently small we can assume its closure X is a compact subset of $\mathrm{ri}(\mathbb{M}(V))$. As discussed in (11) the map $\pi_V \circ \exp_1 : V \to \mathrm{ri}(\mathbb{M}(V))$ is a real analytic diffeomorphism. Using the inverse mapping, the set $\log_0 \circ \pi_\mathcal{E}(X) \subset V$ is compact and it contains θ_i for large i. It follows $\rho \in \mathcal{E}$.

Otherwise, if $\pi_V(\rho)$ belongs to the boundary of the mean value set, then by Theorem 13.1 in Ref. 36 there is a non-zero vector $v \in V$ such that $\pi_V(\rho) \in F(\mathbb{M}(V), v)$. Then the state ρ lies in the exposed face $F(\mathcal{S}(\mathcal{A}), v)$ for the same vector v. $\qquad\square$
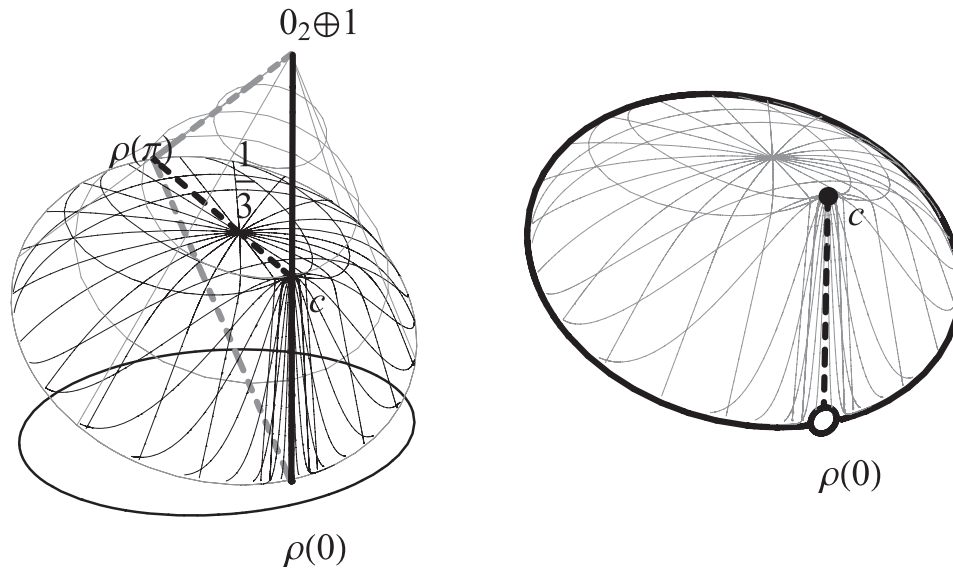
FIG. 3. The Staffelberg family $\mathcal{E}$ sketched by $(+1)$-geodesics. Left: The cone about $\mathcal{E}$ is the state space $\mathcal{S}(\mathcal{A})$. The ellipse below is the boundary of the mean value set $\mathbb{M}(V)$. The generating line $[\rho(0), 0_2 \oplus 1]$ of the cone $\mathcal{S}(\mathcal{A})$, with midpoint c, is perpendicular to V. Right: $\mathcal{E}$ has equal $(+1)$- and rI-closures, they cover the punctured base circle of $\mathcal{S}(\mathcal{A})$ (large circle) with $\rho(0)$ missing (small circle). These closures include c. The norm closure of $\mathcal{E}$ contains in addition the entire segment $[\rho(0), c]$.

B. The Staffelberg family

The exponential family $\mathcal{E}$ discussed in this section is an example of a discontinuous maximum-entropy inference announced in the Introduction. That exponential family has the form of the *Staffelberg* table mountain, in the natural preserve of *Fränkische Schweiz—Veldensteiner Forst*. Its mean value set appears at the angle (16) of $\varphi = \frac{\pi}{3}$ in the metamorphosis of Figure 2. Smaller angles φ have non-exposed faces, larger angles do not. We explain the geometrical components of the closures $\mathrm{cl}^{(+1)}(\mathcal{E}) = \mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) \subsetneq \overline{\mathcal{E}}$. Then we address continuity issues.

Definition 17: The *Staffelberg family*, depicted in Figure 3, is the Gibbsian family

$$\mathcal{E} := \exp_1\left(\mathrm{span}_{\mathbb{R}}\{\sigma_1 \oplus 0, \sigma_2 \oplus 1\}\right)$$

in the *-subalgebra $\mathcal{A} \subset \mathrm{Mat}(3, \mathbb{C})$ defined in Example 3.

The self-adjoint matrices in $\mathcal{A}$ are $\mathcal{A}_{\mathrm{sa}} = \mathrm{span}_{\mathbb{R}}\{\mathbb{1}_2 \oplus 0, \sigma_1 \oplus 0, \sigma_2 \oplus 0, 0_2 \oplus 1\}$, the state space $\mathcal{S}(\mathcal{A})$ is a 3D cone. We use the notation

$$B := \{\rho(\alpha) \mid \alpha \in (0, 2\pi)\}$$

for the *punctured base circle* of $\mathcal{S}(\mathcal{A})$ with $\rho(0) = \frac{1}{2}(\mathbb{1}_2 + \sigma_2) \oplus 0$ missing. The symmetry axis l of $\mathcal{S}(\mathcal{A})$ goes through the tracial state $\frac{1}{3}\mathbb{1}$ and through the apex $0_2 \oplus 1$, where it meets the generating lines of the cone $\mathcal{S}(\mathcal{A})$ under an angle of $\frac{\pi}{6}$. The generating line $[\rho(0), 0_2 \oplus 1]$ is perpendicular to V. We denote its midpoint by

$$c := \frac{1}{2}(\rho(0) + 0_2 \oplus 1).$$

The canonical tangent space $V = \Theta$ of $\mathcal{E}$ is spanned by $v_1 := \sigma_1 \oplus 0$ and $v_2 := \sigma_2 \oplus 1 - \frac{1}{3}\mathbb{1}$. The vector $z = -\frac{1}{2}\mathbb{1}_2 \oplus 1$ is perpendicular to v_1, so

$$\varphi = \angle(V, z) = \angle(v_2, z) = \arccos(\tfrac{1}{2}) = \tfrac{\pi}{3}$$

as claimed. The basis vectors of V connect special points in $\mathcal{S}(\mathcal{A})$,

$$v_1 = \rho(\tfrac{\pi}{2}) - \rho(\tfrac{3}{2}\pi) \quad \text{and} \quad v_2 = \tfrac{4}{3}(c - \rho(\pi)).$$

The *-algebra generated by $\sigma_2 \oplus 1$ is isomorphic to $\mathbb{R}^2$ and it has the segment $[\rho(\pi), c]$ as its state space. The $(+1)$-geodesic $\{\exp_1(\lambda v_2) \mid \lambda \in \mathbb{R}\}$ is included in $\mathcal{E}$ and it covers the invertible states in $[\rho(\pi), c]$. The *-algebra generated by $\rho(0)$, $\rho(\pi)$, and $0_2 \oplus 1$ is isomorphic to $\mathbb{R}^3$, its state space is the equilateral triangle spanned by these generators, see Figure 3, left.

For discussions of $(+1)$-geodesics in $\mathcal{E}$ we use a redundant parametrization and define for real α, s, t

$$E(\alpha, s, t) := \exp_1 \left\{ t \left[\cos(\alpha)(\sigma_2 \oplus 1) + \sin(\alpha)\sigma_1 \oplus 0 \right] \right. \tag{24}$$
$$\left. + s \left[-\sin(\alpha)(\sigma_2 \oplus 1) + \cos(\alpha)\sigma_1 \oplus 0 \right] \right\}.$$

Let $x := s\cos(\alpha) + t\sin(\alpha)$, $y := -s\sin(\alpha) + t\cos(\alpha)$, $b := \sqrt{x^2 + y^2} = \sqrt{s^2 + t^2}$ and $\eta := 2\cosh(b) + e^{-s\sin(\alpha) + t\cos(\alpha)}$. Then

$$E(\alpha, s, t) = \tfrac{1}{\eta}\left\{ [\cosh(b)\mathbb{1}_2 + \sinh(b)(x\sigma_1 + y\sigma_2)/b] \oplus e^{-s\sin(\alpha) + t\cos(\alpha)} \right\}.$$

The vectors v_1 and v_2 are completed by $v_3 := 0_2 \oplus 1 - \rho(0)$ to an orthogonal basis of the traceless matrices $\mathcal{A}_0 = V + \mathbb{R}z$. We have

$$\langle E(\alpha, s, t), \sigma_1 \oplus 0 \rangle = \tfrac{1}{\eta}\left[2\sinh(b)x/b \right] \tag{25}$$
$$\langle E(\alpha, s, t), \sigma_2 \oplus 1 \rangle = \tfrac{1}{\eta}\left[2\sinh(b)y/b + e^{-s\sin(\alpha) + t\cos(\alpha)} \right]$$
$$\langle E(\alpha, s, t), 0_2 \oplus 1 - \rho(0) \rangle = \tfrac{1}{\eta}\left[-\cosh(b) - \sinh(b)y/b + e^{-s\sin(\alpha) + t\cos(\alpha)} \right].$$

We discuss closures of the Staffelberg family and its entropy distance.

Theorem 18: *The Staffelberg family $\mathcal{E}$ has $(+1)$-closure and rI-closure equal to $\mathrm{cl}^{(+1)}(\mathcal{E})$ $= \mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) = \mathcal{E} \cup B \cup \{c\}$. The norm closure is $\overline{\mathcal{E}} = \mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) \cup [\rho(0), c]$. The entropy distance of ρ $\in [\rho(0), 0_2 \oplus 1]$ from $\mathcal{E}$ is $\mathrm{d}_{\mathcal{E}}(\rho) = S(\rho, c)$. The restricted projection $\pi_V|_{\mathrm{cl}^{\mathrm{rI}}(\mathcal{E})}$ is a bijection onto the mean value set $\mathbb{M}(V)$.*

Proof: By Proposition 10 the $(+1)$-closure of $\mathcal{E}$ is a union of exponential families $\mathcal{E}^q = \left\{ \exp_1^q(q\theta q) \mid \theta \in V \right\}$ for maximal projections q. In place of the maximal projections of $v \neq 0$ in V we consider equivalently the maximal projections of the vectors

$$u(\alpha) := \sin(\alpha)\sigma_1 \oplus 0 + \cos(\alpha)(\sigma_2 \oplus 1), \qquad \alpha \in \mathbb{R}. \tag{26}$$

There are two cases depending on the spectral projections in the orthogonal sum

$$u(\alpha) = \rho(\alpha) - \rho(\alpha + \pi) + 0_2 \oplus \cos(\alpha).$$

The maximal eigenvalue of $u(\alpha)$ is constant one. If $\alpha \neq 0 \bmod 2\pi$, then the maximal projection of $u(\alpha)$ is $\rho(\alpha)$ and has rank one. We get

$$\mathcal{E}^{\rho(\alpha)} = \left\{ \exp_1^{\rho(\alpha)}(\rho(\alpha)\theta\rho(\alpha)) \mid \theta \in V \right\} = \{\rho(\alpha)\}$$

proving $B \subset \mathrm{cl}^{(+1)}(\mathcal{E})$. If $\alpha = 0 \bmod 2\pi$, then the maximal projection of $u(0)$ is $p := \rho(0) + 0_2 \oplus 1 = 2c$. Since $p(\sigma_1 \oplus 0)p = 0$ and $p(\sigma_2 \oplus 1)p = p$ the canonical parameter space of $\mathcal{E}^p$ consists of multiples of the identity p in $p\mathcal{A}p$, $p\Theta p = pVp = \mathbb{R}p$. So is $\mathcal{E}^p = \{c\}$, we conclude $\mathrm{cl}^{(+1)}(\mathcal{E}) = \mathcal{E} \cup B \cup \{c\}$.

Lemma 16 provides an upper bound on the norm closure $\overline{\mathcal{E}}$ in terms of faces of $\mathcal{S}(\mathcal{A})$ exposed by vectors in V, and Lemma 12 describes these faces in terms of maximal projections

$$F(\mathcal{S}(\mathcal{A}), u(\alpha)) = \left\{ \rho \in \mathcal{S}(\mathcal{A}) \mid s(\rho) \preceq p^+(u(\alpha)) \right\}.$$

For $\alpha \neq 0 \bmod 2\pi$ the maximal projection $\rho(\alpha)$ of $u(\alpha)$ has rank one and the exposed face is $F(\mathcal{S}(\mathcal{A}), u(\alpha)) = \{\rho(\alpha)\}$. The projection $p^+(u(0)) = p$ above gives the segment $[\rho(0), 0_2 \oplus 1]$ $= F(\mathcal{S}(\mathcal{A}), u(0))$. We obtain

$$\overline{\mathcal{E}} \subset \mathcal{E} \cup B \cup [\rho(0), 0_2 \oplus 1].$$

The inclusions $\mathrm{cl}^{(+1)}(\mathcal{E}) \subset \overline{\mathcal{E}}$ and $B \subset \mathrm{cl}^{(+1)}(\mathcal{E})$ show $B \subset \overline{\mathcal{E}}$. We prove that exactly the part $[\rho(0), c]$ of the segment $[\rho(0), 0_2 \oplus 1]$ belongs to $\overline{\mathcal{E}}$.

We prove that at most the half segment $[\rho(0), c]$ belongs to $\overline{\mathcal{E}}$ by showing that $\mathcal{E}$ is included in the closed half space $\langle a, v_3 \rangle \leq 0$. This is sufficient because v_3 is parallel to $[\rho(0), 0_2 \oplus 1]$ and

$$\langle \rho(0), v_3 \rangle \;=\; -1\,, \qquad \langle c, v_3 \rangle \;=\; 0\,, \qquad \text{and} \qquad \langle 0_2 \oplus 1, v_3 \rangle \;=\; 1\,.$$

We look at the polar parametrization of $\mathcal{E}$, defined with (24) as

$$\mathbb{R} \times \mathbb{R}_0^+ \to \mathcal{E}\,, \quad (\alpha, t) \mapsto E(\alpha, 0, t)\,.$$

The normalization factor η is strictly positive, so $\langle E(\alpha, 0, t), v_3 \rangle \leq 0$ by (25) is equivalent to

$$z(\alpha, t) \;:=\; \eta \langle E(\alpha, 0, t), v_3 \rangle \;=\; -\cos(\alpha)\sinh(t) - \cosh(t) + e^{\cos(\alpha)t} \;\leq\; 0\,.$$

For $t = 0$ we have $z(\alpha, 0) = 0$ while for $t \geq 0$ and arbitrary $\alpha \in \mathbb{R}$ we have

$$\pm z(\alpha, t) + \tfrac{\partial}{\partial t} z(\alpha, t) \;=\; (\cos(\alpha) \pm 1)\left[e^{\cos(\alpha)t} - e^{\pm t} \right] \;\leq\; 0\,.$$

This implies $\tfrac{\partial}{\partial t} z(\alpha, t) \leq 0$ and by integration $z(\alpha, t) \leq 0$.

We show $[\rho(0), c] \subset \overline{\mathcal{E}}$. The state $\rho(0)$ lies in the closure of B so we still have to approximate for $\lambda \in (0, 1]$ the state $\tau(\lambda) := (1 - \tfrac{\lambda}{2})\rho(0) \oplus \tfrac{\lambda}{2}$ from within $\mathcal{E}$. For $t > 0$ we choose $\alpha(t) := \sqrt{\tfrac{2}{t} \ln(\tfrac{2-\lambda}{\lambda})}$. Then $\lim_{t \to \infty} \alpha(t) = 0$ and $\lim_{t \to \infty} e^{(\cos(\alpha(t))-1)t} = \tfrac{\lambda}{2-\lambda}$ hold. Expanding by e^{-t} we have

$$\lim_{t \to \infty} E(\alpha(t), 0, t) \;=\; \frac{\tfrac{1}{2}(\mathbb{1}_2 + \sigma_2) \oplus \tfrac{\lambda}{2-\lambda}}{1 + \tfrac{\lambda}{2-\lambda}} \;=\; \tau(\lambda)\,.$$

We calculate the rI-closure. This is bounded by Corollary 15 between $(+1)$- and norm closures

$$\mathrm{cl}^{(+1)}(\mathcal{E}) \;=\; \mathcal{E} \cup B \cup \{c\} \;\subset\; \mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) \;\subset\; \mathcal{E} \cup B \cup [\rho(0), c] \;=\; \overline{\mathcal{E}}\,.$$

It remains to discuss states $\rho \in [\rho(0), 0_2 \oplus 1] = F(\mathcal{S}(\mathcal{A}), u(0))$. Proposition 14 and $\mathcal{E}^p = \{c\}$ show

$$\mathrm{d}_{\mathcal{E}}(\rho) = \mathrm{d}_{\mathcal{E}^p}(\rho) = S(\rho, c)\,.$$

So $\rho \in \mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$ holds for $\rho \in [\rho(0), 0_2 \oplus 1]$ if and only if $\rho = c$. This shows $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) = \mathrm{cl}^{(+1)}(\mathcal{E})$.

We show that $\pi_V|_{\mathrm{cl}^{\mathrm{rI}}(\mathcal{E})}$ is a bijection onto $\mathbb{M}(V)$. The boundary of the mean value set $\mathbb{M}(V)$ is by (17) and by Lemma 8 equal to the ellipse

$$\partial \mathbb{M}(V) = \pi_V(B \cup \{\rho(0)\}),$$

so π_V restricted to the circle $B \cup \{\rho(0)\}$ is a bijection. Since c lies on the segment $[\rho(0), 0_2 \oplus 1]$ which is perpendicular to V, it substitutes $\rho(0)$ in that bijection. Another bijection is the mean value chart $\pi_V|_{\mathcal{E}} : \mathcal{E} \to \mathrm{ri}(\mathbb{M}(V))$, see (11). The two latter bijections assembled prove the claim. $\quad\square$

Corollary 19: The entropy distance $\mathrm{d}_{\mathcal{E}} : \mathcal{S}(\mathcal{A}) \to [0, \log(3)]$ *from the Staffelberg family is discontinuous at* $\rho(0)$.

Proof: By the previous theorem we have $\mathrm{d}_{\mathcal{E}}(\rho(0)) = S(\rho(0), c) = \ln(2)$ while $\mathrm{d}_{\mathcal{E}} \equiv 0$ on the punctured base circle B of the cone $\mathcal{S}(\mathcal{A})$. But $\rho(0) \in \overline{B}$. $\quad\square$

Corollary 20: The mean value parametrization $\pi_{\mathcal{E}} : \mathrm{ri}(\mathbb{M}(V)) \to \mathcal{E}$ *of the Staffelberg family has no continuous extension to the mean value set* $\mathbb{M}(V)$; *it has no continuous extension to* $\pi_V(\rho(0))$.

Proof: Since the segment $[\rho(0), 0_2 \oplus 1]$ belongs to the norm closure of $\mathcal{E}$ and since this segment is perpendicular to V, the mean value parametrization $\pi_{\mathcal{E}} : \mathrm{ri}(\mathbb{M}(V)) \to \mathcal{E}$ does not extend continuously to $\pi_V(\rho(0))$. $\quad\square$

We address the maximum-entropy principle.

Theorem 21: *The rI-closure of the Staffelberg family is a set of maximum-entropy density matrices,* $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) = \{\mathrm{argmax}_{\rho \in F(v)} S(\rho) \mid v \in \mathbb{M}(V)\}$. *This holds for fibers* $F(v) := (v + V^{\perp}) \cap \mathcal{S}(\mathcal{A})$ *as well as for*

$$F(v) := (v + V^{\perp}) \cap \mathcal{S}(\mathrm{Mat}(3, \mathbb{C})).$$

Proof: Since the Staffelberg family $\mathcal{E}$ is included in the state space $\mathcal{S}(\mathcal{A})$, the Pinsker-Csiszár inequality, recalled in Corollary 15, shows that $\mathcal{E}$ has the same rI-closure in both algebras $\mathcal{A}$ and $\mathrm{Mat}(3, \mathbb{C})$. The mean value chart (11) shows that the mean value set $\mathbb{M}(V)$ is the same for both algebras. So the bijection $\pi_V|_{\mathrm{cl}^{\mathrm{rI}}(\mathcal{E})}$ from the rI-closure onto the mean value set, proved in Theorem 18, also applies to both algebras.

We discuss the inverse $\mathbb{M}(V) \to \mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$. Its restriction to the interior of the mean value set $\mathrm{ri}(\mathbb{M}(V)) \to \mathcal{E}$ is the mean value parametrization of $\mathcal{E}$ and this is known to have the maximum-entropy property (14).

Let us now consider the boundary of the mean value set $\mathbb{M}(V)$, which is by (17) and by Lemma 8 equal to the ellipse

$$\partial \mathbb{M}(V) = \pi_V\big(B \cup \{\rho(0)\}\big).$$

The fibers $F(\widetilde{v})$ for points $\widetilde{v} \in \partial\mathbb{M}(V)$ are faces of the state space $\mathcal{S}(\mathcal{A})$, see Sec. 5 in Ref. 42. Indeed they are the set of state space faces $F(\mathcal{S}(\mathcal{A}), v)$ which are exposed by a non-zero $v \in V$. Using Lemma 12 and consulting the list of maximal projections of vectors $v \in V$ in Theorem 18 these faces are the points on the punctured circle B and the segment $[\rho(0), 0_2 \oplus 1]$. Maximizers of the von Neumann entropy on these fibers are the points on B and the centroid c in the segment. This set completes $\mathcal{E}$ to its rI-closure by Theorem 18.

In the larger C*-algebra $\mathrm{Mat}(3, \mathbb{C})$ the projection $\rho(0) + 0_2 \oplus 1$ corresponds to the face $\{\rho \in \mathcal{S}(\mathrm{Mat}(3, \mathbb{C})) \mid s(\rho) \preceq \rho(0) + 0_2 \oplus 1\}$ which is isomorphic to the Bloch ball. So the maximizer of the von Neumann entropy in the fiber $(v + V^{\perp}) \cap \mathcal{S}(\mathrm{Mat}(3, \mathbb{C}))$ is c as before. $\square$

We draw two conclusions about asymptotics of the maximum-entropy inference defined in (1)–(3).

Remark 22: First, there exist invertible states ρ that such that $\widehat{\rho}(n)$ has variance of order $\mathcal{O}(1/n)$ with an arbitrary large constant.

We consider the maximum-entropy inference with respect to the observables $a_1 := \sigma_1 \oplus 0$ and $a_2 := \sigma_2 \oplus 1 - \mathbb{1}_3/3$ spanning the canonical tangent space V of the Staffelberg family. To start with, let the state ρ of the quantum system be on the segment $[\rho(0), 0_2 \oplus 1]$, e.g. equal to the maximum-entropy state $c = \frac{1}{2}(\rho(0) + 0_2 \oplus 1)$, and recall the notation $\rho(0) = \frac{1}{2}(\mathbb{1}_2 + \sigma_2) \oplus 0$ in Example 3. The random variable $\overline{a_2}(1)$ becomes sharp $\frac{2}{3}$. The random variable $\overline{a_1}(1)$ becomes sharp 0 in the state $0_2 \oplus 1$ and uniformly distributed on $\{-1, +1\}$ in the state $\rho(0)$. It is governed by the probability vector $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ if the system is in the state c.

Let ρ be invertible and close to c. The random variable $\overline{a_1}(n)$ has variance $\frac{1}{n}$ times the variance of $\overline{a_1}(1)$, the latter is $\approx \frac{1}{2}$. Mean value tuples $m(a) = (m_1(a), m_2(a)) := (\langle a_1, a \rangle, \langle a_2, a \rangle)$ are assigned to self-adjoint matrices a in particular to density matrices. Since the variance of $\overline{a_1}(n)$ is $\approx \frac{1}{2}/n$, it suffices to show that the partial derivative of the inference in the m_1-direction diverges along a sequence of points converging to $m(c) = (0, \frac{2}{3})$. We consider the linear functional $f : \mathcal{A}_{\mathrm{sa}} \to \mathbb{R}$, $a \mapsto \langle 0_2 \oplus 1 - \rho(0), a \rangle$. The functional f completes the m_1-m_2-plane to an orthogonal coordinate system of the state space. Its value on the segment $[\rho(\pi), c]$ of maximum-entropy states is constant to zero. The mean value tuples of this segment lie on a segment s in the m_1-m_2 space, pointing in the m_2-direction. The mean value tuples of the elliptic arc of rank-two states σ having $f(\sigma) = -\frac{1}{2}$, they all lie on an elliptic arc e in the m_1-m_2 space and $s \cap e = \{m(c)\}$. By the geometry of the cone $\mathcal{S}(\mathcal{A})$, the f-value of the inference of points on $e \setminus \{m(c)\}$ is less than or equal $-\frac{1}{2}$.

Through each point x on e we draw the perpendicular to s, it meets s in a point y. The line through the inference of x and the inference of y has a slope of at least $\frac{1}{2}/\|y - x\|$. By the mean

value theorem, the partial derivative of the inference in the m_1-direction is at least $\frac{1}{2}/\|y - x\|$ at some point on the open segment $]x, y[$. The claim follows by letting x converge to $m(c)$.

Second, the non-generic choice of ρ on the segment $[\rho(0), 0_2 \oplus 1]$ makes it possible that $\widehat{\rho}(n)$ diverges or converges to a state which is not a maximum-entropy state. This is suggested by Theorems 18 and 21 because the whole segment $[\rho(0), c]$ belongs to the closure of $\mathcal{E}$ while only c is a state of maximum entropy under the given constraints. It will be interesting to calculate the probability distribution of limits of $\widehat{\rho}(n)$ on the segment $[\rho(0), c]$.

C. (+1)-asymptotics and (−1)-closure of the Staffelberg family

We show that the (-1)-closure of the Staffelberg family $\mathcal{E}$ equals its rI-closure. This follows from an asymptotic analysis of its $(+1)$-geodesics. See (7) and (8) for definitions of these closures.

We use the parametrization $E(\alpha, s, t)$ of $\mathcal{E}$ defined in (24) and a coordinate system spanned by $(\sigma_2 \oplus 1)$ and $(\sigma_1 \oplus 0)$. Coefficients of points on $\mathcal{E}$ are the first two numbers in (25), they describe projection onto V:

$$g := \langle E(\alpha, s, t), \sigma_2 \oplus 1 \rangle = \tfrac{1}{\eta}\left[(e^b - e^{-b})y/b + e^{-s\sin(\alpha)+t\cos(\alpha)}\right]$$

$$h := \langle E(\alpha, s, t), \sigma_1 \oplus 0 \rangle = \tfrac{1}{\eta}\left[(e^b - e^{-b})x/b\right].$$

We consider the asymptotic slope in the $(\sigma_2 \oplus 1)$-$(\sigma_1 \oplus 0)$-coordinate system

$$\kappa(\alpha, s) := \lim_{t \to \infty} \frac{\mathrm{d}h}{\mathrm{d}g} = \lim_{t \to \infty} \frac{\frac{\mathrm{d}h}{\mathrm{d}t}}{\frac{\mathrm{d}g}{\mathrm{d}t}} = \lim_{t \to \infty} \frac{\eta \frac{\mathrm{d}(h\eta)}{\mathrm{d}t} - (h\eta)\frac{\mathrm{d}\eta}{\mathrm{d}t}}{\eta \frac{\mathrm{d}(g\eta)}{\mathrm{d}t} - (g\eta)\frac{\mathrm{d}\eta}{\mathrm{d}t}} . \tag{27}$$

The coordinates $\{(\langle \rho, \sigma_2 \oplus 1 \rangle, \langle \rho, \sigma_1 \oplus 0 \rangle) \mid \rho \in \mathcal{S}(\mathcal{A})\}$ of the mean value set fill the unit disk. Projections of $(+1)$-geodesics hit the unit circle for $s = 0$, they are tangential to the unit circle for every $s \neq 0$:

Lemma 23: For all $\alpha \in \mathbb{R}$ and all $s \in \mathbb{R}$ we have $(g, h) \overset{t \to \infty}{\longrightarrow} (\cos(\alpha), \sin(\alpha))$. The asymptotic slope of $(+1)$-geodesics through the tracial state $(s = 0)$ is

$$\kappa(\alpha, 0) = \begin{cases} 0 & \text{if } \alpha = 0, \\ -\cot(\tfrac{\alpha}{2}) & \text{if } \alpha \neq 0. \end{cases}$$

The asymptotic slope of $(+1)$-geodesics missing the tracial state $(s \neq 0)$ is

$$\kappa(\alpha, s) = -\cot(\alpha).$$

Proof: The $(+1)$-geodesic limit $t \to \infty$ follows from (20) and from the discussion of maximal projections in Theorem 18. Then $\lim_{t \to \infty}(g, h)$ follows.

We first compute the asymptotical slope for $(+1)$-geodesics through the tracial state $s = 0$. We have

$$(\eta \tfrac{\mathrm{d}(h\eta)}{\mathrm{d}t} - (h\eta)\tfrac{\mathrm{d}\eta}{\mathrm{d}t})e^{-t(1+\cos(\alpha))}$$
$$= \sin(\alpha)(1 + e^{-2t} + 4e^{-t-t\cos(\alpha)} - \cos(\alpha) + e^{-2t}\cos(\alpha))$$

and

$$(\eta \tfrac{\mathrm{d}(g\eta)}{\mathrm{d}t} - (g\eta)\tfrac{\mathrm{d}\eta}{\mathrm{d}t})e^{-t(1+\cos(\alpha))}$$
$$= -(1 - \cos(\alpha))^2 + e^{-2t} + \cos(\alpha)(2e^{-2t} + 4e^{-t-t\cos(\alpha)} + e^{-2t}\cos(\alpha)).$$

From this and (27) we get the desired result, studying $\alpha = 0$ and $\alpha = \pi$ apart.
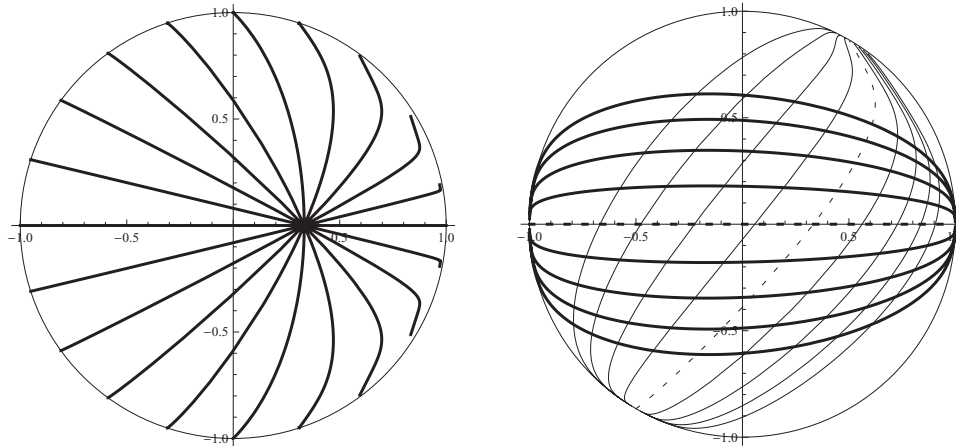
FIG. 4. Projected $(+1)$-geodesics in the Staffelberg family. Left: geodesics through the tracial state; Right: two families of parallel geodesics, those through the tracial state are dashed.

The asymptotical slope for $(+1)$-geodesics missing the tracial state $(s \neq 0)$ follows from a third order Taylor expansion at $t = \infty$. If $\alpha \neq 0$ modulo 2π then

$$(\eta \tfrac{\mathrm{d}(h\eta)}{\mathrm{d}t} - (h\eta)\tfrac{\mathrm{d}\eta}{\mathrm{d}t}) = -\frac{s}{t^2}\cos(\alpha) + O(\tfrac{1}{t^3})$$

$$(\eta \tfrac{\mathrm{d}(g\eta)}{\mathrm{d}t} - (g\eta)\tfrac{\mathrm{d}\eta}{\mathrm{d}t}) = \frac{s}{t^2}\sin(\alpha) + O(\tfrac{1}{t^3}).$$

For $\alpha = 0$ we have

$$(\eta \tfrac{\mathrm{d}(h\eta)}{\mathrm{d}t} - (h\eta)\tfrac{\mathrm{d}\eta}{\mathrm{d}t}) = -\frac{2s}{t^2} + O(\tfrac{1}{t^3})$$

$$(\eta \tfrac{\mathrm{d}(g\eta)}{\mathrm{d}t} - (g\eta)\tfrac{\mathrm{d}\eta}{\mathrm{d}t}) = O(\tfrac{1}{t^3})$$

completing the claim.   □

Some projected $(+1)$-geodesics of the Staffelberg family are drawn in Figure 4. As a fact not used in the sequel, Lemma 23 shows that the two asymptotic tangents $t \to \pm\infty$ of a projected $(+1)$-geodesic through the tracial state $(s = 0)$ intersect orthogonally at $(1, 0)$ for $\alpha \neq 0, \pi$. While the right angle is not invariant under affine reparametrizations, these tangents intersect in V at the projection of the cliff $c = \frac{1}{2}(\rho(0) + 0_2 \oplus 1)$ of the Staffelberg family.

Lemma 24. For all $s \in [-1, 1]$ and all $t \geq 1$ we have uniformly in s

$$\| E(0, s, t) - c \|_2 = \mathcal{O}(t^{-1}).$$

Proof: By Taylor expansion $b = t + \frac{s^2}{2t} + \mathcal{O}(t^{-2})$, we have uniformly for $s \in [-1, 1]$

$$E(0, s, t) = \begin{pmatrix} \cosh(b) & (s - it)\frac{\sinh(b)}{b} & 0 \\ (s + it)\frac{\sinh(b)}{b} & \cosh(b) & 0 \\ 0 & 0 & e^t \end{pmatrix} \Big/ \left(2\cosh(b) + e^t\right) = c + \mathcal{O}(t^{-1}).$$

This proves the statement, since $\|a\|_2 = \sqrt{\sum_{k,\ell} |a_{k,\ell}|^2}$.   □

Theorem 25: *For the Staffelberg family $\mathcal{E}$ the (-1)-closure equals the $(+1)$- and the rI-closure,* $\mathrm{cl}^{(-1)}(\mathcal{E}) = \mathrm{cl}^{(+1)}(\mathcal{E}) = \mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$.

Proof: The equality $\mathrm{cl}^{(+1)}(\mathcal{E}) = \mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$ was shown in Theorem 18. Since (-1)-geodesics are included in $\mathcal{E}$ we clearly have $\mathrm{cl}^{(-1)}(\mathcal{E}) \subset \overline{\mathcal{E}}$. On the other hand, in every fiber $(v + V^\perp) \cap \mathcal{S}(\mathcal{A})$ with

$v \in \mathbb{M}(V)$ there is at least one point of the (-1)-closure (choose a segment $]u, v[\subset \text{ri}(\mathbb{M}(V))$ and lift it to $\mathcal{E}$ through the mean value parametrization). By Theorem 18 there is a bijection

$$\pi_V|_{\overline{\mathcal{E}}\setminus S} : \overline{\mathcal{E}} \setminus S \to \mathbb{M}(V) \setminus \{m\}$$

for the segment $S := [\rho(0), 0_2 \oplus 1]$ and its projection $m := \pi_V(c)$. The three arguments combined show $\overline{\mathcal{E}} \setminus S = \text{cl}^{(-1)}(\mathcal{E}) \setminus S$.

It remains to discuss states $\rho \in S$, whether they belong to $\text{cl}^{(-1)}(\mathcal{E})$. The point c clearly does since the unparametrized (-1)-geodesic $]\rho(\pi), c[$ belongs to $\mathcal{E}$. We finish by showing $\{c\} = S \cap \text{cl}^{(-1)}(\mathcal{E})$.

The (-1)-geodesic from $\rho(\pi)$ to c is also a $(+1)$-geodesic, parametrized for $s = 0$ by

$$g_s(t) := E(0, s, t).$$

Using (20) we see that for all real s the geodesic g_s has the limit c when $t \to +\infty$, its projection $\pi_V(g_s)$ has the limit $m = \pi_V(c)$. For $s \neq 0$ the asymptotic tangent of $\pi_V(g_s)$ is tangential to the elliptical boundary $\partial\mathbb{M}(V)$ of the mean value set by Lemma 23. This implies that the projections $\pi_V(g_{-1})$ and $\pi_V(g_{+1})$ concatenate to a closed smooth curve in $\mathbb{M}(V)$ which is tangential to $\partial\mathbb{M}(V)$ at m. Using the mean value chart (11) of $\mathcal{E}$, it is clear that this curve bounds the set

$$U := \{\pi_V(g_s(t)) \mid -1 \leq s \leq 1, t \in \mathbb{R}\} \subset \mathbb{M}(V).$$

Let h be any (-1)-geodesic in $\mathcal{E}$ with limit ρ in the segment S. If we choose any sequence $\rho_n \subset h$ such that $\rho = \lim_{n\to\infty}\rho_n$, then $\theta_n := \log_0(\rho_n)$ diverges in the norm (otherwise the contradiction $\rho \in \mathcal{E}$ follows). As the boundary of U is tangential to the ellipse $\partial\mathbb{M}(V)$ at m, there is $\epsilon > 0$ such that

$$\pi_V(h) \cap \{v \in V \mid \|v - m\|_2 < \epsilon\} \subset U.$$

So the points $\pi_V(\rho_n)$ lie in U for large n. Since the convergence of the $(+1)$-geodesics g_s to c is uniform (for $-1 \leq s \leq 1$) by Lemma 24, the states ρ_n converge to c. $\qquad\square$

D. The Swallow family

We now consider 2D families $\mathcal{E} = \exp_1(V)$ in the metamorphosis of Figure 2 that have non-exposed faces in the mean value set $\mathbb{M}(V)$. By Lemma 8 this happens for angles $\varphi(V) \in (0, \pi/3)$. We prove that the $(+1)$-closure $\text{cl}^{(+1)}(\mathcal{E})$ is too small to serve as a set of entropy maximizers under linear constraints. The problem is that the two non-exposed points of the mean value set are not covered by $\text{cl}^{(+1)}(\mathcal{E})$ in the projection onto V. Calculations become easy for $\varphi = \arccos(\sqrt{2/5}) \approx 0.28\pi$ and we then call $\mathcal{E}$ the *Swallow family* because it looks like the beak of a bird:

Definition 26: The *Swallow family*, depicted in Figure 5, is the Gibbsian family

$$\mathcal{E} := \exp_1\big(\text{span}_\mathbb{R}\{\sigma_1 \oplus 1, \sigma_2 \oplus 1\}\big)$$

in the $*$-subalgebra $\mathcal{A} \subset \text{Mat}(3, \mathbb{C})$ defined in Example 3.

The canonical tangent space $V = \Theta$ of $\mathcal{E}$ is spanned by the vectors of equal length $v_1 := \sigma_1 \oplus 1 - \frac{1}{3}\mathbb{1}$ and $v_2 := \sigma_2 \oplus 1 - \frac{1}{3}\mathbb{1}$. The vector $z = -\frac{1}{2}\mathbb{1}_2 \oplus 1$ is perpendicular to $v_2 - v_1$, so indeed

$$\varphi = \angle(V, z) = \angle(v_1 + v_2, z) = \arccos(\sqrt{2/5}).$$

The pure states $\rho(0) = \frac{1}{2}(\mathbb{1}_2 + \sigma_2) \oplus 0$ and $\rho(\frac{\pi}{2}) = \frac{1}{2}(\mathbb{1}_2 + \sigma_1) \oplus 0$ on the base circle of the conic state space $\mathcal{S}(\mathcal{A})$ are crucial for the Swallow family.

Theorem 27: *The $(+1)$-closure of the Swallow family $\mathcal{E}$ is the union of $\mathcal{E}$, of the segments $]\rho(0), 0_2 \oplus 1[$ and $]\rho(\frac{\pi}{2}), 0_2 \oplus 1[$ (rank-two states) and of the pure states $0_2 \oplus 1$ and $\{\rho(\alpha) \mid \frac{\pi}{2} < \alpha < 2\pi\}$. The (-1)-rI-and norm closures are*

$$\text{cl}^{(-1)}(\mathcal{E}) = \text{cl}^{\text{rI}}(\mathcal{E}) = \overline{\mathcal{E}} = \text{cl}^{(+1)}(\mathcal{E}) \cup \{\rho(0), \rho(\tfrac{\pi}{2})\}.$$
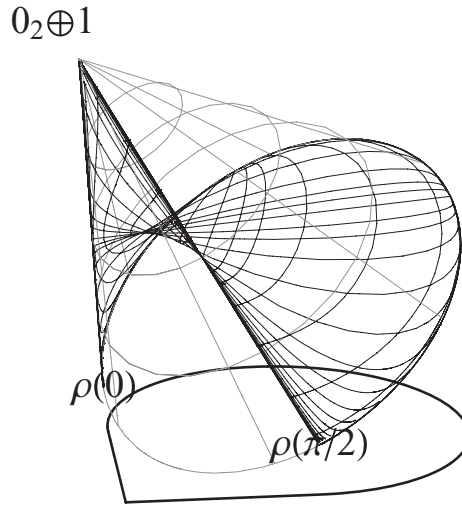
$$0_2 \oplus 1$$

FIG. 5. The Swallow family $\mathcal{E}$ sketched by $(+1)$-geodesics. The cone about $\mathcal{E}$ is the state space $\mathcal{S}(\mathcal{A})$. Its generating lines $[\rho(0), 0_2 \oplus 1]$ and $[\rho(\frac{\pi}{2}), 0_2 \oplus 1]$ belong to the rI-closure of $\mathcal{E}$ but the pure states $\rho(0)$ and $\rho(\frac{\pi}{2})$ do not belong to the $(+1)$-closure of $\mathcal{E}$. They project to the non-exposed points of the mean value set $\mathbb{M}(V)$ whose boundary is drawn below.

Proof: First we calculate the $(+1)$-closure $\mathrm{cl}^{(+1)}(\mathcal{E})$ using Proposition 10. For $\alpha \in \mathbb{R}$ we have the orthogonal sum

$$u(\alpha) := \sin(\alpha)(\sigma_1 \oplus 1) + \cos(\alpha)(\sigma_2 \oplus 1) = \rho(\alpha) - \rho(\alpha + \pi) + 0_2 \oplus \sqrt{2}\cos(\alpha - \tfrac{\pi}{4}).$$

The maximal projections for $\alpha = 0$ and $\frac{\pi}{2}$ are

$$p := p^+(u(0)) = \rho(0) + 0_2 \oplus 1 \quad \text{and} \quad q := p^+(u(\tfrac{\pi}{2})) = \rho(\tfrac{\pi}{2}) + 0_2 \oplus 1.$$

For $0 < \alpha < \frac{\pi}{2}$ we have $p^+(u(\alpha)) = 0_2 \oplus 1$ and for $\frac{\pi}{2} < \alpha < 2\pi$ we have $p^+(u(\alpha)) = \rho(\alpha)$.

Calculating the corresponding exponential families we observe $p\mathcal{A}p \cong \mathbb{C}^2$ and since $p(\sigma_1 \oplus 1)p = 0_2 \oplus 1$, the exponential family $\mathcal{E}^p = \exp_1^p(p\Theta p)$ has the canonical parameter space

$$\mathbb{R}(0_2 \oplus 1 - \rho(0)) \cong \mathbb{R}(1, -1) \subset \mathbb{C}^2.$$

The analogue arguments apply to q, so the exponential family

$$\mathcal{E}^p = \,]\rho(0), 0_2 \oplus 1[\quad \text{resp.} \quad \mathcal{E}^q = \,]\rho(\tfrac{\pi}{2}), 0_2 \oplus 1[$$

consists of the invertible states in the compressed algebra $p\mathcal{A}p$ resp. $q\mathcal{A}q$. All other maximal projections r of elements of $v \neq 0$ of V have rank one and produce the exponential family $\mathcal{E}^r = \left\{ \exp_1^r(r\theta r) \mid \theta \in V \right\} = \{r\}$. This completes the calculation of the $(+1)$-closure of $\mathcal{E}$.

In the second step we prove that the points $\rho(0)$ and $\rho(\frac{\pi}{2})$ missing in the $(+1)$-closure belong to the rI-closure of $\mathcal{E}$. Lemma 12 describes the exposed face $F(\mathcal{S}(\mathcal{A}), u(0)) = [\rho(0), 0_2 \oplus 1] = \mathcal{S}(p\mathcal{A}p)$, containing the pure state $\rho(0)$. Then Proposition 14 shows

$$\mathrm{d}_{\mathcal{E}}(\rho(0)) = \mathrm{d}_{\mathcal{E}^p}(\rho(0)) = \mathrm{d}_{\mathrm{cl}^{(+1)}(\mathcal{E}^p)}(\rho(0)).$$

Since

$$\mathrm{cl}^{(+1)}(\mathcal{E}^p) = \mathrm{cl}^{(+1)}(\,]\rho(0), 0_2 \oplus 1[\,) = [\rho(0), 0_2 \oplus 1]$$

we get $\mathrm{d}_{\mathcal{E}}(\rho(0)) = \mathrm{d}_{[\rho(0), 0_2 \oplus 1]}(\rho(0)) = 0$ and this implies $\rho(0) \in \mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$. The analogue arguments show $\rho(\frac{\pi}{2}) \in \mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$.

By the same method as in Theorem 18 an upper bound on the norm closure $\overline{\mathcal{E}}$ can be stated in terms of maximal projections in V. These projections are listed above, the corresponding faces are the pure state $0_2 \oplus 1$, the arc of pure states $\rho(\alpha)$ for $\frac{\pi}{2} < \alpha < 2\pi$ and the two segments $[\rho(0), 0_2 \oplus 1]$ and $[\rho(\frac{\pi}{2}), 0_2 \oplus 1]$ (the state spaces of the algebras $p\mathcal{A}p \cong q\mathcal{A}q \cong \mathbb{C}^2$). Thus $\overline{\mathcal{E}} \subset \mathrm{cl}^{(+1)}(\mathcal{E}) \cup \{\rho(0), \rho(\frac{\pi}{2})\}$ follows from the above description of the $(+1)$-closure. Since $\rho(0)$ and $\rho(\frac{\pi}{2})$ belong

to the rI-closure and since $\mathrm{cl}^{(+1)}(\mathcal{E}) \subset \mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) \subset \overline{\mathcal{E}}$ holds by Corollary 15 we have shown $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$ $= \overline{\mathcal{E}} = \mathrm{cl}^{(+1)}(\mathcal{E}) \cup \{\rho(0), \rho(\frac{\pi}{2})\}$. $\qquad\qquad\square$

Theorem 28: *The projection $\pi_V|_{\mathrm{cl}^{\mathrm{rI}}(\mathcal{E})}$ is a bijection onto the mean value set $\mathbb{M}(V)$, the non-exposed points of $\mathbb{M}(V)$ are $\pi_V(\rho(0))$ and $\pi_V(\rho(\frac{\pi}{2}))$. The rI-closure of the Swallow family is a set of maximum-entropy density matrices,*

$$\mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) = \{\mathrm{argmax}_{\rho \in F(v)} S(\rho) \mid v \in \mathbb{M}(V)\}$$

for fibers $F(v) := (v + V^{\perp}) \cap \mathcal{S}(\mathcal{A})$.

Proof: The relative interiors of faces of the mean value set $\mathbb{M}(V)$ are a partition of $\mathbb{M}(V)$.[36] Each face F of $\mathbb{M}(V)$ is the projection to V of the inverse projection $(F + V^{\perp}) \cap \mathcal{S}(\mathcal{A})$, which is a face of $\mathcal{S}(\mathcal{A})$. The relative interior of the inverse projection of F projects onto the relative interior of F; we show that these projections are bijections for the algebra $\mathcal{A}$, for the Swallow family $\mathcal{E}$ and for all faces F in the boundary of the mean value set $\mathbb{M}(V)$.

The two non-exposed points $\pi_V(\rho(0))$ and $\pi_V(\rho(\frac{\pi}{2}))$ at the ellipse with corner $\mathbb{M}(V)$ are computed in case 3 of Example 1.2 in Ref. 43 studying tangents. The present setting fits into Example 1.2 in Ref. 43 by choosing there $g := \frac{1}{\sqrt{2}}(1, -1, 0)$ and $h := \frac{1}{\sqrt{2}}(1, 1, 0)$. The inverse projections $(\rho(0) + V^{\perp}) \cap \mathcal{S}(\mathcal{A})$ and $(\rho(\frac{\pi}{2}) + V^{\perp}) \cap \mathcal{S}(\mathcal{A})$ are faces of the state space $\mathcal{S}(\mathcal{A})$ and it is proved in case 3 of Sec. 3.3 in Ref. 43 that these faces are the extremal points $\rho(0)$ and $\rho(\frac{\pi}{2})$ and that they are not larger.

Every exposed face $F = F(\mathbb{M}(V), v)$ for non-zero $v \in V$ is actually the projection of the exposed face $F(\mathcal{S}(\mathcal{A}), v)$, see Sec. 3.1 in Ref. 43. These faces are computed in the last paragraph of Theorem 27. A missing bijectivity of their projections onto V is only possible for the two segments, but it does not occur because the two segments cover the two boundary segments of $\mathbb{M}(V)$.

The maximum-entropy problem is solved for points in $\mathrm{ri}(\mathbb{M}(V))$ in (14). Since the projection of $(\partial\mathbb{M}(V) + V^{\perp}) \cap \mathcal{S}(\mathcal{A})$ onto V is a bijection onto $\partial\mathbb{M}(V)$, the maximum-entropy problem is trivial for boundary points of $\mathbb{M}(V)$. $\qquad\qquad\square$

Remark 29:

a) The Swallow family is suitable to demonstrate that the extreme points of a mean value set $\mathbb{M}(V)$ are in general not covered by the projections $\pi_V(\frac{p}{\mathrm{tr}(p)})$ for maximal projections $p = p^+(v)$, $v \in V$, as is claimed in Theorem 1 (e) in Ref. 45.

    Let $\mathcal{B}$ denote one of the algebras $\mathcal{A}$ or $\mathrm{Mat}(3, \mathbb{C})$ where $\mathcal{A} \subset \mathrm{Mat}(3, \mathbb{C})$ is the *-subalgebra defined in Example 3. Since $\mathcal{A}$ and $\mathrm{Mat}(3, \mathbb{C})$ have the same identities $\mathbb{1} = \mathbb{1}_3$ we can argue with eigenvalues to calculate the maximal projections of vectors in V. Moreover, the mean value set $\mathbb{M}_{\mathcal{B}}(V)$ is well-defined, see Sec. 3.4 in Ref. 43. For faces F of the mean value set the lifted faces $(F + V^{\perp}) \cap \mathcal{S}(\mathcal{B})$ are of the form $\{\rho \in \mathcal{S}(\mathcal{B}) \mid s(\rho) \preceq p\}$ for projections $p \in \mathcal{B}$, see Sec. 2.3 in Ref. 43. The necessary projections p are computed recursively from V, see Theorem 3.7 or Remark 3.10 in Ref. 43. This gives the same set of projections for both algebras $\mathcal{A}$ and $\mathrm{Mat}(3, \mathbb{C})$.

    Now, the pure state $\rho(0)$ (and $\rho(\frac{\pi}{2})$) is not on the list of maximal projections of vectors in V provided in the first paragraph of Theorem 27. On the other hand, as discussed in the second paragraph of Theorem 28, the state $\rho(0)$ is the unique state in $\mathcal{S}(\mathcal{A})$ that projects to the non-exposed point $\pi_V(\rho(0))$ of the mean value set.

b) There is no $(+1)$-geodesic in the Swallow family $\mathcal{E}$ that meets $\rho(0)$ asymptotically. Calculation of $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$ in Theorem 27 is done by two limits of $(+1)$-geodesics. One of the limits is implicit in the equation $\mathrm{d}_{\mathcal{E}}(\rho(0)) = \mathrm{d}_{\mathcal{E}^p}(\rho(0))$. Only a second $(+1)$-geodesic in $\mathcal{E}^p$ meets $\rho(0)$ asymptotically.

V. MAXIMIZERS OF THE ENTROPY DISTANCE

We now study local maximizers of the entropy distance $\mathrm{d}_{\mathcal{E}}$ from an exponential family $\mathcal{E}$, a question which was motivated in Sec. I A in the context of infomax principles. We have to restrict

to Gibbsian families since the mean value chart (11) is only available for these exponential families in the present article.

We show that a local maximizer ρ of $d_{\mathcal{E}}$ carries a clear imprint from its projection $\pi_{\mathcal{E}}(\rho)$ to $\mathcal{E}$. This generalizes the commutative case, where ρ is the conditional probability distribution of $\pi_{\mathcal{E}}(\rho)$ conditioned on its own support $\mathrm{supp}(\rho)$

$$\rho = \pi_{\mathcal{E}}(\rho)(\cdot \,|\mathrm{supp}(\rho)) . \tag{28}$$

Remark 30: In the commutative case the assertion (28) was proved for a local maximizer $\rho \in \mathrm{dom}\mathcal{E} = \mathcal{S}(\mathcal{A}) \cap (\mathcal{E} + V^{\perp})$ in Ref. 5. The articles[6, 26, 27, 35] contain further characterizations of local and global maximizers that can be interesting also in the non-commutative case.

The derivative of the logarithm is derived for $\mathcal{A} = \mathrm{Mat}(N, \mathbb{C})$ in Ref. 24. It may be generalized to any *-subalgebra $\mathcal{A}$ of $\mathrm{Mat}(N, \mathbb{C})$ using an algebra embedding $\phi : \mathcal{A} \to \mathrm{Mat}(n, \mathbb{C})$ such that $\phi(\mathbb{1})$ is invertible. If $p \in \mathcal{A}$ is a projection then for invertible $\rho \in \mathcal{S}(p\mathcal{A}p)$ and self-adjoint $u \in p\mathcal{A}p$ we have

$$\mathrm{D}|_{\rho} \ln^{p}(u) = \int_{0}^{\infty} (\rho + sp)^{-1} u (\rho + sp)^{-1} \mathrm{d}s . \tag{29}$$

Here we denote functions in $p\mathcal{A}p$ by a superscript like in the paragraph before Lemma 9.

Theorem 31: *Suppose $\mathcal{A}$ is a *-subalgebra of $\mathrm{Mat}(N, \mathbb{C})$ and $\mathcal{E}$ a Gibbsian family in $\mathcal{A}$ with canonical tangent space V. Let $\rho \in \mathrm{dom}\,\mathcal{E}$, let p denote the support projection of ρ and put $\theta := \ln_{0} \circ \pi_{\mathcal{E}}(\rho) \in V$. If u is a traceless self-adjoint matrix in $p\mathcal{A}p$, then $\mathrm{D}|_{\rho}d_{\mathcal{E}}(u) = \langle u, \ln^{p}(\rho) - \theta \rangle$. If ρ is a local maximizer of $d_{\mathcal{E}}$, then $\rho = \exp_{1}^{p}(p\,\theta\,p)$ and $d_{\mathcal{E}}(\rho) = F(\theta) - F^{p}(p\,\theta\,p)$.*

Proof: As discussed in the paragraph following (11), the mean value parametrization $\pi_{\mathcal{E}}$ defined for $a \in \mathcal{E} + V^{\perp}$ by intersection $a \mapsto (a + V^{\perp}) \cap \mathcal{E}$ is real analytic. This gives a real analytic mapping

$$L : \quad \mathcal{E} + V^{\perp} \longrightarrow V, \quad a \longmapsto \ln_{0} \circ \pi_{\mathcal{E}}(a) .$$

We can use $\pi_{\mathcal{E}}(a) = \exp_{1} \circ L(a)$ and rewrite the entropy distance (13) of a state $\rho \in \mathcal{E} + V^{\perp}$ from $\mathcal{E}$ in the form

$$d_{\mathcal{E}}(\rho) = S(\rho, \pi_{\mathcal{E}}(\rho)) = S(\rho, \exp_{1} \circ L(\rho)) \tag{30}$$

$$= -S(\rho) - \mathrm{tr}(\rho \ln \circ \exp_{1} \circ L(\rho)) = -S(\rho) - \mathrm{tr}(\rho L(\rho)) + F \circ L(\rho)$$

with the free energy F and von Neumann entropy S. As ρ is invertible in the algebra $p\mathcal{A}p$, we can differentiate at ρ the logarithm $\ln^{p}$ in the direction of any self-adjoint matrix $u \in p\mathcal{A}p$. By (29) and cyclic reordering under the trace we get

$$\mathrm{D}|_{\rho}S(u) = -\langle u, \ln^{p}(\rho) \rangle - \mathrm{tr}(u) .$$

Using the derivative of the free energy (23), which is for $a, b \in \mathcal{A}$ given by $\mathrm{D}|_{a}F(b) = \langle b, \exp_{1}(a) \rangle$, the chain rule leads to

$$\mathrm{D}|_{\rho}(F \circ L)(u) = \mathrm{D}|_{L(\rho)}F \circ \mathrm{D}|_{\rho}L(u)$$

$$= \langle \mathrm{D}|_{\rho}L(u), \exp_{1} \circ L(\rho) \rangle = \langle \mathrm{D}|_{\rho}L(u), \pi_{\mathcal{E}}(\rho) \rangle .$$

Since the image of L is V we have $\mathrm{D}|_{\rho}L(u) \in V$ and thus by definition of the projection $\pi_{\mathcal{E}}$ follows $\langle \mathrm{D}|_{\rho}L(u), \pi_{\mathcal{E}}(\rho) - \rho \rangle = 0$. Differentiation of (30) in the direction of a traceless self-adjoint matrix $u \in p\mathcal{A}p$ gives

$$\mathrm{D}|_{\rho}d_{\mathcal{E}}(u) = \langle u, \ln^{p}(\rho) \rangle + \mathrm{tr}(u) - \langle u, L(\rho) \rangle - \langle \rho, \mathrm{D}|_{\rho}L(u) \rangle$$

$$+ \langle \mathrm{D}|_{\rho}L(u), \pi_{\mathcal{E}}(\rho) \rangle = \langle u, \ln^{p}(\rho) - L(\rho) \rangle .$$

This completes the asserted directional derivative.

If ρ is a local maximizer of $d_{\mathcal{E}}$, then $\ln^{p}(\rho) = p\,L(\rho)p + \lambda p$ for some real λ because p spans the orthogonal complement of the space of traceless self-adjoint matrices in $p\mathcal{A}p$. If follows that

ρ must be proportional to $p\exp(pL(\rho)p)$ as claimed. If we write $\theta := L(\rho) = \ln_0 \circ \pi_{\mathcal{E}}(\rho)$, then we have $\rho = \exp_1^p(p\,\theta\,p)$ and $\pi_{\mathcal{E}}(\rho) = \exp_1(\theta)$. We get

$$\begin{aligned}
\mathrm{d}_{\mathcal{E}}(\rho) \;=\; S(\rho, \pi_{\mathcal{E}}(\rho)) \;=\; &\mathrm{tr}[\rho(\ln^p(\rho) - \ln\circ\pi_{\mathcal{E}}(\rho))] \\
= &\mathrm{tr}\left[\rho(p\,\theta\,p - p\ln\circ\mathrm{tr}\circ\exp^p(p\,\theta\,p) - \theta + \mathbb{1}\ln\circ\mathrm{tr}\circ\exp(\theta))\right] \\
= &\ln(\mathrm{tr}(e^\theta)) - \ln(\mathrm{tr}(p\,e^{p\,\theta\,p})).
\end{aligned}$$

$\square$

ACKNOWLEDGMENTS

[1] E. M. Alfsen and F. W. Shultz, *State Spaces of Operator Algebras* (Birkhäuser, Boston, 2001).
[2] S. Amari, "Information geometry on hierarchy of probability distributions," IEEE Trans. Inf. Theory **47**, 1701–1711 (2001).
[3] S. Amari and H. Nagaoka, *Methods of Information Geometry*, Translations of Mathematical Monographs Vol. 191 (American Mathematical Society, Providence, 2000).
[4] K. M. R. Audenaert, M. Nussbaum, A. Szkoła, and F. Verstraete, "Asymptotic error rates in quantum hypothesis testing," Commun. Math. Phys. **279**, 251–283 (2008).
[5] N. Ay, "An information-geometric approach to a theory of pragmatic structuring," Ann. Probab. **30**, 416–436 (2002).
[6] N. Ay and A. Knauf, "Maximizing multi-information," Kybernetika **42**, 517–538 (2006).
[7] N. Ay, E. Olbrich, N. Bertschinger, and J. Jost, "A geometric approach to complexity," Chaos **21**, 037103 (2011).
[8] Barndorff-Nielsen, O., *Information and Exponential Families in Statistical Theory* (Wiley, New York, 1978).
[9] Baumgratz, T., Gross, D., Cramer, M., and Plenio, M. B., "Scalable reconstruction of density matrices," e-print arXiv:1207.0358 [quant-ph].
[10] Bengtsson, I., Weis, S., and Życzkowski, K., "Geometry of the set of mixed quantum states: An apophatic approach," in Proceedings of the XXX Workshop on Geometric Methods in Physics, Białowieża, 26.06–02.07.2011 (in press).
[11] Bjelaković, I. and Szkoła, A., "The data compression theorem for ergodic quantum information sources," Quantum. Inf. Process. **4**, 49–63 (2005).
[12] A. Caticha, *Entropic Inference and the Foundations of Physics*, Brazilian Chapter of the International Society for Bayesian Analysis-ISBrA, Sao Paulo, Brazil (2012).
[13] Csiszár, I. and Matúš, F., "Information projections revisited," IEEE Trans. Inf. Theory **49**, 1474–1490 (2003).
[14] Davidson, K. R., *C*-Algebras by Example* (American Mathematical Society, Providence, 1996).
[15] Ellis, R., *Entropy, Large Deviations, and Statistical Mechanics*, Classics in Mathematics (Springer, 2006).
[16] Erb, I. and Ay, N., "Multi-information in the thermodynamic limit," J. Stat. Phys. **115**, 949–976 (2004).
[17] Grünbaum, B., *Convex Polytopes*, 2nd ed. (Springer-Verlag, New York, 2003).
[18] Heinosaari, T., Mazzarella, L., and Wolf, M. M., "Quantum tomography under prior information," e-print arXiv:1109.5478 [quant-ph].
[19] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, "Quantum entanglement," Rev. Mod. Phys. **81**, 865–942 (2009).
[20] Ingarden, R. S., Kossakowski, A., and Ohya, M., *Information Dynamics and Open Systems* (Kluwer, Dordrecht, 1997).
[21] Jaynes, E. T., "Information theory and statistical mechanics I/II," Phys. Rev. **106**, 620–630; **108**, 171–190 (1957).
[22] Kojima, M., Kojima, S., and Hara, S., "Linear algebra for semidefinite programming," Sūrikaisekikenkyūsho Kōkyūroku **1004**, 1–23 (1997).
[23] Kuperberg, G., "The capacity of hybrid quantum memory," IEEE Trans. Inf. Theory **49**, 1465–1473 (2003).
[24] Lieb, E. H., "Convex trace functions and the Wigner-Yanase-Dyson conjecture," Adv. Math. **11**, 267–288 (1973).
[25] Matsuda, H., Kudo, K., Kiyoshi, N., Nakamura, R., Yamakawa, O., and Murata, T., "Mutual information of Ising systems," Int. J. Theor. Phys. **35**, 839–845 (1996).
[26] Matúš, F., "Optimality conditions for maximizers of the information divergence from an exponential family," Kybernetika **43**, 731–746 (2007).
[27] Matúš, F. and Rauh, J., "Maximization of the information divergence from an exponential family and criticality," in *Proceedings of IEEE International Symposium on Information Theory* (July 31–Aug. 5, 2011).
[28] Netzer, T., "Spectrahedra and their shadows," Habilitationsschrift, Universität Leipzig (2011).
[29] Nussbaum, M. and Szkoła, A., "An assymptotic error bound for testing multiple quantum hypothesis," Ann. Stat. **39**, 3211–3233 (2011).
[30] Petz, D., "Geometry of canonical correlation on the state space of a quantum system," J. Math. Phys. **35**, 780–795 (1994).
[31] Petz, D., "Monotone metrics on matrix spaces," Linear Algebr. Appl. **244**, 81–96 (1996).

[32] Petz, D., *Quantum Information Theory and Quantum Statistics*, Theoretical and Mathematical Physics (Springer-Verlag, Berlin, 2008).

[33] Petz, D. and Ruppert, L., "Efficient quantum tomography needs complementary and symmetric measurements," Rep. Math. Phys. **69**, 161–177 (2012).

[34] Rau, J., "Inferring the Gibbs state of a small quantum system," Phys. Rev. A **84**, 012101 (2011).

[35] Rauh, J., "Finding the maximizers of the information divergence from an exponential family," IEEE Trans. Inf. Theory **57**, 3236–3247 (2011).

[36] Rockafellar, R. T., *Convex Analysis* (Princeton University Press, Princeton, 1970).

[37] Ruskai, M. B., "Extremal properties of relative entropy in quantum statistical mechanics," Rep. Math. Phys. **26**, 143–150 (1988).

[38] Schumacher, B., "Quantum coding," Phys. Rev. A **51**, 2738–2747 (1995).

[39] Shannon, C. E., "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423 (1948).

[40] Vedral, V., Plenio, M. B., Rippin, M. A., and Knight, P. L., "Quantifying entanglement," Phys. Rev. Lett. **78**, 2275–2279 (1997).

[41] Wehrl, A., "General properties of entropy," Rev. Mod. Phys. **50**, 221–260 (1978).

[42] Weis, S., "A note on touching cones and faces," J. Convex Anal. **19**, 323–353 (2012).

[43] Weis, S., "Quantum convex support," Linear. Algebr. Appl. **435**, 3168–3188 (2011).

[44] Weis, S., "Duality of non-exposed faces," J. Convex Anal. **19**, 815–835 (2012).

[45] Wichmann, E. H., "Density matrices arising from incomplete measurements," J. Math. Phys. **4**, 884–896 (1963).

[46] Wootters, W. K. and Fields, B. D., "Optimal state-discrimination by mutually unbiased measurements," Ann. Phys. **191**, 363–381 (1989).