Journal of Convex Analysis
Volume 21 (2014), No. 2, 339–399

Information Topologies on Non-Commutative State Spaces

Stephan Weis
Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany
sweis@mis.mpg.de

Received: May 7, 2012
Revised manuscript received: January 18, 2013
Accepted: April 17, 2013

We define an *information topology* (I-topology) and a *reverse information topology* (rI-topology) on the state space of a C*-subalgebra of $\mathrm{Mat}(n,\mathbb{C})$ in terms of sequential convergence with respect to the relative entropy. Open disks with respect to the relative entropy define a base for the topology. This was not evident since Csiszár has shown in the 1960's that the analogue is wrong for probability measures on a countably infinite set. The I-topology is strictly finer than the norm topology, it disconnects the convex state space into its faces. The rI-topology is intermediate and it allows to complete two fundamental theorems of information geometry to the full state space, by taking the closure in the rI-topology. The norm topology can be too coarse for this aim but for commutative algebras it equals the rI-topology, so the difference belongs to the domain of quantum theory. We apply our results to the maximization of the von Neumann entropy under linear constraints and to the maximization of quantum correlations.

Keywords: Relative entropy, information topology, exponential family, convex support, Pythagorean theorem, projection theorem, maximum entropy, mutual information

1991 Mathematics Subject Classification: 81P45, 81P16, 54D55, 94A17, 90C26

Contents

ISSN 0944-6532 / $ 2.50 © Heldermann Verlag

1. Introduction

Pythagorean and projection theorems in information geometry make statements about the distance of a probability measure from a family of probability measures, see e.g. Amari and Nagaoka [4] §3 and Csiszár and Matúš [18] §I.C. The theorems provide a geometric frame for applications in large deviation theory or maximum-likelihood estimation. While information geometry is often confined to families of mutually absolutely continuous probability measures, some theorems have been extended [7, 16, 18, 20] using the I-/rI-convergence[1] with respect to the relative entropy, also known as Kullback-Leibler divergence. In quantum information theory, see e.g. [4, 8, 10, 31, 32, 33, 46, 50], there is also a relative entropy, the Umegaki relative entropy, and one can ask the analogue questions as in classical probability theory.

[1]Here and in the sequel "I" stands for "information" and "rI" for "reverse information".

An n-level quantum system is described by an algebra of complex $n \times n$-matrices. E.g. probability measures on the sample space $\{1, \ldots, n\}$ are represented by the commutative algebra of diagonal matrices. For a non-commutative subalgebra of the 3×3-matrices it was discovered by Weis and Knauf [61] that the above theorems of information geometry can not be extended in norm closures, which can be too large. We believe that the convex geometry of a quantum state space already makes the norm topology unsuitable, because it does not distinguish between a unit ball, a simplex or any other convex body. This failure of the norm topology was probably ignored previously because the space of probability measures on $\{1, \ldots, n\}$ is a simplex, and the norm topology does have suitable closures for the simplex. E.g. it extends the maximum-likelihood estimation for exponential families, see [7] p. 155. On the other hand, a quantum state space is a convex body but neither a ball nor a simplex [9]. It has a Lie group symmetry [10] and it is studied under the name of *free spectrahedron* [54] in the field of convex algebraic geometry.

This article has two expository sections, §2 and §3, including the main ideas and results. The preparatory section §4 follows and provides techniques for subsequent analysis. The sections §5 and §6 collect the main proofs. The dependence on the matrix representation of the algebra is investigated in §7.

As we shall see in section §2, the rI-topology is always first countable because the open disks of the relative entropy are its base. This is also true for the I-topology discussed already in §2.1 but a base of open disks does not exist for infinite-dimensional algebras. In a sense, the I-/rI-topology combines simple properties of a metric topology and a special compatibility with geometric structure (topologies are decreasing with respect to inclusion):

1. The I-topology recognizes the convex geometry of the quantum state space, which is split into the connected components of (relative interiors of) its faces.
2. The rI-topology is adjusted to information geometry, it extends the Pythagorean and the projection theorem in topological closures.
3. In the norm topology the state space appears as an arbitrary convex body.

In the setting of a non-commutative algebra of $n \times n$-matrices the I-topology is too big and the norm topology is too small to extend the Pythagorean or the projection theorem. In the commutative setting the rI-topology equals the norm topology.

In §3 we show that the rI-topology has the perfect closures to extend the Pythagorean and projection theorem to the full state space of an n-level quantum system. However, the extensions are proved using convex geometry and calculus of matrices. The connection to the rI-closure follows then in a corollary in §3.5. The Pythagorean theorem allows us to give in §3.4 the first complete solution in the literature to the maximization of the von Neumann entropy under linear constraints. This completes partial results from 1963 in Wichmann's article [63]. The projection theorem has applications in the analysis of quantum correlations.

Indeed, in §6.5 we generalize several ideas from Ay's article [5] about local maximizers of correlation into the quantum setting. An essential part of our proof of the projection theorem is to study non-exposed faces of linear images of state spaces using Grünbaum's notion of *poonem* [28]. We argue in §3.6 why poonems are needed, the conceptually equivalent concept of *access sequence* was used in [20] to study exponential families of Borel probability measures on $\mathbb{R}^d$.

A broader usefulness of the I-/rI-topology in quantum statistics and quantum hypothesis testing is not yet clarified. In contrast to the commutative setting there is no canonical choice of a computational basis and the consequences for measurement and observation will have to be taken into consideration. One advance is that the infimum of the relative entropy does not decrease under information closures (21). This is useful e.g. in the Sanov theorem in quantum hypothesis testing [14].

2. Information convergence, information topology

This is an expository section. We recall literature on information convergence and topology in §2.1 where we also define quantum state spaces and have a first discussion about topologies on the state space. After recalling some properties of the relative entropy in §2.2 we give an overview of our purely topological results for finite-level quantum systems in §2.3.

An n-level quantum system is described by the algebra $\mathrm{Mat}(n, \mathbb{C})$. We consider a C*-subalgebra $\mathcal{A}$ of $\mathrm{Mat}(n, \mathbb{C})$, i.e. a complex subalgebra $\mathcal{A}$ of $\mathrm{Mat}(n, \mathbb{C})$ closed under the adjoint map $a \mapsto a^*$. The definition of a C*-subalgebra includes completeness with respect to a norm, but this is clear in finite dimensions. We prefer the term C*-subalgebra because it reminds us of the complex field $\mathbb{C}$ and of the closure under the adjoint map. Unless otherwise stated, $\mathcal{A}$ is a C*-subalgebra of $\mathrm{Mat}(n, \mathbb{C})$.

2.1. Spaces of probability measures and quantum states

We discuss convergence with respect to the relative entropy, called *information convergence*. For finite measurable spaces we discuss the associated topology in detail. Then we generalize to finite-level quantum systems and we finish with a short discussion of infinite-dimensional commutative von Neumann algebras. All proofs and further issues follow in §5. A review of information convergence and its topology in probability theory is given in §I.C in [18].

Let $\mathcal{M}$ be a set of probability measures on a measurable space $(X, \mathcal{X})$. If $P, Q \in \mathcal{M}$ are absolutely continuous with respect to a σ-finite measure λ and $p(x)$ resp. $q(x)$ is the Radon-Nikodym derivative of P resp. Q, then the *relative entropy* is

$$D(P||Q) := \int_X p(x) \log \frac{p(x)}{q(x)} d\lambda \,. \tag{1}$$

This equals zero if and only if $P = Q$ and otherwise $D(P||Q)$ is strictly positive or $+\infty$, see [38]. The *total variation* is

$$\|P-Q\|_1 := \int_X |p(x)-q(x)|d\lambda\,. \tag{2}$$

It is well-known that total variation defines a norm on the space of signed measures having a density with respect to λ. The *Pinsker-Csiszár inequality* [26] shows

$$\|P-Q\|_1^2 \leq 2D(P||Q)\,. \tag{3}$$

Given a sequence $(P_k)_{k\in\mathbb{N}} \subset \mathcal{M}$ and a probability measure $P \in \mathcal{M}$ we have, according to [18], *I-convergence* resp. *rI-convergence* of $(P_k)_{k\in\mathbb{N}}$ to P if

$$\lim_{k\to\infty} D(P_k||P) = 0 \qquad \text{resp. } \lim_{k\to\infty} D(P||P_k) = 0\,. \tag{4}$$

Csiszár has studied these convergences in the context of the f-divergence, generalizing the relative entropy. He has proved in Theorem 3 in [17] that the *information neighborhoods*, defined for $P \in \mathcal{M}$ and $\epsilon > 0$ by

$$\{Q \in \mathcal{M} \mid D(Q||P) < \epsilon\} \qquad \text{resp. } \{Q \in \mathcal{M} \mid D(P||Q) < \epsilon\}\,, \tag{5}$$

are not a base of a topology if $(X,\mathcal{X}) = (\mathbb{N}, 2^{\mathbb{N}})$ where $2^{\mathbb{N}}$ is the power set of $\mathbb{N}$.

In spite of Csiszár's negative result it is possible to define an *I-topology* resp. *rI-topology* on $\mathcal{M}$ in terms of the convergence of (countable) sequences (4), see Dudley and Harremoës [23, 30]. Here a subset $U \subset \mathcal{M}$ is open if for each probability measure $P \in U$ and each sequence $(P_k)_{k\in\mathbb{N}} \subset \mathcal{M}$ that I- resp. rI-converges to P, there exists $N \in \mathbb{N}$ such that for all $k \geq N$ we have $P_k \in U$. It follows from (3) that the I-/rI-topology is finer than the norm topology of the total variation.

We will take the approach by Dudley and Harremoës to study information topologies for an n-level quantum system. The common ground between finite-level quantum systems and spaces of probability measures are spaces of probability measures on a finite measurable space. The *probability simplex* of a non-empty (at most) countable set X is

$$\mathbb{P}(X) := \left\{ p = (p_x)_{x\in X} \in [0,1]^X \mid \sum_{x\in X} p_x = 1 \right\}. \tag{6}$$

Elements p of $\mathbb{P}(X)$ are called *probability vectors* on X and can be identified with probability measures P on $(X, 2^X)$ using $P(A) := \sum_{x\in A} p_x$ for $A \subset X$. For a finite measurable space $X = \{1,\ldots,n\}$ the probability simplex $\mathbb{P}(X)$ is a simplex of dimension $n-1$ and the information neighborhoods (5) are a base of a topology. The rI-topology on $\mathbb{P}(X)$ equals the total variation topology, which is the restriction of the standard Euclidean topology on $\mathbb{R}^X$. The I-topology splits $\mathbb{P}(X)$ into connected components $C(X')$ of constant support $X' \subset X$,

$$C(X') := \{p \in \mathbb{P}(X) \mid p_x > 0 \iff x \in X'\}\,.$$

On each connected component the I-topology equals the norm topology. This decomposition into connected components is the stratification (44) of the probability simplex into relative interiors of its faces.

To describe the I-/rI-topology for a C*-subalgebra $\mathcal{A}$ of $\mathrm{Mat}(n,\mathbb{C})$ let us introduce some notation. We denote the identity in $\mathrm{Mat}(n,\mathbb{C})$ by $\mathbb{1}_n$ (the zero by 0_n or 0) and the identity in $\mathcal{A}$ by $\mathbb{1}$. A *state* on $\mathcal{A}$ is a complex linear functional $f:\mathcal{A}\to\mathbb{C}$, such that $f(a^*a)\geq 0$ for all $a\in\mathcal{A}$ and $f(\mathbb{1})=1$. The standard trace tr turns $\mathrm{Mat}(n,\mathbb{C})$ into a complex Hilbert space with the *Hilbert-Schmidt* inner product $\langle a,b\rangle := \mathrm{tr}(ab^*)$ for $a,b\in\mathrm{Mat}(n,\mathbb{C})$ and we use the *two-norm* $\|a\|_2 := \sqrt{\langle a,a\rangle}$. By $\mathcal{A}_{\mathrm{sa}}$ we denote the real vector space of self-adjoint matrices in $\mathcal{A}$ and $(\mathcal{A}_{\mathrm{sa}},\langle\cdot,\cdot\rangle)$ is a Euclidean vector space. We call its norm topology on any subset simply *norm topology* as all norms are equivalent in finite dimensions.

There is a one-to-one correspondence between states f on $\mathcal{A}$ and matrices in $\mathcal{A}$ which are positive semi-definite ($\rho\succeq 0$) and have trace one ($\mathrm{tr}(\rho)=1$), see e.g. Theorem 2.4.21 in [15]. The functional f and the matrix ρ are related by

$$f(a)=\langle a,\rho\rangle \quad (a\in\mathcal{A})\,. \tag{7}$$

The matrix representation ρ of f is called *density matrix* in quantum mechanics. We will use the terms of state and density matrix synonymously. The *state space* is

$$\mathcal{S}=\mathcal{S}_{\mathcal{A}} := \{\rho\in\mathcal{A}\mid \rho\succeq 0, \mathrm{tr}(\rho)=1\,\}\,. \tag{8}$$

For the commutative subalgebra $\mathcal{A}$ of complex diagonal matrices of size $n\times n$ the state space is the probability simplex (6),

$$\mathbb{P}(\{1,\ldots,n\})=\mathcal{S}_{\mathcal{A}}\subset\mathrm{Mat}(n,\mathbb{C})\,. \tag{9}$$

We will show for a possibly non-commutative C*-subalgebra $\mathcal{A}$ of $\mathrm{Mat}(n,\mathbb{C})$ that the analogues of information neighborhoods (5) are bases of two topologies. For a non-commutative algebra $\mathcal{A}$ the analogue of the rI-topology is strictly finer than the norm topology (Corollary 5.19) and it defines – unlike the norm topology – useful closures in information theory, as we shall outline in §3.

We show in Theorem 5.18.3 that the analogue of the I-topology splits $\mathcal{S}$ into connected components of states $\rho\in\mathcal{S}$ of constant support $s(\rho)\in\mathcal{A}$. Here we use the partial ordering $\preceq$ on $\mathcal{A}$ defined by $a\preceq b$ for $a,b\in\mathcal{A}$ if and only if $b-a$ is positive semi-definite. The *projection lattice* $\mathcal{P}$ of the algebra $\mathcal{A}$ is

$$\mathcal{P}=\mathcal{P}_{\mathcal{A}} := \{p\in\mathcal{A}\mid p^2=p^*=p\}\,, \tag{10}$$

its elements are *projections*. The partial ordering restricts to $\mathcal{P}$, more details about lattices are discussed in §4.1 and infinite-dimensional algebras are treated e.g. in [1]. The *support projection* of a self-adjoint matrix $a\in\mathcal{A}$ is the infimum

$$s(a) := \bigwedge\{p\in\mathcal{P}\mid pa=a\}\,.$$

Again, like for the probability simplex $\mathbb{P}(X)$, the decomposition of $\mathcal{S}$ into connected components is the stratification (44) of the state space $\mathcal{S}$ into relative interiors of its faces. Of course, $\mathcal{S}$ is homeomorphic in the norm topology to the closed Euclidean unit ball. This is a property of any convex body, i.e. compact and convex subset of Euclidean space, known as the Theorem of Sz. Nagy, see e.g. §VIII.1 in [11].

In the algebraic formalism, the measurable space $(\mathbb{N}, 2^{\mathbb{N}})$ corresponds to the von Neumann algebra of bounded sequences

$$l^\infty := \{x = (x_i)_{i\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}} \mid \sup_{i\in\mathbb{N}} |x_i| < \infty\}$$

acting by multiplication on the Hilbert space $l^2 := \{x \in l^\infty \mid \sum_{i\in\mathbb{N}} |x_i|^2 < \infty\}$ of square summable sequences. The space $l^1 := \{x \in l^\infty \mid \sum_{i\in\mathbb{N}} |x_i| < \infty\}$ of absolutely summable sequences contains the probability simplex[2] $\mathbb{P}(\mathbb{N})$. This has, for the algebra $\mathcal{A} = l^\infty$, the form (8) of a state space if we denote for $x \in l^1$ the set of inequalities $x_i \geq 0$ for all $i \in \mathbb{N}$ simultaneously by $x \succeq 0$ and if we use the *trace* $\mathrm{tr} : l^1 \to \mathbb{C}$, $x \mapsto \sum_{i\in\mathbb{N}} x_i$,

$$\mathbb{P}(\mathbb{N}) = \mathcal{S}_{\mathcal{A}} = \{x \in l^1 \mid x \succeq 0, \mathrm{tr}(x) = 1\}\,.$$

The discussion above shows that the information neighborhoods (5) do not define a topology on $\mathbb{P}(\mathbb{N})$ but two topologies are defined in terms of the convergences (4).

2.2. The relative entropy

The relative entropy is a measure of distance between states. It has an operational meaning e.g. in hypothesis testing [50]. We recall well-known convexity and continuity properties. Although the relative entropy is not continuous in the norm topology, we point out that it is continuous in the I-topology in its first argument and continuous in the rI-topology in its second argument (17).

Definition 2.1. The *relative entropy* of a density matrix $\rho \in \mathcal{S}$ from $\sigma \in \mathcal{S}$ is

$$S(\rho, \sigma) := \mathrm{tr}\,\rho(\log(\rho) - \log(\sigma)) \tag{11}$$

if $\mathrm{Im}(\rho) \subset \mathrm{Im}(\sigma)$. Otherwise $S(\rho, \sigma) := +\infty$. The logarithm can be defined by functional calculus, see Remark 4.23.3.

The relative entropy (11) satisfies $S(\rho, \sigma) \geq 0$ for all $\rho, \sigma \in \mathcal{S}$ with equality if and only if $\rho = \sigma$, see e.g. §11.3 in [50] or §11.3 of [46]. It is discontinuous (in the norm topology) in the first argument already for the algebra $\mathcal{A} = \mathbb{C}^2$ of a bit and in the second argument for the algebra $\mathcal{A} = \mathrm{Mat}(2, \mathbb{C})$ of a qubit.

[2] The probability simplex $\mathbb{P}(\mathbb{N})$ corresponds to the *normal states* on l^∞ (see e.g. Theorem 2.4.21 in [15]). The space of positive linear maps $f : l^\infty \to \mathbb{C}$ with $f(\mathbb{1}) = 1$ is strictly larger than $\mathbb{P}(\mathbb{N})$ and can be represented by bounded additive measures which are not necessarily σ-additive (see e.g. p. 89 in [62] and p. 296 in [24]).

Example 2.2. If $\mathcal{A} = \mathbb{C}^2$ then $S\left((\frac{n-1}{n}, \frac{1}{n}), (1,0)\right) = \infty$ for all $n \in \mathbb{N}$ while $S((1,0),(1,0)) = 0$. If $\mathcal{A} = \mathrm{Mat}(2,\mathbb{C})$, then for real α we have

$$S\left(\frac{1}{2}(\mathbb{1}_2 + \sigma_1), \frac{1}{2}(\mathbb{1}_2 + \cos(\alpha)\sigma_1 + \sin(\alpha)\sigma_2)\right) = \begin{cases} 0 & \text{if } \alpha = 0 \mod 2\pi, \\ \infty & \text{else.} \end{cases}$$

Example 11 in [61] is less trivial: A smooth curve $t \mapsto \sigma_t$ converging in norm to ρ on the boundary of the Bloch ball $\mathcal{S}_{\mathrm{Mat}(2,\mathbb{C})}$ can have any non-negative limit of $S(\rho, \sigma_t)$.

Let us now turn to some well-known properties of the relative entropy.

Definition 2.3. 1. A function $f : X \to (-\infty, \infty]$, defined on a convex subset X of a finite-dimensional Euclidean vector space $\mathbb{E}$, is *convex* if for $x_1, x_2 \in X$ and $\lambda \in [0,1]$

$$f((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)f(x_1) + \lambda f(x_2)\,.$$

In the special case that Y is another convex subset of $\mathbb{E}$ and $f : X \times Y \to (-\infty, \infty]$ is defined, such that for $x_1, x_2 \in X$, $y_1, y_2 \in Y$ and $\lambda \in [0,1]$ we have

$$f((1-\lambda)x_1 + \lambda x_2, (1-\lambda)y_1 + \lambda y_2) \leq (1-\lambda)f(x_1, y_1) + \lambda f(x_2, y_2)$$

then f is called *jointly convex*. A function $f : X \to \mathbb{R}$ is *strictly convex* if for $x, y \in X$, $x \neq y$ and $\lambda \in (0,1)$

$$f((1-\lambda)x + \lambda y) < (1-\lambda)f(x) + \lambda f(y)\,.$$

If f is (strictly) convex, we say that $-f$ is *(strictly) concave*.

2. If (X, d) is a metric space and $f : X \to (-\infty, \infty]$ then f is *lower semi-continuous* if for all $x \in X$ and every sequence $(x_i)_{i \in \mathbb{N}} \subset X$ converging to x we have

$$\liminf_{i \to \infty} f(x_i) \geq f(x)\,.$$

Remark 2.4. 1. The lower semi-continuity of the relative entropy (in the norm topology) is proved e.g. by Wehrl in §III.B in [57], using Lindblad's representation of the relative entropy [40]. Ohya and Petz give another proof in §5 in [47]. They use Kosaki's formula and write the relative entropy as a supremum of affine functionals.

2. The joint convexity of the relative entropy follows from Lieb's theorem [39], see e.g. §11.4 in [46] or §III in [57] for proofs and the historic context. Convexity of the relative entropy is a special case of the joint convexity of quasi-entropies [48]. Also, it follows easily from the monotonicity of the relative entropy under quantum operations, see §3.4 in [50].

3. A convex lower semi-continuous function is continuous along straight lines. More precisely, let $f : X \to (-\infty, \infty]$ be a convex and lower semi-continuous function defined on a closed convex subset X of a finite-dimensional Euclidean

vector space $\mathbb{E}$. We extend f to $\mathbb{E}$ by setting its value to $+\infty$ outside of X. Since X is convex, the extension $\widetilde{f}$ is convex. Since X is closed, $\widetilde{f}$ is lower semi-continuous. Thus, Corollary 7.5.1 in [52] shows for $x, y \in \mathbb{E}$, subject to $\widetilde{f}(x) < +\infty$, that

$$\widetilde{f}(y) = \lim_{\lambda \nearrow 1} \widetilde{f}((1-\lambda)x + \lambda y)\,.$$

Here the values of $\widetilde{f}$ converge in the Alexandroff compactification $(-\infty, \infty]$ of $\mathbb{R}$, see Example 5.12. For example, if $\tau \in \mathcal{S}$ is any invertible density matrix, then for arbitrary $\rho, \sigma \in \mathcal{S}$ we have $S(\rho, \tau) < \infty$ so

$$S(\rho, \sigma) = \lim_{\lambda \nearrow 1} S(\rho, (1-\lambda)\tau + \lambda\sigma)\,. \tag{12}$$

We use (12) in Theorem 5.18.5 to prove that the state space $\mathcal{S}$ is connected in the rI-topology[3].

2.3. New results about the I- and the rI-topology

This section summarizes properties of the I-/rI-topology of a C*-subalgebra $\mathcal{A}$ of $\mathrm{Mat}(n, \mathbb{C})$ with focus on similarities to a metric topology. Reasoning is done within the theory of sequential convergence, recalled in §5, exceptions are the Pinsker-Csiszár inequality (15) and the continuity result of (17) which are from matrix theory. Since the I-topology and the rI-topology share many properties, we use a prefix variable $\omega \in \{\mathrm{I}, \mathrm{rI}\}$ to denote

ω-topology, ω-closure, etc.

Unless otherwise specified we always use the norm topology. We end the section with an application in information theory.

Definition 2.5 (Information topology). 1. We use short-hand notation for the two possible variable orderings of the relative entropy (11),

$$S^{\mathrm{I}}(\rho, \sigma) := S(\sigma, \rho) \quad \text{and} \quad S^{\mathrm{rI}}(\rho, \sigma) := S(\rho, \sigma)\,.$$

If $\{A(i)\}_{i \in \mathbb{N}}$ is a sequence of statements, then we shall say that $A(i)$ is true *for large* i if there is $N \in \mathbb{N}$ such that $A(i)$ holds for all $i \geq N$. We define a family of subsets of the state space $\mathcal{S}$ by

$$\mathcal{T}^{\omega} := \left\{ U \subset \mathcal{S} \;\middle|\; \begin{array}{l} \text{if } \rho \in U, (\rho_i)_{i \in \mathbb{N}} \subset \mathcal{S} \text{ and } \lim\limits_{i \to \infty} S^{\omega}(\rho, \rho_i) = 0\,, \\ \text{then } \rho_i \in U \text{ for large } i \end{array} \right\}.$$

The *open ω-disk* about $\rho \in \mathcal{S}$ with radius $\epsilon \in (0, \infty]$ is

$$V^{\omega}(\rho, \epsilon) := \{\sigma \in \mathcal{S} \mid S^{\omega}(\rho, \sigma) < \epsilon\} \tag{13}$$

[3]The analogue of (12) with flipped arguments is wrong: If σ is not invertible, then $S((1-\lambda)\tau + \lambda\rho, \sigma) = \infty$ for $\lambda < 1$ while the limit $S(\rho, \sigma)$ can be arbitrary.

and the *closed ω-disk* about $\rho \in \mathcal{S}$ with radius $\epsilon \in (0,\infty]$ is

$$W^{\omega}(\rho,\epsilon) := \{\sigma \in \mathcal{S} \mid S^{\omega}(\rho,\sigma) \leq \epsilon\}\,. \tag{14}$$

We denote the (unique) norm topology on $\mathcal{S}$ by $\mathcal{T}^{\|\cdot\|}$. For $a \in \mathcal{A}$ and $(a_i)_{i\in\mathbb{N}} \subset \mathcal{A}$ we denote by $\lim_{i\to\infty} a_i = a$ the convergence of $(a_i)_{i\in\mathbb{N}}$ to a in norm.

2. Let $(X,\mathcal{T})$ be a topological space. The topology $\mathcal{T}$ is a *Hausdorff topology* if each two distinct points of X belong to two disjoint open sets. A family $\mathcal{B} \subset \mathcal{T}$ is a *base* for $(X,\mathcal{T})$ if any non-empty open subset of X is a union of a subfamily of $\mathcal{B}$. A family $\mathcal{B}(x)$ of open sets containing $x \in X$ is called a *base* for $(X,\mathcal{T})$ at x if for any open set V containing x there exists $U \in \mathcal{B}(x)$ such that $U \subset V$. The topological space $(X,\mathcal{T})$ is *first-countable* if there exists a countable base at every point $x \in X$, it is *second-countable* if it has a countable base.

The family $\mathcal{T}^{\omega}$ is easily seen to be a topology on $\mathcal{S}$, which we call the *ω-topology.* The inclusion $\mathcal{T}^{\|\cdot\|} \subset \mathcal{T}^{\omega}$ follows directly from the *Pinsker-Csiszár inequality,* which confirms for $\rho,\sigma \in \mathcal{S}$ that

$$\|\rho-\sigma\|_1^2 \leq 2S(\rho,\sigma)\,. \tag{15}$$

Here the trace norm from Definition 4.1.2 is used, see e.g. §3.4 in [50] for a proof. The inclusion $\mathcal{T}^{\|\cdot\|} \subset \mathcal{T}^{\omega}$ implies that the ω-topology is a Hausdorff topology. The convergence of sequences *a priori*, in terms of the relative entropy, is equivalent to the convergence *a posteriori*, in terms of the ω-topology. This is formalized as the equivalence a) below, which in Theorem 5.18 takes the form of $C(\mathcal{T}^{\omega}) = C^{\omega}$. For sequences $(\rho_i)_{i\in\mathbb{N}} \subset \mathcal{S}$ and states $\rho \in \mathcal{S}$ we have

$$\begin{aligned}\lim_{i\to\infty} S^{\omega}(\rho,\rho_i) = 0 \quad &\overset{\text{a)}}{\iff} \quad \forall U \in \mathcal{T}^{\omega} \text{ with } \rho \in U \text{ we have } \rho_i \in U \text{ for large } i\\ &\overset{\text{b)}}{\iff} \quad \forall \epsilon \in (0,\infty] \text{ we have } \rho_i \in V^{\omega}(\rho,\epsilon) \text{ for large } i\,.\end{aligned} \tag{16}$$

Equivalence a) holds more generally for any divergence function in the sense of §5.1. In particular, a) holds for infinite-dimensional algebras.

The equivalence b) is more restrictive. It follows from a continuity property. In Proposition 5.16 we use (15), and for the case $\omega = \mathrm{rI}$ some perturbation theory, to show for all $\rho,\sigma \in \mathcal{S}$ and $(\sigma_i)_{i\in\mathbb{N}} \subset \mathcal{S}$

$$\lim_{i\to\infty} S^{\omega}(\sigma,\sigma_i) = 0 \quad \Longrightarrow \quad \lim_{i\to\infty} S^{\omega}(\rho,\sigma_i) = S^{\omega}(\rho,\sigma)\,. \tag{17}$$

In Theorem 5.18.2 we show that (17) means that the relative entropy is continuous in the first argument for the I-topology and in the second argument for the rI-topology. Therefore the open ω-disks are a base of $\mathcal{T}^{\omega}$, which is equivalent to b) in (16). Hence $\mathcal{T}^{\omega}$ is first-countable while Corollary 5.19 shows that $\mathcal{T}^{\omega}$ is second-countable if and only if $\mathcal{A}$ is commutative. The continuity (17) is wrong for the infinite-dimensional algebra l^{∞}, where the open ω-disks are not a base of a topology, see §2.1.

In addition to proving distance-like properties (16), we use (17) in Theorem 5.18.4 to show

$$\mathcal{T}^{\|\cdot\|} \subset \mathcal{T}^{\mathrm{rI}} \subset \mathcal{T}^{\mathrm{I}} . \tag{18}$$

Equivalently we have for all sequences $(\rho_i)_{i\in\mathbb{N}} \subset \mathcal{S}$ and states $\rho \in \mathcal{S}$ the ordering of convergences

$$\lim_{i\to\infty} S(\rho_i, \rho) = 0 \quad \Longrightarrow \quad \lim_{i\to\infty} S(\rho, \rho_i) = 0 \quad \Longrightarrow \quad \lim_{i\to\infty} \rho_i = \rho .$$

Later in Corollary 5.19 we prove the proper inclusion $\mathcal{T}^{\|\cdot\|} \subsetneq \mathcal{T}^{\mathrm{rI}}$ for non-commutative algebras, whereas $\mathcal{T}^{\mathrm{rI}} \subsetneq \mathcal{T}^{\mathrm{I}}$ holds already for the algebra $\mathbb{C}^2$ of a bit. Other conditions for commutativity will be mentioned in §6.6.

The infimum of the relative entropy between a state and a set of states is useful in information theory and in quantum information theory, e.g. to compute optimal error rates in hypothesis testing [14].

Definition 2.6 (ω-closure). For $\rho \in \mathcal{S}$ and $X \subset \mathcal{S}$ we write

$$S^{\omega}(\rho, X) := \inf_{\tau \in X} S^{\omega}(\rho, \tau) \tag{19}$$

and we define the ω-*closure* of $X \subset \mathcal{S}$ by

$$\mathrm{cl}^{\omega}(X) := \{\rho \in \mathcal{S} \mid S^{\omega}(\rho, X) = 0\} . \tag{20}$$

In a C*-subalgebra of $\mathrm{Mat}(n, \mathbb{C})$ we can show in Theorem 5.18.2 for an arbitrary subset $X \subset \mathcal{S}$ of states that the ω-closure $\mathrm{cl}^{\omega}(X)$ is the topological closure of X with respect to $\mathcal{T}^{\omega}$. Differently frased, we have for all $\rho \in \mathcal{S}$

$$S^{\omega}(\rho, \mathrm{cl}^{\omega}(X)) = S^{\omega}(\rho, X) . \tag{21}$$

This is also proved in Corollary 5.17 directly from (17). The analogue statement for spaces of probability measures on infinite σ-algebras is wrong by Example 5.4.

3. New results about exponential families

This is an expository section about exponential families in a C*-subalgebra $\mathcal{A}$ of $\mathrm{Mat}(n, \mathbb{C})$. With the exception of some corollaries, all proofs are done in §6. We begin in §3.1 by explaining and proving the Pythagorean theorem and the projection theorem in a restriction where this is easy. In §3.2 and §3.3 we define an extension of every exponential family, by lifting poonems of a convex parameter space. The notion of *poonem* [28] is equivalent to the notion of *face* (see e.g. §1.2.1 in [59] for a proof) which we will define in §4.1. In §3.4 we explain a new Pythagorean theorem, valid for this extension. A corollary solves the problem of maximizing the von Neumann entropy under linear constraints. This is the first complete solution in the literature. In §3.5 we explain a new projection theorem, valid for the extension. A corollary shows that the extension is the rI-closure of the exponential family.

The *Staffelberg family* [61] in §3.5 shows that the norm topology is too coarse to extend an exponential family appropriately, its closures are too large. The *Swallow family* [61] in §3.6 demonstrates why poonems [28] are essential in our proof of the projection theorem.

Issues proved in §6 but not covered in the present section include applications to quantum correlations in §6.5 and equality conditions for closures of exponential families in §6.6.

3.1. A recap of elementary information geometry

We recall the Pythagorean theorem and the projection theorem for the relative entropy in a C*-subalgebra $\mathcal{A}$ of $\mathrm{Mat}(n,\mathbb{C})$. The aim of this article is to extended them from the invertible states to the whole state space. These theorems are exemplary for many elegant ideas in information geometry [2]. The following Pythagorean theorem has first appeared in articles [49, 45] by Petz and Nagaoka. It is well-known [45, 4, 35] that it fits into the differential-geometric context of dually flat spaces, initiated by Amari [2].

The *Pythagorean theorem of relative entropy* applies to states $\rho,\sigma,\tau\in\mathcal{S}$ where σ,τ are invertible and $\rho-\sigma$ is orthogonal to $\log(\tau)-\log(\sigma)$ with respect to the Hilbert-Schmidt inner product. An elementary calculation shows for the relative entropy S

$$S(\rho,\sigma)\,+\,S(\sigma,\tau)=S(\rho,\tau)\,. \tag{22}$$

With relative entropy (11) replaced by squared Euclidean distance, this equation reminds us of the Pythagorean theorem in Euclidean geometry. See also §3.4 in [50] and §3.4 in [4] for further information, as well as §7 in [4] for an overview of applications in estimation theory.

Definition 3.1. We use the real analytic function $R_{\mathcal{A}}:\mathcal{A}_{\mathrm{sa}}\to\mathcal{A}_{\mathrm{sa}}$,

$$R(\theta)=R_{\mathcal{A}}(\theta):=\exp_{\mathcal{A}}(\theta)/\operatorname{tr}(\exp_{\mathcal{A}}(\theta))\,. \tag{23}$$

The exponential $\exp_{\mathcal{A}}$ is defined by functional calculus[4] in the algebra $\mathcal{A}$, see Definition 4.22.3 and Remark 4.23.2. For a non-empty real affine subspace $\Theta\subset\mathcal{A}_{\mathrm{sa}}$ we define an *exponential family* in $\mathcal{A}$ by

$$\mathcal{E}:=R_{\mathcal{A}}(\Theta)=\{R_{\mathcal{A}}(\theta)\mid\theta\in\Theta\,\}\,. \tag{24}$$

The parametrization $R_{\mathcal{A}}$ is the *canonical parametrization* of $\mathcal{E}$. We call a one-dimensional exponential family (with/-out parametrization) *e-geodesic*. We use the translation vector space $U:=\mathrm{lin}(\Theta)=\Theta-\Theta=\{\theta_1-\theta_2\mid\theta_1,\theta_2\in\Theta\}$.

We mention vocabulary in the literature. The analogue of the parametrization $R_{\mathcal{A}}$ of an exponential family is called *canonical parametrization* in probability theory, see §20 in [16]. For an affine map $a:\mathbb{R}\to\Theta$ the curve $\gamma:t\mapsto R_{\mathcal{A}}\circ a(t)$ is

[4]We have $\exp_{\mathcal{A}}(a)=\mathbb{1}+\sum_{i=1}^{\infty}a^i/i!$ where the identity $\mathbb{1}$ in $\mathcal{A}$ can differ from the identity $\mathbb{1}_n$ of $\mathrm{Mat}(n,\mathbb{C})$.

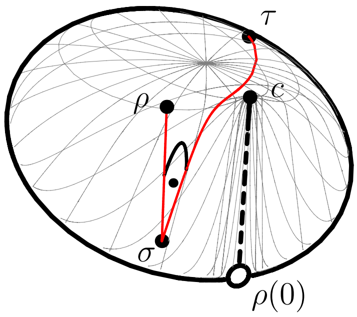

Figure 3.1: The *Staffelberg family* is sketched by e-geodesics (thin curves). The Euclidean geodesic from σ to ρ meets the e-geodesic from σ to τ (red) orthogonally with respect to the BKM-metric: The Pythagorean theorem $S(\rho,\sigma)+S(\sigma,\tau)=S(\rho,\tau)$ holds. Closure components in different topologies are indicated (bold).

called *e-geodesic* in §3.4 in [50]. This curve is called (+1)-*geodesic* in Section 7.2 in [4] while an *e-geodesic* is a more general concept there, see also Remark 4 in [61].

Example 3.2 (The Staffelberg family). We shall use Pauli σ-matrices $\sigma_1 := \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $\sigma_2 := \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right)$ and $\sigma_3 := \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. The *Staffelberg family*, studied in [61], is the exponential family

$$R(\mathrm{span}_{\mathbb{R}}(\sigma_1 \oplus 0,\, \sigma_2 \oplus 1\,)) \subset \mathrm{Mat}(2,\mathbb{C}) \oplus \mathbb{C} \cong \left(\begin{smallmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & * \end{smallmatrix}\right) \subset \mathrm{Mat}(3,\mathbb{C})$$

embedded into $\mathrm{Mat}(3,\mathbb{C})$ by block diagonal matrices. The Staffelberg family is depicted in Figure 3.1. The pointed circle about the family is an equator of the Bloch ball $\mathcal{S}(\mathrm{Mat}(2,\mathbb{C}))$, parametrized for real α by $\rho(\alpha) := \frac{1}{2}(\mathbb{1}_2 + \sin(\alpha)\sigma_1 + \cos(\alpha)\sigma_2) \oplus 0$. The figure also shows $c := \frac{1}{2}(\rho(0) + 0_2 \oplus 1)$. Figure 3.4 shows the Staffelberg family inside the state space $\mathcal{S}_{\mathcal{B}} := \{\rho \in \mathcal{B} \mid \rho \succeq 0, \mathrm{tr}(\rho) = 1\}$ of the real *-subalgebra $\mathcal{B}$ spanned by $\sigma_1 \oplus 0$, $\sigma_2 \oplus 0$, $\mathrm{i}\sigma_3 \oplus 0$, $\mathrm{diag}(1,1,0)$ and $\mathbb{1}_3$. The algebra $\mathcal{B}$ is closed under real scalar multiplication and under the adjoint map. Its state space is a 3D cone with apex $0_2 \oplus 1$ based on the circle $\{\rho(\alpha) \mid \alpha \in \mathbb{R}\}$.

The Pythagorean theorem (22) applies to exponential families. For states $\rho \in \mathcal{S}$ and $\sigma, \tau \in \mathcal{E}$, such that $\rho - \sigma$ is perpendicular to the translation vector space U, with respect to the Hilbert-Schmidt scalar product, we have

$$S(\rho,\sigma) + S(\sigma,\tau) = S(\rho,\tau)\,. \tag{25}$$

The condition $\rho - \sigma \perp U$ means that the Euclidean straight line from σ to ρ is perpendicular to the exponential family $\mathcal{E}$ with respect to the BKM-Riemannian metric, see Remark 6.2. This is indicated by the right angle in Figure 3.1.

The *projection theorem*, is now an easy corollary. For every state $\rho \in \mathcal{E} + U^\perp$ the intersection $(\rho + U^\perp) \cap \mathcal{E}$ contains a unique state $\pi_{\mathcal{E}}(\rho)$, which defines a projection to $\mathcal{E}$

$$\pi_{\mathcal{E}} :\; (\mathcal{E} + U^\perp) \cap \mathcal{S} \;\to\; \mathcal{E}\,, \quad \rho \;\mapsto\; \pi_{\mathcal{E}}(\rho)\,. \tag{26}$$

By the choice of ρ the intersection is non-empty. If it contains two states σ, τ, then $\rho - \sigma \perp U$ and $\rho - \tau \perp U$. Equality $\sigma = \tau$ follows if we add the two corresponding Pythagorean equations (25). The minimal relative entropy of ρ from $\mathcal{E}$, called *entropy distance* in [61], is

$$\mathrm{d}_{\mathcal{E}}(\rho) := \inf_{\tau \in \mathcal{E}} S(\rho, \tau)\,. \tag{27}$$

If $\rho \in \mathcal{E} + U^{\perp}$ then (25) implies the projection theorem

$$\mathrm{d}_{\mathcal{E}}(\rho) = S(\rho, \pi_{\mathcal{E}}(\rho))\,. \tag{28}$$

A geometric optimization formula like (28) is called a projection theorem in §3.4 in [4]. We will extend (28) in this article to arbitrary states ρ.

3.2. An algorithm for poonems of the mean value set

We use two equivalent descriptions of the convex set of mean values. Its boundary components are described algebraically in a lattice $\mathcal{P}^U$ of projections.

Definition 3.3. The *mean value set* of a linear subspace $U \subset \mathcal{A}_{\mathrm{sa}}$ of self-adjoint matrices is the orthogonal projection of the state space onto U

$$\mathbb{M}(U) = \mathbb{M}_{\mathcal{A}}(U) := \pi_U(\mathcal{S}_{\mathcal{A}}) \subset U\,. \tag{29}$$

Here $\pi_U : \mathcal{A}_{\mathrm{sa}} \to U$ denotes the orthogonal projection from $\mathcal{A}_{\mathrm{sa}}$ onto U. This linear mapping is characterized for each $a \in \mathcal{A}_{\mathrm{sa}}$ by the equation $a - \pi_U(a) \perp U$. For $u_1, \ldots, u_k \in \mathcal{A}_{\mathrm{sa}}$ we abbreviate $\mathbf{u} := (u_1, \ldots, u_k)$ and we define the *mean value mapping* by

$$m_{\mathbf{u}} : \ \mathcal{A}_{\mathrm{sa}} \ \to \ \mathbb{R}^k\,, \quad a \ \mapsto \ (\langle u_1, a\rangle, \ldots, \langle u_k, a\rangle)\,. \tag{30}$$

The *convex support* of $\mathbf{u}$ is

$$\mathrm{cs}(\mathbf{u}) = \mathrm{cs}_{\mathcal{A}}(\mathbf{u}) := \{m_{\mathbf{u}}(\rho) \mid \rho \in \mathcal{S}_{\mathcal{A}}\} \subset \mathbb{R}^k\,. \tag{31}$$

The concept of convex support was first used by Barndorff-Nielsen [7] in probability theory and later by Čencov [16]. It was refined by Csiszár and Matúš [18, 20] to investigate mean values of exponential families. Barndorff-Nielsen's definition for a finite measurable space is equivalent to (31) if probability measures are embedded into an algebra of diagonal matrices like in (9).

The convex support introduces coordinates on the mean value set $\mathbb{M}(U)$. If $U := \mathrm{span}_{\mathbb{R}}(u_1, \ldots, u_k)$, then the convex bodies $\mathbb{M}(U) \cong \mathrm{cs}(\mathbf{u})$ are "affinely isomorphic" (see Remark 1.1.1 in [60]): The mean value mapping restricts to the bijection

$$m_{\mathbf{u}}|_{\mathbb{M}(U)} : \ \mathbb{M}(U) \ \to \ \mathrm{cs}(\mathbf{u}) \tag{32}$$

such that $m_{\mathbf{u}} \circ \pi_U = m_{\mathbf{u}}$. In the majority of all proofs we are going to use the Hilbert-Schmidt Euclidean geometry in $\mathcal{A}_{\mathrm{sa}}$, using orthogonal projection π_U rather than coordinates.

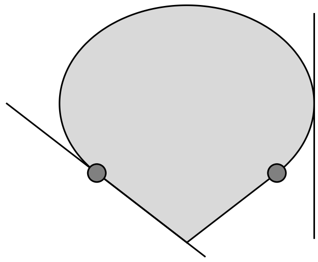

Figure 3.2: This clove shape is the mean value set of the *Swallow family*. It is the convex hull of an ellipse and a point (Example 3.4). The supporting hyperplane to the left resp. right defines an exposed face which is a segment resp. a point. Two non-exposed faces (points) of the clove are indicated by small disks.

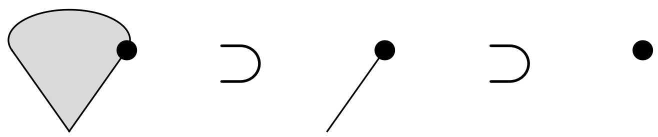

Figure 3.3: Repeated inclusions of exposed faces define a poonem. Here a poonem of point form is visualized by the disk ●.

The simplest boundary component of a convex set C in Euclidean space is an *exposed face* of C, the set of maximizers in C of a linear functional. The empty set is an exposed face by definition. Except for $\emptyset$ and C, every exposed face of C is the intersection of C and a *supporting hyperplane* H, i.e. an affine subspace H of codimension one which intersects C such that $C \setminus H$ is convex. Figure 3.2 shows examples.

A *poonem* [28] of C is a member of a sequence $F_1 \subset \cdots \subset F_k = C$, s.th. F_i is an exposed face of F_{i+1} for $i = 1, \ldots, k-1$. A *non-exposed* face of C is a poonem of C, which is not an exposed face of C. Figure 3.3 shows an example of a poonem, which is a non-exposed face. The set of exposed faces and the set of poonems are partially ordered by inclusion, they are lattices.

The first step to the extension of $\mathcal{E}$ is a lifting construction for poonems of the mean value set $\mathbb{M}(U)$. For every poonem P of $\mathbb{M}(U)$ the inverse image under projection $\{\rho \in \mathcal{S} \mid \pi_U(\rho) \in P\}$ is an exposed face of the state space $\mathcal{S}$. For every exposed face F of the state space $\mathcal{S}$ there exists a unique projection $p \in \mathcal{P}_{\mathcal{A}}$ in the projection lattice (10) of $\mathcal{A}$, such that

$$F = \mathcal{S}_{p\mathcal{A}p} = \{\rho \in p\mathcal{A}p \mid \rho \succeq 0, \mathrm{tr}(\rho) = 1\}\,.$$

The C*-subalgebra $p\mathcal{A}p = \{pap \mid a \in \mathcal{A}\}$ is called *compressed algebra*. Let us collect in the *projection lattice*

$$\mathcal{P}^U \tag{33}$$

all the projections arising in this construction from poonems of $\mathbb{M}(U)$. The projection lattice $\mathcal{P}^U$ ordered by $\preceq$ is isomorphic to the lattice of poonems of $\mathbb{M}(U)$

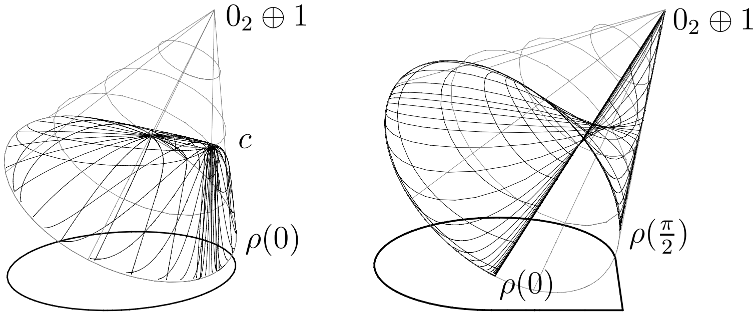

Figure 3.4: The *Staffelberg family* (left) and *Swallow family* (right) are sketched by e-geodesics inside the conic state space (real algebra). The boundary of the mean value set (projection of the state space) is drawn underneath each family.

ordered by inclusion. An algorithm to compute $\mathcal{P}^U$ is described in Remark 4.17. This is the algebraic reformulation of the concept of poonem for the special case of a mean value set.

3.3. The extension of an exponential family

We introduce an extension to the exponential family defined in (23) in terms of a non-empty affine subspace $\Theta \subset \mathcal{A}_{\mathrm{sa}}$ of self-adjoint matrices,

$$\mathcal{E} = R_{\mathcal{A}}(\Theta) = \{\exp_{\mathcal{A}}(\theta)/\mathrm{tr}\exp_{\mathcal{A}}(\theta) \mid \theta \in \Theta\}\,.$$

The translation vector space of Θ is $U := \Theta - \Theta$. To each poonem of the mean value set $\mathbb{M}(U)$ corresponds a projection $p \in \mathcal{P}^U$, we associate an exponential family to p and take the union over all $p \in \mathcal{P}^U$.

Example 3.4. The *Swallow family*, studied in [61], is the exponential family

$$R(\mathrm{span}_{\mathbb{R}}(\sigma_1 \oplus 1,\, \sigma_2 \oplus 1\,))\,.$$

Figure 3.4 shows two exponential families $\mathcal{E}$ inside the conic state space $\mathcal{S}_{\mathcal{B}}$ of the real algebra $\mathcal{B}$ from Example 3.2. The figure also shows the mean value sets $\mathbb{M}(V)$, translated into the drawing frame, with respect to the *canonical tangent space* [61]

$$V := \{\log(\rho) - \mathrm{tr}(\log(\rho))\mathbb{1}_3/3 \mid \rho \in \mathcal{E}\}\,.$$

The projection of the 3D cone $\mathcal{S}_{\mathcal{B}}$ onto V equals the mean value set $\mathbb{M}_{\mathcal{A}}(V)$ for algebras $\mathcal{A} = \mathrm{Mat}(3,\mathbb{C})$ or $\mathcal{A} = \mathrm{Mat}(2,\mathbb{C}) \oplus \mathbb{C}$. This follows from Theorem 6.5 as $\mathcal{E}$ is included in the state space $\mathcal{S}_{\mathcal{B}}$. This also follows from simpler arguments about state spaces, see Lemma 3.13 in [60]. So $\mathbb{M}(V) = \pi_V(\mathcal{S}_{\mathcal{B}})$ is the convex hull of an ellipse and a point.

We now consider an affine space $p\Theta p = \{p\theta p \mid \theta \in \Theta\}$ for each $p \in \mathcal{P}^U$ and we use it to define an exponential family in the compressed algebra $p\mathcal{A}p$ by

$$\mathcal{E}_p := R_{p\mathcal{A}p}(p\Theta p) = \{p\exp(p\theta p)/\mathrm{tr}(p\exp(p\theta p)) \mid \theta \in \Theta\} \tag{34}$$

using the normalized exponential $R_{p\mathcal{A}p}$ in (23). Noticing $\mathbb{1} \in \mathcal{P}^U$ and $\mathcal{E}_{\mathbb{1}} = \mathcal{E}$, the disjoint union

$$\operatorname{ext}(\mathcal{E}) := \bigcup_{p \in \mathcal{P}^U \setminus \{0\}} \mathcal{E}_p \tag{35}$$

contains $\mathcal{E}$ and we will prove in Lemma 6.9 a bijection between this extension and the mean value set

$$\pi_U|_{\operatorname{ext}(\mathcal{E})} : \ \operatorname{ext}(\mathcal{E}) \ \to \ \mathbb{M}(U)\,. \tag{36}$$

This allows us to define a projection

$$\pi_{\mathcal{E}} : \ \mathcal{S} \ \to \ \operatorname{ext}(\mathcal{E})\,, \tag{37}$$

such that $\pi_U(\rho) = \pi_U \circ \pi_{\mathcal{E}}(\rho)$ holds for all states ρ.

The bijection (36) is based on the lattice analysis outlined above and on the mean value chart of exponential families, proved in [63] and generalized in §6.1.

Example 3.5 (Extension). The Staffelberg family, shown in Figure 3.1, is extended by the pointed circle $\{\rho(\alpha) \mid \alpha \in (0, 2\pi)\}$ of one-point exponential families, the states $\rho(\alpha)$ being defined in Example 3.2, and by $c = \frac{1}{2}(0_2 \oplus 1 + \rho(0))$. Their union is in one-to-one correspondence with the elliptical boundary of the mean value set. Figure 3.4 shows the Staffelberg family with its mean value set. It also shows the Swallow family, where the extension contains two one-dimensional exponential families, the open segments between $0_2 \oplus 1$ and $\rho(0)$ resp. $\rho(\frac{\pi}{2})$. The extension of the Swallow family is the norm closure. The extension of the Staffelberg family is strictly smaller than the norm closure.

The projection lattice $\mathcal{P}^V$ of these exponential families $\mathcal{E}$ is computed in §3.3 in [60]. The rI-closure $\operatorname{cl}^{\mathrm{rI}}(\mathcal{E})$ is computed in §IV.B and §IV.D in [61]. By Theorem 6.16 the rI-closure $\operatorname{cl}^{\mathrm{rI}}(\mathcal{E})$ is the extension $\operatorname{ext}(\mathcal{E})$ which is computed directly from the projection lattice $\mathcal{P}^V$ in [61].

3.4. The Complete Pythagorean theorem

We will extend in Theorem 6.12 the Pythagorean theorem (25) to the full state space using the extension previously defined. A corollary is the maximization of the von Neumann entropy under linear constraints, a fundamental problem in quantum statistical mechanics, see e.g. [33, 53, 50, 4].

Theorem (Complete Pythagorean theorem). *For any $\rho \in \mathcal{S}$ and $\sigma, \tau \in \operatorname{ext}(\mathcal{E})$ such that $\rho - \sigma \perp U$ we have $S(\rho, \sigma) + S(\sigma, \tau) = S(\rho, \tau)$.*

Definition 3.6. The *von Neumann entropy* of $\rho \in \mathcal{S}$ is defined by

$$S(\rho) := -\operatorname{tr} \rho \log(\rho)\,, \tag{38}$$

using functional calculus, see Remark 4.23.3. The *free energy* of $\theta \in \mathcal{A}_{\mathrm{sa}}$ is

$$F(\theta) = F_{\mathcal{A}}(\theta) := \log \operatorname{tr} \exp_{\mathcal{A}}(\theta)\,, \tag{39}$$

where $\exp_{\mathcal{A}}$ is defined by functional calculus in $\mathcal{A}$, see (23).

According to Jaynes [34] the state which maximizes the von Neumann entropy under arbitrary constraints is the least biased choice of a state compatible with the constraints. Exponential families with linear canonical parameter space $\Theta = U$ are called *Gibbsian families* [50], they maximize the von Neumann entropy under linear constraints. We can show this for their extension, too. Let $u_1, \ldots, u_k \in \mathcal{A}_{\rm sa}$, put $\Theta := U := \mathrm{span}_{\mathbb{R}}(u_1, \ldots, u_k)$, $\mathcal{E} := R_{\mathcal{A}}(U)$ and denote $\mathbf{u} = (u_1, \ldots, u_k)$. We consider the mean value map $m_{\mathbf{u}}$ and the convex support $\mathrm{cs}(\mathbf{u}) = m_{\mathbf{u}}(\mathcal{S})$, defined respectively in (30) and (31).

Corollary 3.7. *For all $x \in \mathrm{cs}(\mathbf{u})$ there exists a unique state $\sigma \in \mathrm{ext}(\mathcal{E})$ such that $\sigma = \mathrm{argmax}\{S(\rho) \mid \rho \in \mathcal{S}, m_{\mathbf{u}}(\rho) = x\}$.*

Proof. Using the bijections (36) and (32) there exists a unique state $\sigma \in \mathrm{ext}(\mathcal{E})$ such that $m_{\mathbf{u}}(\sigma) = x$. Let $\rho \in \mathcal{S}$ such that $m_{\mathbf{u}}(\rho) = m_{\mathbf{u}}(\sigma) = x$. Then by (32) we have $\rho - \sigma \perp U$. Using $S(\rho) = \log(\mathrm{tr}\,\mathbb{1}) - S(\rho, \frac{1}{\mathrm{tr}\,\mathbb{1}}\mathbb{1})$, the Complete Pythagorean theorem, applied to the states ρ, σ and $\tau := \frac{1}{\mathrm{tr}\,\mathbb{1}}\mathbb{1} \in \mathcal{E}$, shows

$$S(\sigma) - S(\rho) = S(\rho, \tau) - S(\sigma, \tau) = S(\rho, \sigma)\,.$$

The claim follows from the distance-like properties of the relative entropy, see §2.2. □

There exist coordinates $\beta_1, \ldots, \beta_k$ for a Gibbsian family $\mathcal{E}$ analogous to *inverse temperatures* in statistical mechanics [33] – this explains the sign convention of $-\beta$ (not $+\beta$) below. The extension $\mathrm{ext}(\mathcal{E})$ has a second parameter specifying the support projection. We consider the projection lattice $\mathcal{P}^U$ and the free energy F, defined respectively in (33) and (39).

Corollary 3.8. *For every mean value tuple $x = (\xi_1, \ldots, \xi_k) \in \mathrm{cs}_{\mathcal{A}}(\mathbf{u})$ there is a unique maximizer $\rho(x)$ of the von Neumann entropy among all states $\rho \in \mathcal{S}_{\mathcal{A}}$ with mean values $m_{\mathbf{u}}(\rho) = x$. There exists a unique projection $p \in \mathcal{P}^U \setminus \{0\}$ and there exist (generally not unique) numbers $\beta_1, \ldots, \beta_k \in \mathbb{R}$, such that*

$$\frac{\partial}{\partial \beta_j} F_{p\mathcal{A}p}\left(-\sum_{i=1}^{k} \beta_i p u_i p\right) = -\xi_j\,, \quad j = 1, \ldots, k\,.$$

For each solution $p, \beta_1, \ldots, \beta_k$ we have

$$\rho(x) = R_{p\mathcal{A}p}\left(-\sum_{i=1}^{k} \beta_i p u_i p\right)$$

and $\rho(x)$ has the von Neumann entropy $S(\rho(x)) = \sum_{i=1}^{k} \beta_i \xi_i + F_{p\mathcal{A}p}(-\sum_{i=1}^{k} \beta_i p u_i p)$.

Proof. Corollary 3.7 shows that the extension $\mathrm{ext}(\mathcal{E})$ is the set of unique maximizers of the von Neumann entropy under the linear constraints. Corollary 6.11 provides the described coordinates. It remains to compute the von Neumann entropy of $\rho := \rho(x) = R_{p\mathcal{A}p}(p(-\sum_{i=1}^k \beta_i u_i)p)$. By definition of $R_{p\mathcal{A}p}$ in (23) we have

$$\begin{aligned} S(\rho) = & -\operatorname{tr}\rho\log(\rho) \\ = & -\operatorname{tr}\rho\left[-\sum_{i=1}^k \beta_i p u_i p - F_{p\mathcal{A}p}\left(-\sum_{i=1}^k \beta_i p u_i p\right)\right] \\ = & \sum_{i=1}^k \beta_i \langle p u_i p, \rho\rangle + F_{p\mathcal{A}p}\left(-\sum_{i=1}^k \beta_i p u_i p\right). \end{aligned}$$

We have $\langle p u_i p, \rho\rangle = \langle u_i, \rho\rangle = \xi_i$ for $i = 1, \dots, k$ because ρ has support $s(\rho) = p$. $\square$

In the paragraph following Corollary 6.11 we comment on the uniqueness of parameters $\beta_1, \dots, \beta_k$. The special case of Corollary 3.8 for states of full support $s(\rho) = \mathbb{1}$ in $\mathcal{A}$ was known already in 1963. To explain it, let $\operatorname{ri}(C)$ denote the *relative interior* of a convex subset C in Euclidean space, i.e. $\operatorname{ri}(C)$ is the interior of C in the norm topology of the affine hull of C. Wichmann [63] has proved the bijection

$$\pi_U|_{\mathcal{E}} : \mathcal{E} \to \operatorname{ri}\mathbb{M}(U). \tag{40}$$

It implies the Pythagorean theorem (25) and the projection theorem (28) for all states $\rho \in U^\perp + \operatorname{ri}\mathbb{M}(U)$, in particular for all invertible states ρ. This holds since $\pi_U(\operatorname{ri}\mathcal{S}) = \operatorname{ri}\mathbb{M}(U)$ and because the relative interior $\operatorname{ri}\mathcal{S}$ consists of the invertible states, see e.g. Proposition 4.5. The map

$$\Phi : \operatorname{cs}(\mathbf{u}) \to \mathcal{S}, \quad x \mapsto \operatorname{argmax}\{S(\rho) \mid \rho \in \mathcal{S}, m_{\mathbf{u}}(\rho) = x\} \tag{41}$$

parametrizes $\operatorname{ext}(\mathcal{E})$ by mean values. Wichmann has shown the following:

- $\Phi(\operatorname{ri}\operatorname{cs}(\mathbf{u})) = \mathcal{E}$,
- $\Phi|_{\operatorname{ri}\operatorname{cs}(\mathbf{u})} : \operatorname{ri}\operatorname{cs}(\mathbf{u}) \to \mathcal{E}$ is real analytic,
- $\Phi(\operatorname{cs}(\mathbf{u})) \subset \overline{\mathcal{E}}$, where $\overline{\mathcal{E}}$ denotes the closure of $\mathcal{E}$ in the norm topology.

Our result of $\Phi(\operatorname{cs}(\mathbf{u})) = \operatorname{ext}(\mathcal{E})$ in Corollary 3.8 resolves $\Phi(\operatorname{cs}(\mathbf{u})) \subset \overline{\mathcal{E}}$ to an equality. However we do not know how tight the upper bound $\Phi(\operatorname{cs}(\mathbf{u})) \subset \overline{\mathcal{E}}$ is. This question is related to the continuity of Φ. A discontinuous Φ is presented in Example 3.10, see also Proposition 6.22.

3.5. The Complete projection theorem

We extend in Theorem 6.16 the projection theorem (28) to the full state space using the extension $\operatorname{ext}(\mathcal{E})$ previously defined. As a corollary we write $\operatorname{ext}(\mathcal{E})$ as the topological closure in the rI-topology. Both statements hold for our standard assumptions of an affine space Θ and $\mathcal{E} = R_{\mathcal{A}}(\Theta)$.

For technical reasons we write $S_\rho(\sigma) := S(\rho, \sigma)$ for the relative entropy with fixed state ρ and variable state σ. We use the entropy distance

$$\mathrm{d}_{\mathcal{E}}(\rho) = \inf_{\sigma \in \mathcal{E}} S_\rho(\sigma)$$

and the projection $\pi_{\mathcal{E}}$ defined in (27) and (37) respectively.

Theorem (Complete projection theorem). *For each $\rho \in \mathcal{S}$ the relative entropy S_ρ has on the extension* $\mathrm{ext}(\mathcal{E})$ *a unique local minimizer at $\pi_{\mathcal{E}}(\rho)$. We have*

$$\mathrm{d}_{\mathcal{E}}(\rho) = \min_{\sigma \in \mathrm{ext}(\mathcal{E})} S_\rho(\sigma) = S_\rho(\pi_{\mathcal{E}}(\rho)) \,.$$

The Complete projection theorem was proved in the thesis [58]. It shows $\mathrm{ext}(\mathcal{E}) = \{\rho \in \mathcal{S} \mid \mathrm{d}_{\mathcal{E}}(\rho) = 0\}$, where the right-hand side is the rI-closure $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$ of $\mathcal{E}$ defined in (20). Theorem 5.18.2 shows that $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$ is the topological closure of $\mathcal{E}$ with respect to the rI-topology. This topology can be defined by the base $\{V^{\mathrm{rI}}(\rho,\epsilon) \mid \epsilon \in (0,\infty]\}$ of open rI-disks $V^{\mathrm{rI}}(\rho,\epsilon) = \{\sigma \in \mathcal{S} \mid S(\rho,\sigma) < \epsilon\}$. To sum up:

Corollary 3.9. *We have* $\mathrm{ext}(\mathcal{E}) = \mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$.

We have seen in (18) that the rI-topology is finer than the norm topology. So $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) \subset \overline{\mathcal{E}}$ is obvious. A proper inclusion is possible.

Example 3.10. The *Staffelberg family* $\mathcal{E}$ from Example 3.2 satisfies $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) \subsetneq \overline{\mathcal{E}}$. This exponential family is depicted in Figure 3.1: The norm closure $\overline{\mathcal{E}}$ is the union of $\mathcal{E}$ with the bold circle around $\mathcal{E}$ and the dashed upright segment from $\rho(0)$ to $0_2 \oplus 1$. The upright segment is missing in the rI-closure $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$ except for its top endpoint c. See §IV.B in [61] for this analysis.

The Staffelberg is a Gibbsian family (with $\frac{1}{\mathrm{tr}\mathbb{1}}\mathbb{1} \in \mathcal{E}$). The rI-closure $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$ is a set of maximizers of the von Neumann entropy under linear constraints. The norm closure $\overline{\mathcal{E}}$ is too large for this aim.

3.6. Why poonems are essential

Poonems are essential for the Complete projection theorem. Strictly speaking, the bijection in Lemma 6.9

$$\pi_U|_{\mathrm{ext}(\mathcal{E})} : \ \mathrm{ext}(\mathcal{E}) \ \to \ \mathbb{M}(U)$$

as well as the Complete Pythagorean theorem can be proved without any need for poonems – using the equivalent notion of face.

The first reason to use poonems is that the algebraic algorithm in Remark 4.17 to compute the projection lattice $\mathcal{P}^U$ proceeds along chains of consecutively exposed faces of $\mathbb{M}(U)$

$$\mathbb{M}(U) = F_0 \ \supsetneq \ F_1 \ \supsetneq \ \cdots \ \supsetneq \ F_m \tag{42}$$

such that F_{i+1} is an exposed face of F_i for $i = 0, \ldots, m-1$. Elements of such chains are poonems by definition. Though the projection lattice $\mathcal{P}^U$ is essential for defining the extension $\mathrm{ext}(\mathcal{E})$ in (35) we are not forced to use the algorithm, since $\mathcal{P}^U$ is defined in (46) in terms of lattice isomorphisms.

Chains (42) and the algorithm are essential in the Complete projection theorem. Certainly, if ρ is a state, $p \in \mathcal{P}^U$ and $\pi_U(\rho) \in \pi_U(\mathcal{E}_p)$ then we can compute the minimum

$$\mathrm{d}_{\mathcal{E}_p}(\rho) = \inf\{S_\rho(\sigma) \mid \sigma \in \mathcal{E}_p\}\,.$$

This will be done in Proposition 6.15 in the paragraph following (80). But it is not clear how the value $\mathrm{d}_{\mathcal{E}_p}(\rho)$ is related to $\mathrm{d}_{\mathcal{E}_q}(\rho)$ for $q \succeq p$, $q \in \mathcal{P}^U$. The equality $\mathrm{d}_{\mathcal{E}_p}(\rho) = \mathrm{d}_{\mathcal{E}_q}(\rho)$ will be proved by approximation of $\mathcal{E}_p$ from within $\mathcal{E}_q$, using e-geodesics. E-geodesics guarantee a controlled limit of the (discontinuous) relative entropy (Lemma 6.14). It is not possible to exhaust the extension $\mathrm{ext}(\mathcal{E})$ in one step of such an approximation. In fact, we show in Proposition 6.21 that the union X of an exponential family $\mathcal{E}$ with the two limit points of all in $\mathcal{E}$ included e-geodesics covers only part of the mean value set $\mathbb{M}(U)$ under the projection π_U. Relative interiors of non-exposed faces of $\mathbb{M}(U)$ are missing.

The Swallow family $\mathcal{E} = R_{\mathcal{A}}(V)$ in Example 3.4 makes this clear. Two non-exposed faces of the mean value set $\mathbb{M}(V)$ are depicted in Figure 3.2. As depicted in Figure 3.4, each of these non-exposed faces has a unique inverse projection under π_U, which is $\rho(\alpha)$ for $\alpha \in \{0, \frac{\pi}{2}\}$. So there exists no e-geodesic in $\mathcal{E}$ with limit $\rho(\alpha)$ although $\rho(\alpha)$ belongs to the extension $\mathrm{ext}(\mathcal{E})$ and *a fortiori* to the norm closure $\overline{\mathcal{E}}$. On the other hand, the open segment between $0_2 \oplus 1$ and $\rho(\alpha)$ is included in X and it is an exponential family $\mathcal{E}_p$, corresponding to the rank-two projection $p = \rho(\alpha) + 0_2 \oplus 1 \in \mathcal{P}^V$. Now $\rho(\alpha)$ can be approximated in a second step by the e-geodesic $\mathcal{E}_p$. The corresponding poonems of the mean value set $\mathbb{M}(V)$ are

$$\pi_V(\mathcal{S}) = \mathbb{M}(V) \;\supsetneq\; \pi_V(\text{segment between } 0_2 \oplus 1 \text{ and } \rho(\alpha)) \;\supsetneq\; \pi_V(\rho(\alpha))\,.$$

They are depicted schematically in Figure 3.3 and they form a chain (42).

For details about missing e-geodesic limits at non-exposed faces we refer to the proof of Proposition 6.21. For explicit calculations on the Swallow family we refer to Remark 29 a) in [61].

4. Analysis on the state space

This a preparatory section summarizing techniques from convex geometry, geometry of state spaces and mean value sets including their algebraic formulation. Detailed perturbation theoretical proofs are provided in §4.3, where the standard theory is not sufficient.

Let us introduce important notations. The use of spectral values may seem unnecessary. We have decided to work with the extension $\mathrm{ext}(\mathcal{E}) = \bigcup_{p \in \mathcal{P}^U \setminus \{0\}} \mathcal{E}_p$ of an exponential family (35) as a single object in the Euclidean space $\mathcal{A}_{\mathrm{sa}}$. Thus we need to distinguish between invertibility in several compressed algebras $p\mathcal{A}p$.

Definition 4.1 (Spectral values, norms and Euclidean vector spaces).
1. Let $a \in \mathcal{A}$. The *spectrum* of a is

$$\mathrm{spec}_{\mathcal{A}}(a) := \{\lambda \in \mathbb{C} \mid a - \lambda \mathbb{1} \text{ is not invertible in } \mathcal{A}\},$$

its elements are the *spectral values* of a in $\mathcal{A}$. The matrix a is *positive semi-definite* if $a \in \mathcal{A}_{\rm sa}$ and if a has no negative spectral values, we then write $a \succeq 0$. If $a \succeq 0$, then there exists $b \in \mathcal{A}$, $b \succeq 0$ with $a = b^2$, see e.g. §2.2 in [44]. The matrix b is unique and one defines $\sqrt{a} := b$. We have $a^*a \succeq 0$ and put $|a| := \sqrt{a^*a}$.

2. In addition to the two-norm $\|\cdot\|_2$ introduced in §2.1 we consider the *spectral norm* $\|a\|$, which is the square root of the largest eigenvalue of a^*a and the *trace norm* $\|a\|_1 := \operatorname{tr}|a|$. The topology of any norm is the *norm topology* and convergence of a sequence $(a_i)_{i\in\mathbb{N}} \subset \mathcal{A}$ to $a \in \mathcal{A}$ in any norm will be denoted by $\lim_{i\to\infty} a_i = a$. The three norms restrict to the real vector space of self-adjoint matrices $\mathcal{A}_{\rm sa}$.

3. In any Euclidean vector space $(\mathbb{E}, \langle\cdot,\cdot\rangle)$ we denote the *two-norm* by $\|x\|_2 := \sqrt{\langle x, x\rangle}$ and write $x \perp y :\iff \langle x, y\rangle = 0$ for $x, y \in \mathbb{E}$. For subsets $X, Y \subset \mathbb{E}$ we write $X \perp Y :\iff \langle x, y\rangle = 0\, \forall x \in X, y \in Y$ and $X^\perp := \{y \in \mathbb{E} \mid x \perp y\, \forall x \in X\}$ (if $z \in \mathbb{E}$ then $z \perp Y :\iff \{z\} \perp Y$ and $z^\perp := \{z\}^\perp$). If $\mathbb{A}$ is a non-empty affine subspace of $\mathbb{E}$, then we denote the translation vector space of $\mathbb{A}$ by

$$\operatorname{lin}(\mathbb{A}) := \mathbb{A} - \mathbb{A} = \{a - b \mid a, b \in \mathbb{A}\}\,.$$

We denote the orthogonal projection from $\mathbb{E}$ onto $\mathbb{A}$ by $\pi_{\mathbb{A}} : \mathbb{E} \to \mathbb{A}$. This affine mapping is characterized for each $x \in \mathbb{E}$ by the equation $x - \pi_{\mathbb{A}}(x) \perp \operatorname{lin}(\mathbb{A})$.

4.1. Lattices of faces and projections

In this section we settle the convex geometry and the algebraic description of the mean value set $\mathbb{M}(U) = \pi_U(\mathcal{S})$, referring to [59, 60]. The mean value set (29) is the orthogonal projection of the state space $\mathcal{S}$ onto a linear subspace $U \subset \mathcal{A}_{\rm sa}$. Notice that [60] erroneously uses eigenvalues and not spectral values. The corrected statements are cited below from the copy on the arXiv.

A map $f : X \to Y$ between two partially ordered sets $(X, \leq)$ and $(Y, \leq)$ is *isotone* if for all $x, y \in X$ such that $x \leq y$ we have $f(x) \leq f(y)$. A lattice is a partially ordered set $(\mathcal{L}, \leq)$ where the *infimum* $x \wedge y$ and *supremum* $x \vee y$ of each two elements $x, y \in \mathcal{L}$ exist. A *lattice isomorphism* is a bijection between two lattices that preserves the lattice structure. A lattice $\mathcal{L}$ is *complete* if for an arbitrary subset $S \subset \mathcal{L}$ the infimum $\bigwedge S$ and the supremum $\bigvee S$ exist. The *least element* $\bigwedge \mathcal{L}$ and the *greatest element* $\bigvee \mathcal{L}$ in a complete lattice $\mathcal{L}$ are *improper* elements of $\mathcal{L}$, all other elements of $\mathcal{L}$ are *proper* elements. An *atom* of a complete lattice $\mathcal{L}$ is an element $x \in \mathcal{L}$, $x \neq \bigwedge \mathcal{L}$, such that $y \leq x$ and $y \neq x$ implies $y = \bigwedge \mathcal{L}$ for all $y \in \mathcal{L}$.

The projection lattice $\mathcal{P} = \{p \in \mathcal{A} \mid p^2 = p^* = p\}$, defined in (10), with the partial ordering $\preceq$ is a complete lattice. For this and the following two statements see e.g. [1] or Remark 2.6 in [60]. For a self-adjoint (also normal) matrix $a \in \mathcal{A}_{\rm sa}$ we have

$$a = pap \quad\iff\quad pa = a \quad\iff\quad s(a) \preceq p\,. \tag{43}$$

Hence the partial ordering for projections $p, q \in \mathcal{P}$ simplifies to $p \preceq q \iff pq = p$.

We need two distinct notions of "face" of a convex set, each defining a lattice of subsets ordered by inclusion. We begin with a general convex set.

Definition 4.2. Let $(\mathbb{E}, \langle\cdot,\cdot\rangle)$ be a finite-dimensional Euclidean vector space.

1. The *closed segment* between $x, y \in \mathbb{E}$ is $[x,y] := \{(1-\lambda)x + \lambda y \mid \lambda \in [0,1]\}$, the *open segment* is $]x,y[:= \{(1-\lambda)x + \lambda y \mid \lambda \in (0,1)\}$. A subset $C \subset \mathbb{E}$ is *convex* if $x, y \in C \implies [x,y] \subset C$.

2. Let C be a convex subset of $\mathbb{E}$. A *face* of C is a convex subset F of C, such that whenever for $x, y \in C$ the open segment $]x,y[$ intersects F, then the closed segment $[x,y]$ is included in F. If $x \in C$ and $\{x\}$ is a face, then x is called an *extreme point*. The set of faces of C will be denoted by $\mathcal{F}(C)$, called the *face lattice* of C.

3. The *support function* of a convex subset $C \subset \mathbb{E}$ is defined by $\mathbb{E} \to \mathbb{R} \cup \{\pm\infty\}$, $u \mapsto h(C,u) := \sup_{x\in C}\langle u, x\rangle$. For non-zero $u \in \mathbb{E}$ the set

$$H(C,u) := \{x \in \mathbb{E} : \langle u, x\rangle = h(C,u)\}$$

is an affine hyperplane unless it is empty, which can happen if $C = \emptyset$ or if C is unbounded in u-direction. If $C \cap H(C,u) \neq \emptyset$, then we call $H(C,u)$ a *supporting hyperplane* of C. The *exposed face* of C by u is

$$F_\perp(C,u) := C \cap H(C,u)$$

and we put $F_\perp(C,0) := C$. The faces $\emptyset$ and C are exposed faces of C by definition. The set of exposed faces of C will be denoted by $\mathcal{F}_\perp(C)$, called the *exposed face lattice* of C. A face of C, which is not an exposed face is a *non-exposed face* and we then say the face F is not *exposed*.

4. Some topology is needed. Let $X \subset \mathbb{E}$ be an arbitrary subset. The *affine hull* of X, denoted by $\operatorname{aff}(X)$, is the smallest affine subspace of $\mathbb{E}$ that contains X. The interior of X with respect to the relative topology of $\operatorname{aff}(X)$ is the *relative interior* $\operatorname{ri}(X)$ of X. The complement $\operatorname{rb}(X) := X \setminus \operatorname{ri}(X)$ is the *relative boundary* of X. If $C \subset \mathbb{E}$ is a non-empty convex subset then we consider the vector space $\operatorname{lin}(C) = \{x - y \mid x, y \in \operatorname{aff}(C)\}$. We define the *dimension* $\dim(C) := \dim(\operatorname{lin}(C))$ and $\dim(\emptyset) := -1$.

Remark 4.3. 1. As observed e.g. in [61, 59], the mean value set $\mathbb{M}(U)$ can have non-exposed faces even though all faces of $\mathcal{S}$ are exposed. An example is shown in Figure 3.2.

2. Let $C \subset \mathbb{E}$ be a convex subset. Different to Rockafellar or Schneider [52, 55] we always include $\emptyset$ and C to $\mathcal{F}_\perp(C)$ so that this set is a lattice. The inclusion $\mathcal{F}_\perp(C) \subset \mathcal{F}(C)$ is easy to show and there are various ways to see that $\mathcal{F}_\perp(C)$ and $\mathcal{F}(C)$ are complete lattices ordered by inclusion where the infimum is the intersection, see e.g. §1.1 in [59] or §2.1 in [60]. The convex set C admits by Theorem 18.2 in [52] a partition into relative interiors of its faces

$$C = \dot{\bigcup}_{F\in\mathcal{F}(C)} \operatorname{ri}(F)\,. \tag{44}$$

In particular, every proper face of C is included in the relative boundary of C and its dimension is strictly smaller than the dimension of C.

We recall the algebraic description of the face lattice $\mathcal{F}(\mathcal{S}_{\mathcal{A}}) = \mathcal{F}_{\perp}(\mathcal{S}_{\mathcal{A}})$ of the state space $\mathcal{S}_{\mathcal{A}}$.

Definition 4.4. Extreme points of $\mathcal{S}$ are called *pure states.* For every orthogonal projection $p \in \mathcal{P}_{\mathcal{A}}$ we set

$$\mathbb{F}(p) = \mathbb{F}_{\mathcal{A}}(p) := \mathcal{S}_{p\mathcal{A}p}$$

and we denote the face lattice of the state space by $\mathcal{F} = \mathcal{F}_{\mathcal{A}} := \mathcal{F}(\mathcal{S}_{\mathcal{A}})$. We use notation $\mathcal{A}_0 = \{a \in \mathcal{A}_{\rm sa} \mid \operatorname{tr}(a) = 0\}$ resp. $\mathcal{A}_1 = \{a \in \mathcal{A}_{\rm sa} \mid \operatorname{tr}(a) = 1\}$ for the spaces of trace-less resp. trace one self-adjoint matrices.

Proposition 4.5 (Proposition 2.9 in [60]). *The state space $\mathcal{S}$ is a convex body of dimension $\dim(\mathcal{A}_{\rm sa}) - 1$, the affine hull is $\operatorname{aff}(\mathcal{S}) = \mathcal{A}_1$, the translation vector space is $\operatorname{lin}(\mathcal{S}) = \mathcal{A}_0$ and the relative interior consists of all invertible states. The support function at $a \in \mathcal{A}_{\rm sa}$ is the maximal spectral value $h(\mathcal{S}, a) = \lambda^+(a)$ of a. If $a \in \mathcal{A}_{\rm sa}$ is non-zero, then the exposed face $F_{\perp}(\mathcal{S}, a) = \mathbb{F}(p)$ by a is the state space of the compressed algebra $p\mathcal{A}p$, where $p = p^+(a)$ is the maximal projection of a.*

Corollary 4.6 (Corollary 2.10 in [60]). *All faces of the state space $\mathcal{S}$ are exposed. The mapping $\mathbb{F} : \mathcal{P} \to \mathcal{F}$, $p \mapsto \mathbb{F}(p)$ is an isomorphism of complete lattices.*

Remark 4.7. It follows from Corollary 4.6, Proposition 4.5 and (43) that every face of $\mathcal{S}$ can be written as $\mathbb{F}(p) = \{\rho \in \mathcal{S} \mid s(\rho) \preceq p\}$ for some $p \in \mathcal{P}$ and the relative interior is $\operatorname{ri}\mathbb{F}(p) = \{\rho \in \mathcal{S} \mid s(\rho) = p\}$.

Let us turn to the mean value set $\mathbb{M}(U) = \pi_U(\mathcal{S})$ defined in (29), where $U \subset \mathcal{A}_{\rm sa}$ is a linear subspace. A lifting construction connects to the isomorphism $\mathbb{F} : \mathcal{P} \to \mathcal{F}$. This leads to algebraic descriptions of the two face lattices $\mathcal{F}_{\perp}(\mathbb{M}(U)) \subset \mathcal{F}(\mathbb{M}(U))$.

Definition 4.8. We define for subsets $C \subset \mathcal{A}_{\rm sa}$ the (set-valued) *lift* by

$$L^U(C) = L^U_{\mathcal{A}}(C) := \mathcal{S}_{\mathcal{A}} \cap (C + U^{\perp}) .$$

We define the *lifted face lattice*

$$\mathcal{L}^U = \mathcal{L}^U_{\mathcal{A}} := \{\, L^U(F) \mid F \in \mathcal{F}(\mathbb{M}(U)) \,\}$$

and the *lifted exposed face lattice*

$$\mathcal{L}^{U,\perp} = \mathcal{L}^{U,\perp}_{\mathcal{A}} := \{\, L^U(F) \mid F \in \mathcal{F}_{\perp}(\mathbb{M}(U)) \,\} .$$

Lemma 4.9 (§5 in [59]). *The lift L restricts to the bijection $\mathcal{F}(\mathbb{M}(U)) \xrightarrow{L} \mathcal{L}^U$ and to the bijection $\mathcal{F}_{\perp}(\mathbb{M}(U)) \xrightarrow{L} \mathcal{L}^{U,\perp}$. These are isomorphisms of complete lattices with inverse π_U. For $u \in U$ we have $\pi_U\left[F_{\perp}(\mathcal{S}, u)\right] = F_{\perp}(\mathbb{M}(U), u)$ and $L^U\left[F_{\perp}(\mathbb{M}(U), u)\right] = F_{\perp}(\mathcal{S}, u)$.*

The lifting construction defines useful lattice isomorphisms, if we use appropriate lattices of projections:

Definition 4.10. The *projection lattice* resp. *exposed projection lattice* of U is

$$\mathcal{P}^U = \mathcal{P}^U_{\mathcal{A}} := \mathbb{F}^{-1}(\,\mathcal{L}^U_{\mathcal{A}}\,) \qquad \text{resp. } \mathcal{P}^{U,\perp} = \mathcal{P}^{U,\perp}_{\mathcal{A}} := \mathbb{F}^{-1}(\,\mathcal{L}^{U,\perp}_{\mathcal{A}}\,)\,. \tag{45}$$

Corollary 4.6 and Lemma 4.9 imply two lattice isomorphisms defined for suitable projections p by $p \mapsto \pi_U(\mathbb{F}(p))$:

$$\mathcal{P}^U \;\longrightarrow\; \mathcal{F}(\mathbb{M}(U)) \qquad \text{resp. } \mathcal{P}^{U,\perp} \;\longrightarrow\; \mathcal{F}_\perp(\mathbb{M}(U)) \tag{46}$$

between $\mathcal{P}^U$ and the face lattice of the mean value set resp. between $\mathcal{P}^{U,\perp}$ and the exposed face lattice. Lemma 4.9 characterizes the lifted exposed face lattice by

$$\mathcal{L}^{U,\perp} = \{F_\perp(\mathcal{S},u) \mid u \in U\} \cup \{\emptyset\}\,.$$

The algebraic description in Proposition 4.5 of faces $F_\perp(\mathcal{S},u)$ of the state space $\mathcal{S}$ translates therefore to the exposed faces of the mean value set $\mathbb{M}(U)$:

Corollary 4.11. *The exposed projection lattice is* $\mathcal{P}^{U,\perp}_{\mathcal{A}} = \{p^+_{\mathcal{A}}(u) \mid u \in U\} \cup \{0\}$.

Sequences of faces allows an algebraical description of non-exposed faces of $\mathbb{M}(U)$.

Definition 4.12 (Access sequence). 1. Let C be a convex subset C of the finite-dimensional Euclidean vector space $(\mathbb{E}, \langle\cdot,\cdot\rangle)$. We call a finite sequence $F_0,\dots,F_m \subset C$ an *access sequence* (of faces) for C if $F_0 = C$ and if F_{i+1} is a properly included exposed face of F_i for $i = 0,\dots,m-1$,

$$F_0 \supsetneq F_1 \supsetneq \cdots \supsetneq F_m\,. \tag{47}$$

2. For $p \in \mathcal{P}$ and $a \in \mathcal{A}_{\mathrm{sa}}$ the orthogonal projection $\mathcal{A}_{\mathrm{sa}} \to (p\mathcal{A}p)_{\mathrm{sa}}$ is

$$c^p(a) := \pi_{(p\mathcal{A}p)_{\mathrm{sa}}}(a) = pap\,. \tag{48}$$

3. We call a finite sequence $p_0,\dots,p_m \subset \mathcal{P}^U$ an *access sequence* (of projections) for U if $p_0 = \mathbb{1}$ and if p_{i+1} belongs to the exposed projection lattice $\mathcal{P}^{c^{p_i}(U),\perp}_{p_i\mathcal{A}p_i}$ for $i = 0,\dots,m-1$ and such that ($p_i \succ p_{i+1} :\Longleftrightarrow\; p_i \succeq p_{i+1}$ and $p_i \neq p_{i+1}$)

$$p_0 \succ p_1 \succ \cdots \succ p_m\,.$$

Access sequences of faces are also used in [20]. Grünbaum [28] defines a *poonem* as an element of an access sequence of faces. An example is depicted in Figure 3.3. In finite dimensions the notion of poonem is equivalent to the notion of face, see e.g. §1.2.1 in [59].

Theorem 4.13 (§3.2 in [60]). *The lattice isomorphism* $\mathcal{P}^U \to \mathcal{F}(\mathbb{M}(U))$ *in* (46) *extends to a bijection from the set of access sequences of projections for* U *to the set of access sequences of faces for* $\mathbb{M}(U)$ *by assigning*

$$(p_0,\dots,p_m) \;\mapsto\; (\pi_U(\mathbb{F}(p_0)),\dots,\pi_U(\mathbb{F}(p_m)))\,.$$

We will also use the following results from §3.2 in [60].

Lemma 4.14. *If $p \in \mathcal{P}$ is a projection, then $c^p(U) \xrightarrow{\pi_U} \pi_U((p\mathcal{A}p)_{\mathrm{sa}})$ is a real linear isomorphism and the following diagrams commute.*

$$\begin{tikzcd}
(p\mathcal{A}p)_{\mathrm{sa}} \arrow[rr, two heads, "\pi_U"] \arrow[dr, "\pi_{c^p(U)}"'] & & \pi_U((p\mathcal{A}p)_{\mathrm{sa}}) \\
& c^p(U) \arrow[ur, hook, "\pi_U"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbb{F}(p) \arrow[rr, two heads, "\pi_U"] \arrow[dr, "\pi_{c^p(U)}"'] & & \pi_U(\mathbb{F}(p)) \\
& \mathbb{M}_{p\mathcal{A}p}(c^p(U)) \arrow[ur, hook, "\pi_U"'] &
\end{tikzcd}$$

$$\begin{tikzcd}
\mathrm{ri}(\mathbb{F}(p)) \arrow[rr, two heads, "\pi_U"] \arrow[dr, "\pi_{c^p(U)}"'] & & \mathrm{ri}(\pi_U(\mathbb{F}(p))) \\
& \mathrm{ri}(\mathbb{M}_{p\mathcal{A}p}(c^p(U))) \arrow[ur, hook, "\pi_U"'] &
\end{tikzcd}$$

Corollary 4.15. *A projection $p \in \mathcal{P}$ belongs to the projection lattice $\mathcal{P}^U$ if and only if p belongs to an access sequence of projections for U.*

Corollary 4.16. *For each two projections $p, q \in \mathcal{P}^U$ such that $p \preceq q$ there exists an access sequence for U including p and q.*

Remark 4.17. Corollary 4.15 implies a computation method for $\mathcal{P}^U$, which is an algebraic reformulation of the concept of poonem for the special case of a mean value set. One has to compute the maximal projection (see Definition 4.22.2) of all elements of U, then the maximal projections of elements of $c^p(U)$ for each previously calculated projection p and so on (see Remark 3.10 and §3.3 in [60]).

Lemma 4.18 (§3.2 in [60]). *If $\rho \in \mathcal{S}$, then $\rho \in \mathrm{ri}(\mathbb{F}(p)) + U^\perp$ holds for a unique projection $p \in \mathcal{P}^U$. We have $p = \bigwedge\{q \in \mathcal{P}^U \mid s(\rho) \preceq q\}$.*

Although we will not need it in the following, let us point out an advantage that the coordinates of the convex support have over mean value sets. The algebraic decomposition of mean value sets in Lemma 4.14 into its faces becomes a simple inclusion $\mathrm{cs}_{p\mathcal{A}p} \subset \mathrm{cs}_{\mathcal{A}}$ of convex support sets:

Lemma 4.19. *Let $u_1, \dots, u_k \in \mathcal{A}_{\mathrm{sa}}$ and put $U := \mathrm{span}_{\mathbb{R}}(u_1, \dots, u_k)$. Then for all $p \in \mathcal{P}^U$ the following diagram commutes.*

$$\begin{tikzcd}
& \pi_U(\mathbb{F}(p)) \arrow[r, hook, two heads, "m_{u_1,\dots,u_k}"] & m_{u_1,\dots,u_k}(\mathbb{F}(p)) & \subset & \mathrm{cs}_{\mathcal{A}}(u_1, \dots, u_k) \\
\mathbb{F}(p) \arrow[ur, "\pi_U"] \arrow[dr, "\pi_{c^p(U)}"'] \arrow[urr, "m_{u_1,\dots,u_k}"] \arrow[drr, "m_{c^p(u_1),\dots,c^p(u_k)}"'] & & & & \\
& \mathbb{M}_{p\mathcal{A}p}(c^p(U)) \arrow[uu, hook, "\pi_U"] \arrow[r, hook, two heads, "m_{c^p(u_1),\dots,c^p(u_k)}"'] & \mathrm{cs}_{p\mathcal{A}p}(c^p(u_1), \dots, c^p(u_k)) \arrow[uu, hook, "\mathrm{Id}|_{\mathbb{R}^k}"'] & &
\end{tikzcd}$$

Proof. We extend the second diagram in Lemma 4.14. We can use the bijection $m_{u_1,\dots,u_k}|_{\mathbb{M}(U)} : \mathbb{M}(U) \to \mathrm{cs}_{\mathcal{A}}(u_1, \dots, u_k)$ in (32), it satisfies $m_{u_1,\dots,u_k} \circ \pi_U =$

$m_{u_1,\dots,u_k}$. In the algebra $p\mathcal{A}p$ this means $m_{c^p(u_1),\dots,c^p(u_k)} \circ \pi_{c^p(U)} = m_{c^p(u_1),\dots,c^p(u_k)}$ and

$$m_{c^p(u_1),\dots,c^p(u_k)}|_{\mathbb{M}(c^p(U))} : \mathbb{M}(c^p(U)) \to \mathrm{cs}_{p\mathcal{A}p}(c^p(u_1),\dots,c^p(u_k))$$

is a bijection. For all $a \in p\mathcal{A}p$ and $i = 1,\dots,k$ we have

$$\langle u_i, a\rangle = \langle u_i, pap\rangle = \langle c^p(u_i), a\rangle$$

hence $m_{u_1,\dots,u_k}(a) = m_{c^p(u_1),\dots,c^p(u_k)}(a)$ holds and completes the proof. □

4.2. Projections and functional calculus

We recall functional calculus for normal matrices. Definitions are somewhat technical because we want to work in subalgebras of $\mathrm{Mat}(n,\mathbb{C})$ not containing $\mathbb{1}_n$. The partial ordering on $\mathcal{A}_{\mathrm{sa}}$ induced by the positive semi-definite cone is central in the following. We will consider this ordering in its restriction to the lattice of projections.

Definition 4.20 (The projection lattice and spectral decomposition). 1. Let $a \in \mathcal{A}$ be a normal matrix, i.e. $a^*a = aa^*$. Let $N \in \mathbb{N}$, $\{c_i\}_{i=1}^N \subset \mathbb{C}$ be mutually distinct numbers and let $\{p_i\}_{i=1}^N \subset \mathcal{P}_{\mathcal{A}}$ be a family of non-zero projections such that for $i,j = 1,\dots,N$ we have $p_ip_j = p_i\delta_{ij}$, where $\delta_{ij} = 0$ unless $i = j$ with $\delta_{ii} = 1$. If $\sum_{i=1}^N p_i = \mathbb{1}$ and

$$a = \sum_{i=1}^{N} c_i p_i\,, \tag{49}$$

then the sum (49) is called *spectral decomposition* of a in $\mathcal{A}$, $\{p_i\}_{i=1}^N$ is a *spectral family* for a in $\mathcal{A}$ and its members are *spectral projections* of a in $\mathcal{A}$.

Remark 4.21 (Spectral decomposition). 1. It is a classical result of linear algebra, see e.g. §§79–80 in [29], that a normal matrix $a \in \mathrm{Mat}(n,\mathbb{C})$ has a unique spectral decomposition $a = \sum_{\lambda\in\mathrm{spec}_{\mathrm{Mat}(n,\mathbb{C})}(a)} \lambda p_\lambda$ in $\mathrm{Mat}(n,\mathbb{C})$. Moreover, for every $\lambda \in \mathrm{spec}_{\mathrm{Mat}(n,\mathbb{C})}(a)$ there exists a polynomial f_λ in one variable and with complex coefficients, such that $p_\lambda = f_\lambda(a)$.

2. Let $\mathcal{A} \subset \mathrm{Mat}(n,\mathbb{C})$ be a C*-subalgebra of $\mathrm{Mat}(n,\mathbb{C})$ with identity $\mathbb{1}$ and $a \in \mathcal{A}$ a normal matrix. If $a = \sum_{\lambda\in\mathrm{spec}(a)} \lambda p_\lambda$ is the spectral decomposition of a in $\mathrm{Mat}(n,\mathbb{C})$ then it is easy to show that $a = \sum_{\lambda\in\mathrm{spec}_{\mathcal{A}}(a)} \lambda(\mathbb{1}p_\lambda)$ is the unique spectral composition of a in $\mathcal{A}$. Either $\mathrm{spec}_{\mathcal{A}}(a) = \mathrm{spec}_{\mathrm{Mat}(n,\mathbb{C})}(a)$ or $\mathrm{spec}_{\mathcal{A}}(a) \subsetneq \mathrm{spec}_{\mathcal{A}}(a) \cup \{0\} = \mathrm{spec}_{\mathrm{Mat}(n,\mathbb{C})}(a)$. For all non-zero $\lambda \in \mathrm{spec}_{\mathcal{A}}(a)$ we have $\mathbb{1}p_\lambda = p_\lambda$. But $\mathbb{1}p_0 \neq p_0$ is possible.

Definition 4.22 (Special projections and functional calculus). 1. If $a \in \mathcal{A}$ is a normal matrix then we denote the spectral projections of a by $p^\lambda(a) = p^\lambda_{\mathcal{A}}(a)$ for $\lambda \in \mathrm{spec}_{\mathcal{A}}(a)$. The *support projection* of a, also called *support* of a, is $s(a) := \sum_{\lambda\in\mathrm{spec}_{\mathcal{A}}(a)\setminus\{0\}} p^\lambda(a)$. The *kernel projection* of a in $\mathcal{A}$ is $k_{\mathcal{A}}(a) := \mathbb{1} - s(a)$.

2. If a is self-adjoint, then the maximum of $\mathrm{spec}_{\mathcal{A}}(a)$ is denoted by $\lambda^+(a) = \lambda^+_{\mathcal{A}}(a)$ and the corresponding spectral projection in $\mathcal{A}$ is denoted by $p^+(a) = p^+_{\mathcal{A}}(a)$ and $p^+_{\mathcal{A}}(a)$ is called the *maximal projection* of a in $\mathcal{A}$.

3. If a complex valued function f is defined on the spectrum of a normal matrix $a \in \mathcal{A}$, then $f(a) = f_{\mathcal{A}}(a) := \sum_{\lambda \in \mathrm{spec}_{\mathcal{A}}(a)} f(\lambda) p^{\lambda}_{\mathcal{A}}(a)$ is defined by *functional calculus* in $\mathcal{A}$. If $p \in \mathcal{P}$ and f is defined on the spectrum of a normal matrix $a \in p\mathcal{A}p$, we abbreviate functional calculus in $p\mathcal{A}p$ by $f^{[p]}(a) = f^{[p]}_{\mathcal{A}}(a) := f_{p\mathcal{A}p}(a)$.

Remark 4.23 (Projections and functional calculus). 1. By Remark 4.21.2, the support projection $s(a)$ of a normal matrix $a \in \mathcal{A}$ does not depend on the algebra $\mathcal{A} \subset \mathrm{Mat}(n, \mathbb{C})$. But $k_{\mathcal{A}}(a)$ depends on $\mathcal{A}$. Similarly, if a is self-adjoint then $p^+_{\mathcal{A}}(a)$ depends on $\mathcal{A}$.

2. If $a \in \mathcal{A}$ is a normal matrix and $f : \mathbb{C} \to \mathbb{C}$ is defined on $\mathrm{spec}_{\mathcal{A}}(a)$ and on $\mathrm{spec}_{\mathrm{Mat}(n,\mathbb{C})}(a)$, then we have $f_{\mathcal{A}}(a) = \mathbb{1} f_{\mathrm{Mat}(n,\mathbb{C})}(a)$. E.g. for $a, u \in \mathcal{A}$

$$\frac{\partial}{\partial t}|_{t=0} \exp_{\mathcal{A}}(a + tu) = \int_0^1 \exp_{\mathcal{A}}((1-y)a) u \exp_{\mathcal{A}}(ya) \mathrm{d}y \tag{50}$$

holds. This follows from the analogue equation in $\mathrm{Mat}(n, \mathbb{C})$, see e.g. [39], by multiplication with $\mathbb{1} \in \mathcal{A}$. This method has to be applied carefully, e.g. $\log(1, 0)$ is undefined in $\mathcal{A} = \mathbb{C}^2$ and the term $\log^{[(1,0)]}(1, 0) = (0, 0)$ is an example of functional calculus in the compressed algebra $\mathbb{C} \oplus \{0\}$.

3. The von Neumann entropy (38) of $\rho \in \mathcal{S}$ can be defined by $S(\rho) := \mathrm{tr}\, \eta(\rho)$ in terms of functional calculus in the algebra $\mathrm{Mat}(n, \mathbb{C})$. Since the function $\eta : [0, 1] \to \mathbb{R}$, $\eta(x) = -x \log(x)$, is continuous on $[0, 1]$, it follows that the von Neumann entropy is a continuous function, see e.g. Theorem VIII.20 in [51]. The detailed definition of the relative entropy (11) is for $\rho, \sigma \in \mathcal{S}$

$$S(\rho, \sigma) = \mathrm{tr}\, \rho(\log^{[s(\rho)]}(\rho) - \log^{[s(\sigma)]}(\sigma))$$

if $s(\rho) \preceq s(\sigma)$ and otherwise $S(\rho, \sigma) := \infty$. By part 1 this definition restricts from $\mathrm{Mat}(n, \mathbb{C})$ to any C*-subalgebra of $\mathrm{Mat}(n, \mathbb{C})$. The relative entropy is not continuous.

4.3. Two perturbative statements

The following perturbation analysis is essential since the relative entropy can not be defined directly in terms of functional calculus with respect to a continuous function, like e.g. the von Neumann entropy, see Remark 4.23.3. The analysis will allow us to consider logarithmic functions depending on density matrices with some eigenvalues converging to zero. In spite of considering a C*-subalgebra $\mathcal{A} \subset \mathrm{Mat}(n, \mathbb{C})$ we can argue mainly within the algebra $\mathrm{Mat}(n, \mathbb{C})$.

Definition 4.24. 1. In this section we denote the set of eigenvalues of $a \in \mathrm{Mat}(n, \mathbb{C})$ by $\mathrm{spec}(a) = \mathrm{spec}_{\mathrm{Mat}(n,\mathbb{C})}(a)$ and we shall write ζ in place of $\zeta \mathbb{1}_n$ for scalars $\zeta \in \mathbb{C}$.

2. The *resolvent set* of a matrix $a \in \mathrm{Mat}(n, \mathbb{C})$ is the complement of the spectrum $\mathrm{res}(a) := \mathbb{C} \setminus \mathrm{spec}(a)$.

3. The *resolvent* of $a \in \mathrm{Mat}(n, \mathbb{C})$ is defined for $\zeta \in \mathrm{res}(a)$ by $(a - \zeta)^{-1}$.

4. The *second resolvent equation* for $a, b \in \mathrm{Mat}(n, \mathbb{C})$ and $\zeta \in \mathrm{res}(a) \cap \mathrm{res}(b)$ is

$$(a - \zeta)^{-1} - (b - \zeta)^{-1} = (a - \zeta)^{-1}(b - a)(b - \zeta)^{-1} . \tag{51}$$

Remark 4.25. 1. If $a, b \in \mathrm{Mat}(n, \mathbb{C})$ are self-adjoint matrices, let $\lambda_1^{\downarrow}(a), \dots, \lambda_n^{\downarrow}(a)$ denote the eigenvalues of a arranged in decreasing order and counting multiplicities. *Weyl's perturbation theorem*, proved e.g. in §III.2 in [12], states that

$$\max_{k=1}^{n} |\lambda_k^{\downarrow}(a) - \lambda_k^{\downarrow}(b)| \leq \|a - b\| . \tag{52}$$

Here the spectral norm from Definition 4.1.2 is used.

2. According to Problem 5.7 on page 40 in [36], if ζ belongs to $\mathrm{res}(a)$ for a normal matrix $a \in \mathrm{Mat}(n, \mathbb{C})$, then the resolvent of a is bounded by

$$\|(a - \zeta)^{-1}\| \leq \mathrm{dist}(\zeta, \mathrm{spec}(a))^{-1} \tag{53}$$

where $\mathrm{dist}(z, M) := \inf\{|z - m| \mid m \in M\}$ for $z \in \mathbb{C}$ and $M \subset \mathbb{C}$.

3. Given a normal matrix $a \in \mathrm{Mat}(n, \mathbb{C})$, let $\Gamma \subset \mathrm{res}(a)$ be a positively oriented circular curve of radius $r > 0$. It is well-known, see e.g. Chapter 2 §1.4 in [36], that

$$P_\Gamma(a) := -\frac{1}{2\pi i} \int_\Gamma (a - \zeta)^{-1} \mathrm{d}\zeta \tag{54}$$

is the sum of all spectral projections $p^\lambda(a)$ of a in $\mathrm{Mat}(n, \mathbb{C})$, such that λ lies inside Γ.

4. Let $a, b \in \mathrm{Mat}(n, \mathbb{C})$ be self-adjoint matrices and let Γ_λ be disjoint circular curves of radius $r > 0$ centered at $\lambda \in \mathrm{spec}(b)$. If $\|b - a\| < r$, then by Weyl's perturbation theorem (52) every eigenvalue of a lies in exactly one of the circles $\{\Gamma_\lambda\}_{\lambda \in \mathrm{spec}(b)}$. The projections $Q^\lambda(a) := P_{\Gamma_\lambda}(a)$ in (54) are defined and $\mathbb{1}_n = \sum_\lambda Q^\lambda(a)$ holds (with summation over the eigenvalues $\lambda \in \mathrm{spec}(b)$ of b). The second resolvent equation (51) and the inequality (53) imply for $\lambda \in \mathrm{spec}(b)$

$$\|Q^\lambda(a) - p^\lambda(b)\| \leq \frac{1}{2\pi} \int_{\Gamma_\lambda} \|(b - \zeta)^{-1}(b - a)(a - \zeta)^{-1}\| \mathrm{d}\zeta \leq \frac{\|b - a\|}{r(r - \|b - a\|)} . \tag{55}$$

Hence for fixed b, if $\|b - a\| \to 0$ then $Q^\lambda(a)$ converges in spectral norm to $p^\lambda(b)$.

The next proposition will characterize the rI-convergence in Proposition 5.16. By Remark 4.23.1 the support projection $s(a)$ of a self-adjoint matrix $a \in \mathcal{A}$ does not depend on $\mathcal{A}$, so we assume $\mathcal{A} = \mathrm{Mat}(n, \mathbb{C})$ in the proof.

Lemma 4.26. *Let $\rho, \sigma \in \mathcal{S}$ and $(\tau_i)_{i \in \mathbb{N}} \subset \mathcal{S}$ such that $s(\rho) \preceq s(\sigma) \preceq s(\tau_i)$ holds for all $i \in \mathbb{N}$. Then $\lim_{i \to \infty} S(\sigma, \tau_i) = 0$ implies $\lim_{i \to \infty} S(\rho, \tau_i) = S(\rho, \sigma)$.*

Proof. By the Pinsker-Csiszár inequality (15) the sequence $(\tau_i)_{i\in\mathbb{N}}$ converges to σ in norm. We view τ_i as a perturbation of σ and take a sufficiently small circle Γ of radius $r>0$ about $0\in\mathbb{C}$. Then, for large $i\in\mathbb{N}$ the projection $P_\Gamma(\tau_i)$ in (54) is defined and satisfies $k(\tau_i)\preceq P_\Gamma(\tau_i)$ where $k(\tau_i)=k_{\mathrm{Mat}(n,\mathbb{C})}(\tau_i)$ is the kernel projection. Then two projections $p_i,q_i\in\mathcal{A}$ are defined by $p_i:=P_\Gamma(\tau_i)-k(\tau_i)$ and $q_i:=\mathbb{1}_n-P_\Gamma(\tau_i)$, they satisfy $q_i+p_i=s(\tau_i)$. We think of p_i as the negligible contribution to $s(\tau_i)$.

According to Definition 4.22.3 we split the functional calculus into two compressed algebras $p_i\mathcal{A}p_i$ and $q_i\mathcal{A}q_i$,

$$S(\sigma,\tau_i)=-S(\sigma)-\mathrm{tr}\,\sigma\log^{[p_i]}(p_i\tau_i)-\mathrm{tr}\,\sigma\log^{[q_i]}(q_i\tau_i)\,.$$

We have $\tau_j\stackrel{j\to\infty}{\longrightarrow}\sigma$, by (55) we have $q_j\stackrel{j\to\infty}{\longrightarrow}s(\sigma)$ and the spectral values of $q_j\tau_jq_j$ in $q_j\mathcal{A}q_j$ are strictly larger than $r>0$, hence the term $\log^{[q_i]}(q_i\tau_i)\stackrel{i\to\infty}{\longrightarrow}\log^{[s(\sigma)]}(\sigma)$ converges. Using the assumption $S(\sigma,\tau_i)\stackrel{i\to\infty}{\longrightarrow}0$ gives $\lim_{i\to\infty}\mathrm{tr}\,\sigma\log^{[p_i]}(p_i\tau_i)=0$.

Now we use a monotonicity argument. It is clear that $\rho/\lambda^+(\rho)\preceq s(\rho)$ holds and by assumption we have $s(\rho)\preceq s(\sigma)$. If $\lambda>0$ is the smallest non-zero eigenvalue of σ then $\lambda s(\sigma)\preceq\sigma$. Hence $0\preceq\frac{\lambda}{\lambda^+(\rho)}\rho\preceq\sigma$. For all $i\in\mathbb{N}$ we have $\log^{[p_i]}(p_i\tau_i)\preceq 0$ hence

$$0=\lim_{i\to\infty}\mathrm{tr}\,\sigma\log^{[p_i]}(p_i\tau_i)\leq\frac{\lambda}{\lambda^+(\rho)}\lim_{i\to\infty}\mathrm{tr}\,\rho\log^{[p_i]}(p_i\tau_i)\leq 0$$

proves $\lim_{i\to\infty}\mathrm{tr}\,\rho\log^{[p_i]}(p_i\tau_i)=0$. Now

$$\begin{aligned}S(\rho,\tau_i)=&\;-S(\rho)-\mathrm{tr}\,\rho\log^{[p_i]}(p_i\tau_i)-\mathrm{tr}\,\rho\log^{[q_i]}(q_i\tau_i)\\ \stackrel{i\to\infty}{\longrightarrow}&\;-S(\rho)-0-\mathrm{tr}\,\rho\log^{[s(\sigma)]}(\sigma)=S(\rho,\sigma)\end{aligned}$$

completes the proof. □

The following statement is used in Proposition 6.3 to set up the mean value chart of an exponential family and in Lemma 6.13 to study rI-closures of exponential families. Part 1 is used implicitly in Lemma 7 in [63].

Lemma 4.27.

1. *Let* $(x_j)_{j\in\mathbb{N}}\subset\mathcal{A}_{\mathrm{sa}}\setminus\{0\}$ *such that* $\lim_{j\to\infty}\|x_j\|=\infty$. *We assume there exist* $u,a\in\mathcal{A}_{\mathrm{sa}}$ *such that* $\lim_{j\to\infty}\frac{x_j}{\|x_j\|}=u$ *and* $\lim_{j\to\infty}\exp_{\mathcal{A}}(x_j)=a$. *Then* $\mathrm{spec}_{\mathcal{A}}(u)\subset(-\infty,0]$ *and* $s(a)\preceq k_{\mathcal{A}}(u)$.
2. *Let* $\theta,u\in\mathcal{A}_{\mathrm{sa}}$ *such that* $\mathrm{spec}_{\mathcal{A}}(u)\subset(-\infty,0]$. *Then*

$$\lim_{t\to+\infty}\exp_{\mathcal{A}}(\theta+tu)=\exp_{\mathcal{A}}^{[k_{\mathcal{A}}(u)]}(k_{\mathcal{A}}(u)\theta k_{\mathcal{A}}(u))\,.$$

Proof. Using a C*-algebra embedding we can assume $\mathbb{1}=\mathbb{1}_n$ in the proof. Then eigenvalues can be used in $\mathcal{A}$. The strategy in the first part is to consider $y_j:=\frac{x_j}{\|x_j\|}$ as perturbations of u and to estimate spectral values of e^{x_j} in suitable compressed subalgebras. We choose disjoint circular curves Γ_λ in the complex plane about the eigenvalues $\lambda \in \mathrm{spec}(u)$. Using Weyl's perturbation theorem (52), the projections in (54)

$$Q^\lambda(x_j) := P_{\Gamma_\lambda}(x_j) = P_{\Gamma_\lambda}(y_j)$$

are defined for large j. Let $\lambda \in \mathrm{spec}(u)$ and $\lambda \neq 0$. The projection $Q^\lambda(x_j)$ is a sum of spectral projections of x_j in $\mathrm{Mat}(n,\mathbb{C})$ for non-zero eigenvalues of x_j, so $Q^\lambda(x_j) \in \mathcal{A}$ by Remark 4.21.2. We consider functional calculus in the compressed algebra $Q^\lambda(x_j)\mathcal{A}Q^\lambda(x_j)$,

$$h^\lambda(x_j) := \exp_{\mathcal{A}}^{[Q^\lambda(x_j)]}(Q^\lambda(x_j)x_j) = Q^\lambda(x_j)\exp(x_j)\,.$$

The spectral values of the self-adjoint matrix $Q^\lambda(x_j)y_j$ in $Q^\lambda(x_j)\mathcal{A}Q^\lambda(x_j)$ converge for $j \to \infty$ to $\lambda \neq 0$ because there is only one eigenvalue of u in the circle Γ_λ. Since $x_j = y_j\|x_j\|$ we have for $\lambda < 0$ and for large j the bound $\|h^\lambda(x_j)\| \leq e^{\frac{\lambda}{2}\|x_j\|}$. Then

$$\|x_j\| \overset{j\to\infty}{\to} \infty \;\text{ implies } h^\lambda(x_j) \overset{j\to\infty}{\to} 0\,. \tag{56}$$

If $\lambda > 0$ then the analogous arguments show that the spectral norm $\|h^\lambda(x_j)\| \geq e^{\frac{\lambda}{2}\|x_j\|}$ diverges to $+\infty$.

For $\lambda \neq 0$ the projection $Q^\lambda(x_j)$ converges to $p^\lambda(u)$ by (55). Hence with summation over $\lambda \in \mathrm{spec}(u) \setminus \{0\}$ we have $s(u) = \lim_{j\to\infty}\sum_{\lambda\neq 0} Q^\lambda(x_j)$. Now the assumed convergence of $\exp_{\mathcal{A}}(x_j) \overset{j\to\infty}{\to} a$ gives

$$s(u)a = \lim_{j\to\infty}\sum_{\lambda\neq 0} Q^\lambda(x_j)\exp(x_j) = \lim_{j\to\infty}\sum_{\lambda\neq 0} h^\lambda(x_j) = 0$$

and $\mathrm{spec}(u) \subset (-\infty, 0]$. Then (43) and the equation

$$k_{\mathcal{A}}(u)a = (\mathbb{1} - s(u))a = a$$

show $s(a) \preceq k_{\mathcal{A}}(u)$.

We prove convergence and calculate the limit in the second statement. For small real parameter $c > 0$ let $x_c := u + c\theta$, then $x_c \overset{c\to 0}{\to} u$. For $\lambda \in \mathrm{spec}(u) \cup \{0\}$ we choose disjoint circular curves Γ_λ in the complex plane about each such λ and we define

$$Q^\lambda(x_c) := P_{\Gamma_\lambda}(x_c)\,.$$

For all $\lambda < 0$ the argument in (56) shows $Q^\lambda(x_c)\exp(x_c) \overset{c\to 0}{\to} 0$. Since $\mathbb{1}_n = \sum_{\lambda\in\mathrm{spec}(u)\cup\{0\}} Q^\lambda(x_c)$ holds for small c, we have

$$\lim_{t\to+\infty}\exp(\theta + tu) = \lim_{c\to 0}\exp\left(\frac{1}{c}x_c\right) = \lim_{c\to 0} Q^0(x_c)\exp\left(Q^0(x_c)\frac{1}{c}x_c\right).$$

By (55) we have $Q^0(x_c) \stackrel{c\to 0}{\to} k(u) \in \mathrm{Mat}(n,\mathbb{C})$. The first order expansion is calculated in Chapter II §1 equation (1.17) in [36]: With $\widetilde{Q} := \frac{1}{2\pi i}\int_{\Gamma_0}(u-\zeta)^{-1}\theta(u-\zeta)^{-1}\mathrm{d}\zeta$ we have[5]

$$Q^0(x_c) = k(u) + c\widetilde{Q} + o(c)\,.$$

We compute

$$Q^0(x_c)\frac{1}{c}x_c = \frac{1}{c}Q^0(x_c)x_cQ^0(x_c) = k(u)\theta k(u) + o(1)$$

and the continuity of the exponential gives

$$\lim_{t\to+\infty}\exp(\theta+tu) = \lim_{c\to 0}Q^0(x_c)\exp\left(Q^0(x_c)\frac{1}{c}x_c\right) = k(u)\exp(k(u)\theta k(u))\,.$$

Multiplication of this formula with the identity $\mathbb{1}_n$ of $\mathcal{A}$ completes the proof. $\square$

5. Information topologies

We study in §5.2 the I-/rI-topology on the state space $\mathcal{S}$ of a C*-subalgebra $\mathcal{A}$ of $\mathrm{Mat}(n,\mathbb{C})$. The analysis is based on a socalled divergence function and its L*-convergence, that we customize in §5.1.

5.1. The sequential topology of a divergence function

We consider a divergence function defined for pairs of elements in some set. A topology is defined by a natural convergence of countable sequences in terms a divergence function. Finally, we explore divergence functions having two properties, which are available for the relative entropy on $\mathrm{Mat}(n,\mathbb{C})$.

Definition 5.1 (L*-convergence[6]). Let X be any set. A relation $C \subset X^{\mathbb{N}}\times X$ between sequences and members of X is a *convergence* on X. If $((x_n)_{n\in\mathbb{N}},x)\in C$ then we write $x_n \xrightarrow{C} x$ and we say $(x_n)_{n\in\mathbb{N}}$ *C-converges* to x and x is the *C-limit* of $(x_n)_{n\in\mathbb{N}}$. A convergence C on X is an *L*-convergence* and (X,C) is an *L*-space* if

a) $x_n = x$ for all n implies $x_n \xrightarrow{C} x$,

b) $x_n \xrightarrow{C} x$ and $(y_n)_{n\in\mathbb{N}}$ is a subsequence of $(x_n)_{n\in\mathbb{N}}$ then $y_n \xrightarrow{C} x$.

c) $x_n \not\xrightarrow{C} x$ (i.e. it is false that $x_n \xrightarrow{C} x$) implies the existence of a subsequence $(y_n)_{n\in\mathbb{N}}$ of $(x_n)_{n\in\mathbb{N}}$, such that for any subsequence $(z_n)_{n\in\mathbb{N}}$ of $(y_n)_{n\in\mathbb{N}}$ we have $z_n \not\xrightarrow{C} x$.

A convergence C on X is said to have *unique limits* if

d) $x_n \xrightarrow{C} x$ and $x_n \xrightarrow{C} y$ implies $x = y$.

[5]If g is a positive real valued function and f is any function (here with values in $\mathrm{Mat}(n,\mathbb{C})$), then $f = o(g)$ means $\frac{f}{g}\to 0$ and $o(g)$ is called *Landau symbol.*

We consider the family $\mathcal{T}(C)$ of subsets $U \subset X$ such that $x \in U$ and $x_n \xrightarrow{C} x$ implies $x_n \in U$ for large n.

Remark 5.2 (Sequential topologies and closures). It is well-known [22] that $\mathcal{T}(C)$ is a topology on X if C is a convergence on X. Moreover, if $Y \subset X$ is $\mathcal{T}(C)$ closed then $(y_n)_{n\in\mathbb{N}} \subset Y$ and $y_n \xrightarrow{C} y$ imply $y \in Y$. Important for our purpose is: If b) above holds, then the converse is also true, $Y \subset X$ is $\mathcal{T}(C)$ closed if and only if $(y_n)_{n\in\mathbb{N}} \subset Y$ and $y_n \xrightarrow{C} y$ imply $y \in Y$. If (X, C) is an L*-space then $\mathcal{T}(C)$ is called the *sequential topology* induced by C.

We consider closures in a sequential space.

Definition 5.3 (Sequential closures). Let C be a convergence on X. The *sequential closure* of $Y \subset X$ is

$$\mathrm{cl}_C(Y) := \{x \in X \mid (x_n)_{n\in\mathbb{N}} \xrightarrow{C} x \text{ for a sequence } (x_n)_{n\in\mathbb{N}} \subset Y\}\,. \tag{57}$$

The following property [22], will be proved in the context of the relative entropy:

e) if $x_n \xrightarrow{C} x$ and $(x^{(m)})_n \xrightarrow{C} x_m$ for all $m \in \mathbb{N}$, then there exists a function $n : \mathbb{N} \to \mathbb{N}$, such that $(x^{(m)})_{n(m)} \xrightarrow{C} x$.

A weaker property is defined in Problem 1.7.18 in [25]:

e') if $x_n \xrightarrow{C} x$ and $(x^{(m)})_n \xrightarrow{C} x_m$ for all $m \in \mathbb{N}$, then there exists sequences of natural numbers $m_1, m_2, \dots$ and $n_1, n_2, \dots$, such that $(x^{(m_k)})_{n_k} \xrightarrow{C} x$.

Sequential closures in L*-spaces need not be topological closures.

Example 5.4 (Information closures). The I-/rI-convergence of probability measures in (4) is an L*-convergence [30, 23]. Harremoës discusses a triangle D in $\mathbb{P}(\mathbb{N})$, the probability simplex (6), where $\mathrm{cl}_C(D) \subsetneq \mathrm{cl}_C(\mathrm{cl}_C(D))$ holds for the I-convergence C. Csiszár and Matúš [19] discuss an exponential family $\mathcal{E}$ of Borel probability measures in $\mathbb{R}^3$ where $\mathrm{cl}_C(\mathcal{E}) \subsetneq \mathrm{cl}_C(\mathrm{cl}_C(\mathcal{E}))$ holds for the rI-convergence C.

Sequential closures and topological closures in L*-spaces are related as follows.

Remark 5.5 (Idempotent sequential closure). Let C be a convergence on X satisfying b). Then for each $Y \subset X$ the $\mathcal{T}(C)$ closure of Y equals the sequential closure $\mathrm{cl}_C(Y)$ if and only if e') holds for C. Indeed, by Remark 5.2, since b) holds, a subset $Y \subset X$ is $\mathcal{T}(C)$ closed if and only if $\mathrm{cl}_C(Y) = Y$. Hence $\mathrm{cl}_C(Y)$ is the $\mathcal{T}(C)$ closure of Y if and only if $\mathrm{cl}_C(\mathrm{cl}_C(Y)) = \mathrm{cl}_C(Y)$. The equation $\mathrm{cl}_C(\mathrm{cl}_C(Y)) = \mathrm{cl}_C(Y)$ is easily seen to be equivalent to e') for a arbitrary convergence C.

Every L*-convergence C can be computed from the topology $\mathcal{T}(C)$.

Definition 5.6 (The convergence of a topology). If $(X,\mathcal{T})$ is a topological space then the convergence $C(\mathcal{T})$ is defined for sequences $(x_i)_{i\in\mathbb{N}}\subset X$ and $x\in X$ by

$$(x_i)_{i\in\mathbb{N}}\xrightarrow{C(\mathcal{T})}x\quad:\Longleftrightarrow\qquad\text{if } x\in U\in\mathcal{T}\text{ then } x_i\in U\text{ for large } i.$$

For any topological space $(X,\mathcal{T})$ it is easy to show $\mathcal{T}\subset\mathcal{T}(C(\mathcal{T}))$. Similarly, if C is a convergence on X, then $C\subset C(\mathcal{T}(C))$ holds. An equality condition was proved by Kisyński, see Problem 1.7.19 in [25]:

Theorem 5.7. *If (X,C) is an L*-space, then $C(\mathcal{T}(C))=C$.*

Divergence functions in Definition 5.13 will generalize metric spaces.

Example 5.8 (Metric spaces). Let (X,d) be a metric space for $d:X\times X\to\mathbb{R}$. Then $x_n\xrightarrow{C_d}x:\Longleftrightarrow\lim_{i\to\infty}d(x,x_i)=0$ defines an L*-convergence C_d on X, such that e) holds. Moreover, a base of the *metric topology* $\mathcal{T}(C_d)$ at $x\in X$ is given by the open disks $B(x,\epsilon):=\{y\in X\mid d(x,y)<\epsilon\}$ for $\epsilon>0$. Since rational values of ϵ suffice, a metric topology is first countable.

Continuity will allow to generalize the idea that open disks define a base.

Definition 5.9 (Continuity). Let $f:X\to X'$ be a function and C resp. C' be a convergence on X resp. X'. Then f is *continuous* for C and C' at $x\in X$ if $f(x_n)\xrightarrow{C'}f(x)$ whenever $x_n\xrightarrow{C}x$. The function f is *continuous* for C and C' if f is continuous for C and C' at every $x\in X$. If $\mathcal{T}$ resp. $\mathcal{T}'$ is a topology on X resp. X', then f is continuous for $\mathcal{T}$ and $\mathcal{T}'$ if $f^{-1}(U')$ is $\mathcal{T}$ open for every $\mathcal{T}'$ open set $U'\subset X'$.

The following statement is an excerpt of Theorem 2.2 in [22]. Dudley restricts to convergences satisfying a) and b) in Definition 5.1. But the proof works for arbitrary convergences as well.

Theorem 5.10. *Let $f:X\to X'$ be a function and C resp. C' be a convergence on X resp. X'.*

1. *If f is continuous for C and C', then f is continuous for $\mathcal{T}(C)$ and $\mathcal{T}(C')$.*
2. *If (X',C') is an L*-space then f is continuous for C and C' if and only if f is continuous for $\mathcal{T}(C)$ and $\mathcal{T}(C')$.*

Subspaces will allow us to relate several topologies to each other.

Definition 5.11 (Subspaces). Let $B\subset X$. If $\mathcal{T}$ is a topology on X, then the *subspace topology*

$$\mathcal{T}|_B:=\{B\cap U\mid U\in\mathcal{T}\}$$

is defined. If C is a convergence on X, we have the *subspace convergence*

$$C|_B:=C\cap(B^{\mathbb{N}}\times B)\,.$$

We want to allow infinite "distances" appearing in the relative entropy.

Example 5.12 (One point compactification). Let $I = \mathbb{R}$ or $I = [a, \infty)$ for $a \in \mathbb{R}$. The *Alexandroff compactification* of I is a topology $\mathcal{T}^{\mathrm{c}}$ on $I \cup \{\infty\}$, where $\mathcal{T}^{\mathrm{c}}$ open sets are norm open subsets of I or they are of the form $I \cup \{\infty\} \setminus F$, where $F \subset I$ is norm compact. Theorem 3.5.11 in [25] shows that $(I \cup \{\infty\}, \mathcal{T}^{\mathrm{c}})$ is a compact Hausdorff topological space. The convergence $C^{\mathrm{c}} := C(\mathcal{T}^{\mathrm{c}})$ is

$$\{((x_i), x) \in [0,\infty]^{\mathbb{N}} \times [0,\infty) \mid x_i < \infty \text{ for large } i \text{ and } \lim_{i\to\infty} x_i = x\}$$
$$\cup\ \{((x_i), \infty) \mid (x_i) \subset [0,\infty] \text{ such that } \forall R \in [0,\infty) \text{ we have } x_i \geq R \text{ for large } i\}\,.$$

It is easy to show $\mathcal{T}(C(\mathcal{T}^{\mathrm{c}})) \subset \mathcal{T}^{\mathrm{c}}$ (every $U \in \mathcal{T}(C(\mathcal{T}^{\mathrm{c}}))$ including ∞ has a bounded complement and with each real $x \in U$ there is a disk $B(x, \epsilon)$ in U). The converse inclusion holds for arbitrary topologies so we have

$$\mathcal{T}(C^{\mathrm{c}}) = \mathcal{T}^{\mathrm{c}}\,.$$

It is easy to show that C^{c} is an L*-convergence, hence Theorem 5.10.2 show for every convergence C on a set X and any function $f : X \to I \cup \{\infty\}$ that f is continuous for C and C^{c} if and only if f is continuous for $\mathcal{T}(C)$ and $\mathcal{T}^{\mathrm{c}}$.

We are ready to study the I-/rI-topology abstractly. In the sequel we will use the compactification $[0,\infty]$ of the non-negative half-axis $[0,\infty)$. We shall frequently write $\lim_{i\to\infty} x_i = x$ in place of $x_i \xrightarrow{C^{\mathrm{c}}} x$ for $(x_i)_{i\in\mathbb{N}} \subset [0,\infty]$ and $x \in [0,\infty]$.

Definition 5.13 (Divergence functions). 1. A *divergence function* on a set X is a function $f : X \times X \to [0,\infty]$, such that for all $x \in X$ we have $f(x,x) = 0$. Let C_f be the convergence on X defined by

$$x_n \xrightarrow{C_f} x \ :\Longleftrightarrow\ \lim_{n\to\infty} f(x, x_n) = 0\,.$$

Two extra assumptions on a divergence function f on X suffice for our purpose to analyze the I-/rI-convergence:

A) An abstract *Pinsker-Csiszár inequality* holds, i.e. (X, d) is a metric space and there is a function $g : [0,\infty] \to [0,\infty]$, continuous for C^{c} and C^{c} at 0, such that $g(0) = 0$ and such that for all $x, y \in X$ we have $d(x,y) \leq g(f(x,y))$.

B) The divergence function f is continuous in the second argument, i.e. for all $x \in X$ the function $X \to [0,\infty]$, $y \mapsto f(x,y)$ is continuous for C_f and C^{c}.

2. For $x \in X$ and $\epsilon \in (0,\infty]$ we define the *open* resp. *closed f-disk*

$$V^f(x,\epsilon) := \{y \in X \mid f(x,y) < \epsilon\} \qquad \text{resp. } W^f(x,\epsilon) := \{y \in X \mid f(x,y) \leq \epsilon\}\,.$$

If f is the relative entropy between probability measures on $(\mathbb{N}, 2^{\mathbb{N}})$, then property B) fails and property A) holds by the Pinsker-Csiszár inequality, see §2.1.

Lemma 5.14 (Divergence functions). *Let f be a divergence function on a set X. Then the convergence C_f is an L^*-convergence. In particular $C(\mathcal{T}(C_f)) = C_f$. The sequential closure of $Y \subset X$ is*

$$\begin{aligned}\mathrm{cl}_{C_f}(Y) &= \{x \in X \mid \lim_{n\to\infty} f(x, y_n) = 0 \text{ for a sequence } (y_n)_{n\in\mathbb{N}} \subset Y\}\\ &= \{x \in X \mid \inf_{n\in\mathbb{N}} f(x, y_n) = 0 \text{ for a sequence } (y_n)_{n\in\mathbb{N}} \subset Y\}\\ &= \{x \in X \mid \inf_{y\in Y} f(x, y) = 0\}\,.\end{aligned}$$

1. Let f satisfy property A) *in Definition* 5.13.1 *for a metric $d : X \times X \to \mathbb{R}$. Then C_f has unique limits. We have $C_f \subset C_d$ and $\mathcal{T}(C_f) \supset \mathcal{T}(C_d)$. In particular $\mathcal{T}(C_f)$ is a Hausdorff topology.*

2. The property B) *in Definition* 5.13.1 *is equivalent with the property that for all $x \in X$ the function $X \to [0, \infty]$, $y \mapsto f(x, y)$ is continuous for $\mathcal{T}(C_f)$ and $\mathcal{T}^{\mathrm{c}}$.*

Property B) *implies that for each $x \in X$ and $\epsilon \in (0, \infty]$ the open f-disk $V^f(x, \epsilon)$ is $\mathcal{T}(C_f)$ open and the closed f-disk $W^f(x, \epsilon)$ is $\mathcal{T}(C_f)$ closed. It follows that the open f-disks $\{V^f(x, \epsilon) \mid \epsilon > 0\}$ are a base for $(X, \mathcal{T}(C_f))$ at x. This shows that $\mathcal{T}(C_f)$ is first countable and for any subset $Y \subset X$ we have $\mathcal{T}(C_f)|_Y = \mathcal{T}(C_f|_Y)$.*

Property B) *implies that the L^*-convergence C_f has property* e) *in Definition* 5.3. *This shows for any subset $Y \subset X$ that the sequential closure $\mathrm{cl}_{C_f}(Y)$ is the $\mathcal{T}(C_f)$ closure of Y.*

Proof. Clearly C_f is an L*-convergences and then $C(\mathcal{T}(C_f)) = C_f$ follows from Theorem 5.7. The statements about the sequential closure are clear.

Property A). Let us prove unique limits, i.e. d) in Definition 5.1. Let $x \in X$ and $(x_i)_{i\in\mathbb{N}} \subset X$. Assuming $x_n \xrightarrow{C_f} x$, i.e. $\lim_{i\to\infty} f(x, x_i) = 0$, the continuity of g at zero (for C^{c}) gives

$$\lim_{i\to\infty} g \circ f(x, x_i) = 0\,.$$

For all $i \in \mathbb{N}$ we have by assumption $d(x, x_i) \le g \circ f(x, x_i)$, so $\lim_{i\to\infty} d(x, x_i) = 0$. Limits are unique in a metric space so this translates to the convergence C_f. We have thereby proved $C_f \subset C_d$. If follows that $\mathcal{T}(C_f) \supset \mathcal{T}(C_d)$. Since $\mathcal{T}(C_d)$ is Hausdorff, so is $\mathcal{T}(C_f)$.

Property B). For all $x \in X$ the continuity of the function $X \to [0, \infty]$, $y \mapsto f(x, y)$ for C_f and C^{c} is equivalent to the continuity for $\mathcal{T}(C_f)$ and $\mathcal{T}^{\mathrm{c}}$ according to the discussion in the last paragraph of Example 5.12.

Hence, if property B) holds, then the preimage of every $\mathcal{T}^{\mathrm{c}}$ open resp. closed subset of $[0, \infty]$ is $\mathcal{T}(C_f)$ open resp. closed. In particular, every open resp. closed f-disk is $\mathcal{T}(C_f)$ open resp. closed. The open f-disks $\{V^f(x, \epsilon) \mid \epsilon > 0\}$ define a base for $(X, \mathcal{T}(C_f))$ at $x \in X$: By contradiction, let U be $\mathcal{T}(C_f)$ open, $x \in U$ and let us assume that U contains no open f-disk about x. Then there exists a sequence $(x_i)_{i\in\mathbb{N}} \subset X \setminus U$ with

$$(x_i)_{i\in\mathbb{N}} \xrightarrow{C_f} x\,.$$

But $X \setminus U$ is $\mathcal{T}(C_f)$ closed and so by Remark 5.5 it contains all C_f-limits of sequences in $X \setminus U$. So $x \in X \setminus U$ contradicts the assumption $x \in U$. The space $(X, \mathcal{T}(C_f))$ is first countable, e.g. $\{V^f(x, 1/n) \mid n \in \mathbb{N}\}$ is a base at $x \in X$.

Let us consider a subspace $Y \subset X$. Then $\mathcal{T}(C)|_Y \subset \mathcal{T}(C|_Y)$ is easy to show. Conversely, for all $y \in Y$ and $\epsilon > 0$ we have

$$V^{f|_{Y\times Y}}(y,\epsilon) = V^f(y,\epsilon) \cap Y\,.$$

The divergence function $f|_{Y\times Y}$ on Y satisfies B), hence a set $U \in \mathcal{T}(C|_Y)$ equals

$$U = \bigcup_{\alpha\in I} V^{f|_{Y\times Y}}(y_\alpha,\epsilon_\alpha) = \left(\bigcup_{\alpha\in I} V^f(y_\alpha,\epsilon_\alpha)\right) \cap Y$$

for some $y_\alpha \in Y$ and $\epsilon_\alpha > 0$, $\alpha \in I$. We have proved $U \in \mathcal{T}(C_f)|_Y$.

To prove property e) we use for each $x \in X$ the continuity of the function $X \to [0,\infty]$, $y \mapsto f(x,y)$ for C_f and C^{c} in an open f-disk $V^f(x,\epsilon)$ for some $\epsilon > 0$. If $(x_i)_{i\in\mathbb{N}} \xrightarrow{C_f} x$ then there exists a sequence of positive numbers $(\epsilon_i)_{i\in\mathbb{N}} \xrightarrow{i\to\infty} 0$, such that $f(x,x_i) < \epsilon_i$ for all i. For every $i \in \mathbb{N}$ we choose a sequence $(x^i_j)_{j\in\mathbb{N}} \subset X$ such that $(x^i_j)_{j\in\mathbb{N}} \xrightarrow{C_f} x_i$. By continuity of $f(x,\cdot)$ for C_f and C^{c} there exists $m_i \in \mathbb{N}$ for all i such that $f(x,x^i_j) < \epsilon_i$ for all $j \geq m_i$. Then $f(x,x^i_{m_i}) \leq \epsilon_i$ for all i implies

$$\lim_{i\to\infty} f(x,x^i_{m_i}) \leq \lim_{i\to\infty} \epsilon_i = 0\,.$$

This proves property e) for C_f. A consequence for any $Y \subset X$ is that $\mathrm{cl}_{C_f}(Y)$ is the $\mathcal{T}(C_f)$ closure of Y (see Remark 5.5). $\square$

5.2. The I-topology and the rI-topology

The relative entropy $S : \mathcal{S} \times \mathcal{S} \to [0,\infty]$ defines two divergence functions. Some results are formulated in terms of the convex geometry of the state space. Corollary 5.19 collects topological conditions for a commutative algebra. Several definitions appear already in §2.3, e.g. for $\rho,\sigma \in \mathcal{S}_{\mathcal{A}}$ the functions $S^{\mathrm{I}}(\rho,\sigma) = S(\sigma,\rho)$ and $S^{\mathrm{rI}}(\rho,\sigma) = S(\rho,\sigma)$ are defined. In the sequel let $\omega \in \{\mathrm{I}, \mathrm{rI}\}$.

Definition 5.15. Let $(\rho_i)_{i\in\mathbb{N}} \subset \mathcal{S}$ be a sequence and let $\rho \in \mathcal{S}$. We define the *ω-convergence* C^ω on $\mathcal{S}$ by

$$\rho_i \xrightarrow{C^\omega} \rho \;:\Longleftrightarrow\; \lim_{i\to\infty} S^\omega(\rho,\rho_i) = 0\,.$$

The topology $\mathcal{T}^\omega = \mathcal{T}(C^\omega)$ on $\mathcal{S}$ is called *ω-topology.* We denote the norm convergence on $\mathcal{S}$ by $C^{\|\cdot\|}$ and the norm topology on $\mathcal{S}$ by $\mathcal{T}^{\|\cdot\|} := \mathcal{T}(C^{\|\cdot\|})$.

We begin with continuity of the relative entropy, using the L*-convergence C^{c} on $[0,\infty]$ corresponding to the Alexandroff compactification, see Example 5.12.

Proposition 5.16. *For every state $\rho\in\mathcal{S}$ the mapping $\mathcal{S}\to[0,\infty]$, $\sigma\mapsto S^{\omega}(\rho,\sigma)$ is continuous for C^{ω} and C^{c}.*

Proof. Concerning the I-convergence, we have to show for $\rho,\sigma\in\mathcal{S}$ and $(\tau_i)_{i\in\mathbb{N}}\subset\mathcal{S}$ that $\lim_{i\to\infty}S(\tau_i,\sigma)=0$ implies $\lim_{i\to\infty}S(\tau_i,\rho)=S(\sigma,\rho)$. Let us first assume that $s(\rho)\succeq s(\sigma)$ holds, i.e. $S(\sigma,\rho)<\infty$. Since $\lim_{i\to\infty}S(\tau_i,\sigma)=0$ we have $s(\sigma)\succeq s(\tau_i)$ for large i and hence $s(\rho)\succeq s(\tau_i)$ holds for large i. By the Pinsker-Csiszár inequality (15) the sequence $(\tau_i)_{i\in\mathbb{N}}$ converges to σ in norm. Hence the continuity of the von Neumann entropy, see e.g. §II.A in [57], proves

$$S(\tau_i,\rho)=-S(\tau_i)-\operatorname{tr}\tau_i\log(\rho)\overset{i\to\infty}{\longrightarrow}-S(\sigma)-\operatorname{tr}\sigma\log(\rho)=S(\sigma,\rho)\,.$$

Second, we consider $s(\rho)\not\succeq s(\sigma)$, i.e. $S(\sigma,\rho)=\infty$. By Remark 2.4.1 the relative entropy is lower semi-continuous. We obtain $\liminf_{i\to\infty}S(\tau_i,\rho)\geq S(\sigma,\rho)=\infty$ and this implies $\lim_{i\to\infty}S(\tau_i,\rho)=\infty$.

Concerning the rI-convergence, we have to show that $\lim_{i\to\infty}S(\sigma,\tau_i)=0$ implies $\lim_{i\to\infty}S(\rho,\tau_i)=S(\rho,\sigma)$. If $s(\rho)\not\preceq s(\sigma)$ then $S(\rho,\sigma)=\infty$ and the lower semi-continuity of the relative entropy proves $\lim_{i\to\infty}S(\rho,\tau_i)=\infty$ as in the previous paragraph. Finally we consider $s(\rho)\preceq s(\sigma)$ with $S(\rho,\sigma)<\infty$. Since $S(\sigma,\tau_i)\overset{i\to\infty}{\longrightarrow}0$ we have $s(\sigma)\preceq s(\tau_i)$ for large i. Lemma 4.26 completes the proof. □

The norm topology is too coarse for a similar continuity result, see e.g. Example 2.2. We now prove that ω-closures do not decrease the relative entropy. For $X\subset\mathcal{S}_{\mathcal{A}}$ we use $S^{\omega}(\rho,X)=\inf_{\sigma\in X}S^{\omega}(\rho,\sigma)$ and the ω-closure from (20).

Corollary 5.17. *Let $\rho\in\mathcal{S}$ and $X\subset\mathcal{S}$. Then $S^{\omega}(\rho,X)=S^{\omega}(\rho,\mathrm{cl}^{\omega}(X))$ holds.*

Proof. For every state $\sigma\in\mathrm{cl}^{\omega}(X)$ there exists by (20) a sequence $(\sigma_i)_{i\in\mathbb{N}}\subset X$, such that $\sigma_i\overset{C^{\omega}}{\longrightarrow}\sigma$. Proposition 5.16 shows that the relative entropies converge, $\lim_{i\to\infty}S^{\omega}(\rho,\sigma_i)=S^{\omega}(\rho,\sigma)$. Hence

$$S^{\omega}(\rho,X)=\inf_{\tau\in X}S^{\omega}(\rho,\tau)\leq\inf_{i\in\mathbb{N}}S^{\omega}(\rho,\sigma_i)\leq\lim_{i\to\infty}S^{\omega}(\rho,\sigma_i)=S^{\omega}(\rho,\sigma)\,.$$

Taking the infimum over all $\sigma\in\mathrm{cl}^{\omega}(X)$, we get $S(\rho,X)\leq S(\rho,\mathrm{cl}^{\omega}(X))$. The converse inequality is trivial. □

We now investigate the ω-topology of the state space $\mathcal{S}$. For $\rho\in\mathcal{S}$ and $\epsilon\in(0,\infty]$ we use the open ω-disk resp. closed ω-disk defined in (13) resp. (14) and the face lattice $\mathcal{F}$ of the state space $\mathcal{S}$, introduced in §4.1.

Theorem 5.18 (Information topology and reverse information topology). *The convergence C^{ω} is an L*-convergence. In particular $C(\mathcal{T}^{\omega})=C^{\omega}$.*

The sequential closure (57) *of* $X \subset \mathcal{S}$ *equals the* ω*-closure* $\mathrm{cl}^\omega(X)$ *from* (20),

$$\begin{aligned}\mathrm{cl}^\omega(X) &= \{\rho \in \mathcal{S} \mid \lim_{i\to\infty} S^\omega(\rho,\rho_i) = 0 \textit{ for a sequence } (\rho_i)_{i\in\mathbb{N}} \subset X\} \\ &= \{\rho \in \mathcal{S} \mid \inf_{i\in\mathbb{N}} S^\omega(\rho,\rho_i) = 0 \textit{ for a sequence } (\rho_i)_{i\in\mathbb{N}} \subset X\} \\ &= \{\rho \in \mathcal{S} \mid S^\omega(\rho, X) = 0\}\,.\end{aligned} \tag{58}$$

1. *The* L^**-convergence* C^ω *has unique limits. We have* $C^\omega \subset C^{\|\cdot\|}$ *and* $\mathcal{T}^\omega \supset \mathcal{T}^{\|\cdot\|}$. *In particular* $\mathcal{T}^\omega$ *is a Hausdorff topology.*
2. *For every* $\rho \in \mathcal{S}$ *the mapping* $\mathcal{S} \to [0,\infty]$, $\sigma \mapsto S^\omega(\rho,\sigma)$ *is continuous for* C^ω *and* C^{c} *and continuous for* $\mathcal{T}^\omega$ *and* $\mathcal{T}^{\mathrm{c}}$.
   For each $\rho \in \mathcal{S}$ *and* $\epsilon \in (0,\infty]$ *the open* ω*-disk* $V^\omega(\rho,\epsilon)$ *is* $\mathcal{T}^\omega$ *open and the closed* ω*-disk* $W^\omega(\rho,\epsilon)$ *is* $\mathcal{T}^\omega$ *closed. The open* ω*-disks* $\{V^\omega(\rho,\epsilon) \mid \epsilon \in (0,\infty]\}$ *are a base for* $(\mathcal{S},\mathcal{T}^\omega)$ *at* ρ. *In particular* $\mathcal{T}^\omega$ *is first countable and for any subset* $X \subset \mathcal{S}$ *we have* $\mathcal{T}^\omega|_X = \mathcal{T}(C^\omega|_X)$.
   For any subset $X \subset \mathcal{S}$ *the sequential closure* $\mathrm{cl}^\omega(X)$ *is the* $\mathcal{T}^\omega$ *closure of* X.
3. *Every term in the partition* $\mathcal{S} = \bigcup_{F\in\mathcal{F}} \mathrm{ri}\, F$ *is a* $\mathcal{T}^{\mathrm{I}}$ *connected component of* $\mathcal{S}$. *For all faces* $F \in \mathcal{F}$ *we have* $C^{\mathrm{I}}|_{\mathrm{ri}\, F} = C^{\|\cdot\|}|_{\mathrm{ri}\, F}$ *and* $\mathcal{T}^{\mathrm{I}}|_{\mathrm{ri}\, F} = \mathcal{T}^{\|\cdot\|}|_{\mathrm{ri}\, F}$.
4. *We have* $\mathcal{T}^{\|\cdot\|} \subset \mathcal{T}^{\mathrm{rI}} \subset \mathcal{T}^{\mathrm{I}}$ *and* $C^{\|\cdot\|} \supset C^{\mathrm{rI}} \supset C^{\mathrm{I}}$.
5. *The* $\mathcal{T}^{\mathrm{rI}}$ *closure of* $\mathrm{ri}\,\mathcal{S}$ *is* $\mathcal{S}$ *and the topological space* $(\mathcal{S},\mathcal{T}^{\mathrm{rI}})$ *is connected.*

Proof. Both divergence functions $S^{\mathrm{I}}(\rho,\sigma) = S(\sigma,\rho)$ and $S^{\mathrm{rI}}(\rho,\sigma) = S(\rho,\sigma)$ defined for $\rho,\sigma \in \mathcal{S}$ are divergence functions in the sense of Definition 5.13.1. They satisfy condition A) and B) according to the Pinsker-Csiszár inequality (15) and Proposition 5.16. So Lemma 5.14 proves the theorem up to part *2* inclusive.

We show part 3. According to part *2*, for every $\rho \in \mathcal{S}$ the open I-disk of infinite radius is $\mathcal{T}^{\mathrm{I}}$ open and has by Remark 4.7 the form

$$V^{\mathrm{I}}(\rho,\infty) = \{\sigma \in \mathcal{S} \mid S(\sigma,\rho) < \infty\} = \{\sigma \in \mathcal{S} \mid s(\sigma) \preceq s(\rho)\} = \mathbb{F}(s(\rho))\,. \tag{59}$$

By the lattice isomorphism $\mathbb{F} : \mathcal{P} \to \mathcal{F}$ in Corollary 4.6 we obtain that every face F of $\mathcal{S}$ is $\mathcal{T}^{\mathrm{I}}$ open. Let us show that $\mathrm{ri}\, F$ is $\mathcal{T}^{\mathrm{I}}$ open. The complement $\mathcal{S} \setminus F$ of F is $\mathcal{T}^{\mathrm{I}}$ closed and the relative boundary $\mathrm{rb}\, F$ of F is norm closed. By part *2* we have $\mathcal{T}^{\|\cdot\|} \subset \mathcal{T}^{\mathrm{I}}$ hence $\mathrm{rb}\, F$ is $\mathcal{T}^{\mathrm{I}}$ closed as well. So

$$\mathrm{ri}\, F = \mathcal{S} \setminus (\mathrm{rb}\, F \cup (\mathcal{S} \setminus F))$$

is $\mathcal{T}^{\mathrm{I}}$ open. Finally, the relative interior $\mathrm{ri}\, F$ is also $\mathcal{T}^{\mathrm{I}}$ closed because by the stratification (44) we have $\mathrm{ri}\, F = \mathcal{S} \setminus (\bigcup_{G\neq F} \mathrm{ri}\, G)$, the union extending over faces $G \in \mathcal{F}$.

Let $F \in \mathcal{F}$ be an arbitrary face. Since the relative interior of F consists of states of constant support (see Remark 4.7), the relative entropy is norm continuous on $\mathrm{ri}\, F \times \mathrm{ri}\, F$. Hence we have $C^{\|\cdot\|}|_{\mathrm{ri}\, F} \subset C^{\mathrm{I}}|_{\mathrm{ri}\, F}$. This shows $C^{\|\cdot\|}|_{\mathrm{ri}\, F} = C^{\mathrm{I}}|_{\mathrm{ri}\, F}$ as the converse inclusion follows from the Pinsker-Csiszár inequality. With part *2* we have

$$\mathcal{T}^{\mathrm{I}}|_{\mathrm{ri}\,F} = \mathcal{T}(C^{\mathrm{I}}|_{\mathrm{ri}\,F}) = \mathcal{T}(C^{\|\cdot\|}|_{\mathrm{ri}\,F}) = \mathcal{T}^{\|\cdot\|}|_{\mathrm{ri}\,F}\,.$$

We show part 4. We begin with a proof of $\mathcal{T}^{\mathrm{rI}} \subset \mathcal{T}^{\mathrm{I}}$. We first notice $C^{\mathrm{rI}}|_{\mathrm{ri}\,F} = C^{\mathrm{I}}|_{\mathrm{ri}\,F}$ for every face F of $\mathcal{S}$. This follows from $C^{\|\cdot\|}|_{\mathrm{ri}\,F} = C^{\mathrm{I}}|_{\mathrm{ri}\,F}$ proved in part *3* and from $C^{\|\cdot\|}|_{\mathrm{ri}\,F} = C^{\mathrm{rI}}|_{\mathrm{ri}\,F}$, which can be proved analogously. Also

$$\mathcal{T}(C^{\mathrm{rI}})|_{\mathrm{ri}\,F} = \mathcal{T}(C^{\mathrm{I}})|_{\mathrm{ri}\,F}\,.$$

Let $U \in \mathcal{T}^{\mathrm{rI}}$. Then $U \cap \mathrm{ri}\,F \in \mathcal{T}^{\mathrm{rI}}|_{\mathrm{ri}\,F} = \mathcal{T}^{\mathrm{I}}|_{\mathrm{ri}\,F}$ and since $\mathrm{ri}\,F$ is $\mathcal{T}^{\mathrm{I}}$ open by part *3*, this shows $U \cap \mathrm{ri}\,F \in \mathcal{T}^{\mathrm{I}}$. Now $U = \bigcup_{F\in\mathcal{F}}(U \cap \mathrm{ri}\,F) \in \mathcal{T}^{\mathrm{I}}$ and we have proved $\mathcal{T}^{\mathrm{rI}} \subset \mathcal{T}^{\mathrm{I}}$. Part *1* adds the inequality $\mathcal{T}^{\|\cdot\|} \subset \mathcal{T}^{\mathrm{rI}} \subset \mathcal{T}^{\mathrm{I}}$. Since these three topologies arise from L*-convergences we get $C^{\|\cdot\|} \supset C^{\mathrm{rI}} \supset C^{\mathrm{I}}$ from Theorem 5.7.

We show part 5. We first show that any non-empty $\mathcal{T}^{\mathrm{rI}}$ open set $U \subset \mathcal{S}$ intersects $\mathrm{ri}\,\mathcal{S}$. By part *2* the set U contains an open rI-disk $V^{\mathrm{rI}}(\rho,\epsilon) = \{\sigma \in \mathcal{S} \mid S(\rho,\sigma) < \epsilon\}$ for some density matrix ρ and $\epsilon > 0$. We show that $V^{\mathrm{rI}}(\rho,\epsilon)$ intersects $\mathrm{ri}\,\mathcal{S}$. The relative interior $\mathrm{ri}\,\mathcal{S}$ consists of all invertible density matrices by Proposition 4.5. So for any fixed $\tau \in \mathrm{ri}\,\mathcal{S}$ we have $S(\rho,\tau) < \infty$ and then (12) implies

$$0 = S(\rho,\rho) = \lim_{\lambda\nearrow 1} S(\rho,(1-\lambda)\tau + \lambda\rho)\,,$$

whence $(1-\lambda)\tau+\lambda\rho \in V^{\mathrm{rI}}(\rho,\epsilon)$ for $\lambda \nearrow 1$. Since $(1-\lambda)\tau+\lambda\rho \in \mathrm{ri}\,\mathcal{S}$ is invertible for $\lambda < 1$, this shows that U intersects $\mathrm{ri}\,\mathcal{S}$.

As shown in the previous paragraph, the relative boundary $\mathrm{rb}\,\mathcal{S}$ does not contain a $\mathcal{T}^{\mathrm{rI}}$ open set so the $\mathcal{T}^{\mathrm{rI}}$ closure of $\mathrm{ri}\,\mathcal{S}$ equals $\mathcal{S}$. We show that $\mathcal{S}$ is $\mathcal{T}^{\mathrm{rI}}$ connected. By part *3* and *4* we have $\mathcal{T}^{\mathrm{rI}}|_{\mathrm{ri}\,\mathcal{S}} = \mathcal{T}^{\|\cdot\|}|_{\mathrm{ri}\,\mathcal{S}}$. The convex set $\mathrm{ri}\,\mathcal{S}$ is connected in the norm topology hence in the $\mathcal{T}^{\mathrm{rI}}$ topology. The claim follows since the closure of a connected set is connected, see e.g. §IV.7 in [11]. □

The following conditions have applications to exponential families in §6.6.

Corollary 5.19. *If* $\dim_{\mathbb{C}}(\mathcal{A}) > 1$*, then* $C^{\mathrm{I}} \subsetneq C^{\mathrm{rI}}$*,* $\mathcal{T}^{\mathrm{I}} \supsetneq \mathcal{T}^{\mathrm{rI}}$ *and* $\mathcal{S}$ *is not* $\mathcal{T}^{\mathrm{I}}$ *compact. The following assertions are equivalent.*

1. $\mathcal{A}$ *is commutative,*
2. $\mathcal{T}^{\mathrm{I}}$ *is second countable,*
3. $\mathcal{T}^{\mathrm{rI}}$ *is second countable,*
4. $C^{\mathrm{rI}} = C^{\|\cdot\|}$,
5. $\mathcal{T}^{\mathrm{rI}} = \mathcal{T}^{\|\cdot\|}$,
6. $\mathcal{S}$ *is* $\mathcal{T}^{\mathrm{rI}}$ *compact.*

Proof. *Item 1 implies 4.* If $\mathcal{A}$ is commutative, then by (87) it is isomorphic to $\mathbb{C}^n$. We can argue by convergence in components of $\mathbb{C}^n$ and find $C^{\|\cdot\|} = C^{\mathrm{rI}}$.

Statements in the headline. If $\dim_{\mathbb{C}}(\mathcal{A}) > 1$, then by (87) $\mathcal{A}$ contains a C*-subalgebra $\mathcal{B} \cong \mathbb{C}^2$ and by Example 2.2 we have $C^{\mathrm{I}}|_{\mathcal{S}_\mathcal{B}} \subsetneq C^{\|\cdot\|}|_{\mathcal{S}_\mathcal{B}}$ while $C^{\|\cdot\|}|_{\mathcal{S}_\mathcal{B}} = C^{\mathrm{rI}}|_{\mathcal{S}_\mathcal{B}}$ was shown in the previous paragraph. Now the inclusion $C^{\mathrm{I}} \subset C^{\mathrm{rI}}$ in Theorem 5.18.4 shows $C^{\mathrm{I}} \subsetneq C^{\mathrm{rI}}$. In terms of topology, since $C^\omega = C(\mathcal{T}^\omega)$ holds for $\omega \in \{\mathrm{I}, \mathrm{rI}\}$ by Theorem 5.18.1, we have also $\mathcal{T}^{\mathrm{rI}} \subsetneq \mathcal{T}^{\mathrm{I}}$.

We show that $\mathcal{S}$ is not $\mathcal{T}^{\mathrm{I}}$ compact if $\dim_{\mathbb{C}}(\mathcal{A}) > 1$. Theorem 5.18.3 shows $(\mathrm{ri}\,\mathcal{S}, \mathcal{T}^{\mathrm{I}}|_{\mathrm{ri}\,\mathcal{S}}) = (\mathrm{ri}\,\mathcal{S}, \mathcal{T}^{\|\cdot\|}|_{\mathrm{ri}\,\mathcal{S}})$. But $(\mathrm{ri}\,\mathcal{S}, \mathcal{T}^{\|\cdot\|}|_{\mathrm{ri}\,\mathcal{S}})$ is not a compact topological space since $\mathrm{ri}\,\mathcal{S}$ is the relative interior of a convex set of dimension > 0. Then $\mathcal{S}$ is not $\mathcal{T}^{\mathrm{I}}$ compact because $\mathrm{ri}\,\mathcal{S}$ is its $\mathcal{T}^{\mathrm{I}}$ connected component.

Item 4 implies 5. By definition $\mathcal{T}^{\mathrm{rI}} = \mathcal{T}(C^{\mathrm{rI}})$ and $\mathcal{T}^{\|\cdot\|} = \mathcal{T}(C^{\|\cdot\|})$.

Item 5 implies 3 and 6. By Proposition 4.5 the state space $\mathcal{S}$ is a convex body, hence is a (norm) compact metric space. On the other hand, a compact metric space is second countable, see e.g. §V.4–5 in [11]. Since $\mathcal{T}^{\mathrm{rI}} = \mathcal{T}^{\|\cdot\|}$ is assumed, the state space is $\mathcal{T}^{\mathrm{rI}}$ compact and $\mathcal{T}^{\mathrm{rI}}$ second countable.

Item 1 implies 2. For every face $F \in \mathcal{F}$ we have $\mathcal{T}^{\mathrm{I}}|_{\mathrm{ri}\,F} = \mathcal{T}^{\|\cdot\|}|_{\mathrm{ri}\,F}$ by Theorem 5.18.3. As shown in the previous paragraph, $\mathcal{T}^{\|\cdot\|}|_F$ is second countable. Since $\mathrm{ri}\,F$ is an $\mathcal{T}^{\|\cdot\|}|_F$ open subset of F, the topology $\mathcal{T}^{\mathrm{I}}|_{\mathrm{ri}\,F} = \mathcal{T}^{\|\cdot\|}|_{\mathrm{ri}\,F}$ is second countable. The simplex $\mathcal{S}$ is partitioned into finitely many relative interiors $\mathrm{ri}\,F$ of faces F by (44). These sets are $\mathcal{T}^{\mathrm{I}}$ connected components of $\mathcal{S}$, so the proof is complete.

Auxiliary calculation. To show that each of items *2*, *3* or *6* implies *1*, we show that $\mathcal{S}$ has an open cover, indexed by pure states, without a proper subcover. If the algebra $\mathcal{A}$ is non-commutative, then this cover is uncountable. By Remark 4.7 we can write for any state $\rho \in \mathcal{S}$ the open rI-disk of infinite radius in the form

$$V^{\mathrm{rI}}(\rho, \infty) = \{\sigma \in \mathcal{S} \mid s(\rho) \preceq s(\sigma)\} = \bigcup_{\substack{p \in \mathcal{P} \\ p \succeq s(\rho)}} \mathrm{ri}\,\mathbb{F}(p)\,. \tag{60}$$

Here $\mathcal{P}$ denotes the projection lattice of $\mathcal{A}$. The open rI-disks are $\mathcal{T}^{\mathrm{rI}}$ open by Theorem 5.18.2. For pure state $p, q \in \mathcal{P} \cap \mathcal{S}$ we have

$$p \notin V^{\mathrm{rI}}(q, \infty) \quad \text{if } p \neq q\,.$$

If $\mathcal{A}$ is non-commutative then $\mathcal{A}$ contains a C*-subalgebra isomorphic to $\mathrm{Mat}(2, \mathbb{C})$, see (87), hence $\mathcal{P} \cap \mathcal{S}$ is uncountable.

Item 6 implies 1. The open cover $\bigcup_{p \in \mathcal{P} \cap \mathcal{S}} V^{\mathrm{rI}}(p, \infty)$ of $\mathcal{S}$ has no finite subcover.

Item 3 implies 1. If $\mathcal{B}$ is a base of $\mathcal{T}^{\mathrm{rI}}$, then for all $p \in \mathcal{P} \cap \mathcal{S}$ there is a $\mathcal{T}^{\mathrm{rI}}$ open set $U_p \in \mathcal{B}$ such that $p \in U_p \subset V^{\mathrm{rI}}(p, \infty)$. The map $\mathcal{P} \cap \mathcal{S} \to \mathcal{B}$, $p \mapsto U_p$ is injective. This prove that $\mathcal{B}$ is not countable.

Item 2 implies 1. Theorem 5.18.4 shows $\mathcal{T}^{\mathrm{rI}} \subset \mathcal{T}^{\mathrm{I}}$ so the arguments in the previous paragraph apply unmodified. □

6. Exponential families

We study an exponential family $\mathcal{E}$ in a C*-subalgebra $\mathcal{A}$ of $\mathrm{Mat}(n,\mathbb{C})$. The analysis is based on the mean value parametrization of $\mathcal{E}$, developed in §6.1. The family $\mathcal{E}$ of states is defined by the real analytic function (23)

$$R_{\mathcal{A}} : \mathcal{A}_{\mathrm{sa}} \to \mathcal{A}_{\mathrm{sa}}, \quad R(\theta) = R_{\mathcal{A}}(\theta) = \exp_{\mathcal{A}}(a)/\operatorname{tr}(\exp_{\mathcal{A}}(a)).$$

We consider a non-empty affine subspace $\Theta \subset \mathcal{A}_{\mathrm{sa}}$ of self-adjoint matrices, its translation vector space

$$U := \mathrm{lin}(\Theta) = \Theta - \Theta$$

and we define the exponential family $\mathcal{E} := R_{\mathcal{A}}(\Theta)$. In §6.2 we define an extension $\mathrm{ext}(\mathcal{E})$ of $\mathcal{E}$. And we prove the bijection $\pi_U|_{\mathrm{ext}(\mathcal{E})} : \mathrm{ext}(\mathcal{E}) \to \mathbb{M}(U)$ to the mean value set $\mathbb{M}(U) = \pi_U(\mathcal{S})$, which is a projection of the state space (29). Then we prove the Complete Pythagorean theorem in §6.3. The Complete projection theorem is proved in §6.4. Application to quantum correlations are described in §6.5. In §6.6 we discuss necessary conditions for commutativity of the algebra $\mathcal{A}$.

6.1. The mean value chart

We settle the mean value chart of an exponential family. Its inverse is the real analytic mean value parametrization. The mean value chart was established for linear spaces $\Theta = U$ in [63]. Examples are shown in Figure 3.4.

Restrictions to affine subspaces of $\mathcal{A}_{\mathrm{sa}}$ are the rule in subsequent arguments, hence we accept relatively open convex subsets of $\mathcal{A}_{\mathrm{sa}}$ (in place of open subset of $\mathbb{R}^d$) as domains of differentiable maps and as ranges of diffeomorphisms and charts. We recall that the relative interior of the state space consists of all invertible states,

$$\operatorname{ri}\mathcal{S} = \{\rho \in \mathcal{S} \mid \rho^{-1} \text{ exists in } \mathcal{A}\}, \tag{61}$$

and is (norm) open in $\mathcal{A}_1 = \{a \in \mathcal{A}_{\mathrm{sa}} \mid \operatorname{tr}(a) = 1\}$, see e.g §2.3 in [60].

Proposition 6.1. *Let $\mathbb{1} \notin U$ for the multiplicative identity $\mathbb{1}$ of $\mathcal{A}$.*

1. *The set $\pi_U(\mathcal{E})$ is open relative to U and $\pi_U \circ R_{\mathcal{A}}|_\Theta : \Theta \to \pi_U(\mathcal{E})$ is a real analytic diffeomorphism.*
2. *If Θ has codimension one in $\mathcal{A}_{\mathrm{sa}}$, then $R_{\mathcal{A}}|_\Theta : \Theta \to \operatorname{ri}\mathcal{S}$ is a real analytic diffeomorphism.*
3. *The bijections $(R_{\mathcal{A}}|_\Theta)^{-1} : \mathcal{E} \to \Theta$ and $\pi_U|_{\mathcal{E}} : \mathcal{E} \to \pi_U(\mathcal{E})$ are global charts for $\mathcal{E}$ and $(\pi_U|_{\mathcal{E}})^{-1} : \pi_U(\mathcal{E}) \to \mathcal{E}$ is real analytic.*

Proof. *Part 1.* The derivative of $\mathcal{A}_{\mathrm{sa}} \to \mathbb{R}$, $a \mapsto \operatorname{tr}\exp_{\mathcal{A}}(a)$ can be computed from (50) using cyclic reordering under the trace. For $a, u \in \mathcal{A}_{\mathrm{sa}}$ we have

$$\frac{\partial}{\partial t}|_{t=0} \operatorname{tr}\exp_{\mathcal{A}}(a + tu) = \langle u, \exp_{\mathcal{A}}(a)\rangle .$$

Hence the free energy (39) has the derivative at $\theta \in \mathcal{A}_{\mathrm{sa}}$ in the direction $u \in \mathcal{A}_{\mathrm{sa}}$

$$\frac{\partial}{\partial t}|_{t=0} F_{\mathcal{A}}(\theta + tu) = \langle u, R_{\mathcal{A}}(\theta)\rangle . \tag{62}$$

From the product rule and (50) we get

$$\frac{\partial}{\partial t}|_{t=0} R_{\mathcal{A}}(\theta + tu) = \int_0^1 R_{\mathcal{A}}(\theta)^{1-y} u R_{\mathcal{A}}(\theta)^y \mathrm{d}y - \langle u, R_{\mathcal{A}}(\theta)\rangle R_{\mathcal{A}}(\theta)\,. \tag{63}$$

For $\theta, u, v \in \mathcal{A}_{\text{sa}}$ we consider the real symmetric bilinear form

$$\langle\!\langle u, v\rangle\!\rangle_\theta := \frac{\partial^2}{\partial s \partial t}|_{s=t=0} F_{\mathcal{A}}(\theta + su + tv)\,. \tag{64}$$

If restricted to $\theta \in \Theta$ and $u, v \in U$ this bilinear form is called BKM-metric, see Remark 6.2. We recall the well-known fact that it defines a Riemannian matric. We obtain from (62) and (63)

$$\langle\!\langle u, v\rangle\!\rangle_\theta = \left\langle u, \frac{\partial}{\partial t}|_{t=0} R_{\mathcal{A}}(\theta + tv)\right\rangle = \int_0^1 \langle \xi(u, y), \xi(v, y)\rangle \mathrm{d}y \tag{65}$$

with the not necessarily self-adjoint matrix

$$\xi(u, y) := R_{\mathcal{A}}(\theta)^{\frac{y}{2}} \left[u - \langle u, R_{\mathcal{A}}(\theta)\rangle \mathbb{1}\right] R_{\mathcal{A}}(\theta)^{\frac{1-y}{2}}\,.$$

We have $\langle\!\langle u, u\rangle\!\rangle_\theta > 0$ unless $u \in \mathcal{A}_{\text{sa}}$ is a (real) scalar multiple of $\mathbb{1}$. Hence $\langle\!\langle \cdot, \cdot\rangle\!\rangle_\theta$ is a non-degenerate bilinear form on U.

Since $R|_\Theta$ is real analytic, the composition $\pi_U \circ R|_\Theta$ with the orthogonal projection to U is also real analytic. If $\{u_i\}_{i=1}^k$ is an orthonormal basis of U then the directional derivative at $\theta \in \Theta$ along $u \in U$ is by (65)

$$\frac{\partial}{\partial t}|_{t=0} \pi_U \circ R(\theta + tu) = \pi_U \left(\frac{\partial}{\partial t}|_{t=0} R(\theta + tu)\right) = \sum_{i=1}^k \langle\!\langle u, u_i\rangle\!\rangle_\theta u_i\,. \tag{66}$$

Since $\langle\!\langle \cdot, \cdot\rangle\!\rangle_\theta$ is non-degenerate on U, the Jacobian of $\pi_U \circ R|_\Theta$ is invertible everywhere. Then the inverse function theorem implies that $\pi_U \circ R|_\Theta$ is locally invertible and its local inverses are real analytic functions, see e.g. §2.5 in [37]. This implies that the image $\pi_U \circ R(\Theta)$ is an open subset relative to U.

The global injectivity of $\pi_U \circ R|_\Theta$ follows from the projection theorem (28): If there are $\theta, \theta' \in \Theta$ such that $\pi_U \circ R(\theta) = \pi_U \circ R(\theta')$, then $R(\theta) = R(\theta')$. Taking the logarithm on both sides one has $\theta - F(\theta)\mathbb{1} = \theta' - F(\theta')\mathbb{1}$ so the difference $\theta - \theta'$ is proportional to $\mathbb{1}$. Hence $\theta = \theta'$ by the assumption $\mathbb{1} \not\in U$.

Part 2. If $\Theta = \mathcal{A}_0$ is the space of traceless matrices, then

$$\log_0 : \operatorname{ri} \mathcal{S} \to \mathcal{A}_0\,, \quad \rho \mapsto \log(\rho) - \frac{\operatorname{tr} \log(\rho)}{\operatorname{tr} \mathbb{1}} \mathbb{1} \tag{67}$$

is inverse to $R|_{\mathcal{A}_0}$ and this shows $R(\Theta) = \operatorname{ri} \mathcal{S}$. Since $R(\theta + \mathbb{1}) = R(\theta)$ for all $\theta \in \mathcal{A}_{\text{sa}}$, we have $R(\Theta) = \operatorname{ri} \mathcal{S}$ for every affine subspace $\Theta \subset \mathcal{A}_{\text{sa}}$ of codimension one and with $\mathbb{1} \not\in \operatorname{lin}(\Theta)$.

Part 3. By virtue of the real analytic diffeomorphism in *1* it is sufficient to prove that $R_{\mathcal{A}}|_\Theta : \Theta \to \mathcal{E}$ is a real analytic bijection. The function $R_{\mathcal{A}}$ is real analytic by definition and $R_{\mathcal{A}}|_\Theta$ is invertible on $\mathcal{E}$ by *2*. $\square$

Remark 6.2 (The BKM-metric). If Θ has codimension one in $\mathcal{A}_{\mathrm{sa}}$ and if $\mathbb{1} \not\in U$ then $\Theta \cong \mathrm{ri}\,\mathcal{S}$. On this manifold the family (65) of scalar products, parametrized by $\theta \in \Theta$ and defined for $u, v \in U$ by

$$\langle\!\langle u, v \rangle\!\rangle_\theta = \left\langle u, \frac{\partial}{\partial t}|_{t=0} R_{\mathcal{A}}(\theta + tv) \right\rangle ,$$

is a Riemannian metric, called *BKM-metric* (an acronym for Bogoliubov, Kubo and Mori, see e.g. [4, 50]).

Indeed, by Proposition 6.1.2 the map $R_{\mathcal{A}}|_\Theta : \Theta \to \mathrm{ri}\,\mathcal{S}$ is a diffeomorphism. For $\theta \in \Theta$ and $u \in U$ let us use the curve $\gamma_{\theta,u} : \mathbb{R} \to \Theta$, $t \mapsto \theta + tu$ to represent a tangent vector at the footpoint θ. The (-1)-representation $u^{(-1)}$ of u is by definition taken in the identity chart $\mathrm{id} : \mathrm{ri}\,\mathcal{S} \to \mathrm{ri}\,\mathcal{S}$, so

$$u^{(-1)} = \frac{\partial}{\partial t}|_{t=0} R_{\mathcal{A}}(\theta + tu) .$$

The $(+1)$-representation $u^{(+1)}$ of u is by definition taken in the logarithmic representation $\log : \mathrm{ri}\,\mathcal{S} \to \mathcal{A}_{\mathrm{sa}}$, so

$$u^{(+1)} = D \log_{\mathcal{A}}(u^{(-1)}) = \frac{\partial}{\partial t}|_{t=0} \log_{\mathcal{A}} \circ R_{\mathcal{A}}(\theta + tu) = u + \lambda \mathbb{1}$$

for some $\lambda \in \mathbb{R}$. Since $v^{(-1)}$ has trace zero, we arrive at the *mixed representation* $\langle\!\langle u, v \rangle\!\rangle_\theta = \mathrm{tr}(u^{(+1)} v^{(-1)})$ of the BKM-metric, see e.g. [27].

Let us calculate the range of the chart $\pi_U|_{\mathcal{E}}$. The following statement gives us an upper bound on the norm closure $\overline{\mathcal{E}}$. It is used implicitly in Lemma 7 in [63].

Proposition 6.3. *Let $(x_i)_{n \in \mathbb{N}} \subset \Theta$ and assume that the states $R_{\mathcal{A}}(x_i)$, $i \in \mathbb{N}$, converge in norm to the state ρ. If $\rho \not\in \mathcal{E}$ then $s(\rho) \preceq p^+_{\mathcal{A}}(u)$ holds for every accumulation point of $(\frac{x_i}{\|x_i\|})_{i \in \mathbb{N}}$.*

Proof. If x_i has a bounded subsequence, then by continuity of $R_{\mathcal{A}}$ we have $\rho = \lim_{i\to\infty} R_{\mathcal{A}}(x_i) \in \mathcal{E}$. So we can assume $\|x_i\| \overset{i\to\infty}{\to} \infty$. By selecting a subsequence let us choose an accumulation point

$$u := \lim_{i\to\infty} \frac{x_i}{\|x_i\|} \in U .$$

To apply Lemma 4.27.1 we need to bound the free energy (39). Let $\lambda^-_{\mathcal{A}}(a)$ resp. $\lambda^+_{\mathcal{A}}(a)$ denote the smallest resp. largest spectral value of a self-adjoint matrix $a \in \mathcal{A}_{\mathrm{sa}}$. Then

$$\log(\mathrm{tr}\,\mathbb{1}) + \lambda^-_{\mathcal{A}}(a) \leq F(a) \leq \log(\mathrm{tr}\,\mathbb{1}) + \lambda^+_{\mathcal{A}}(a)$$

and it implies $|F(a)| \leq \log(\mathrm{tr}\,\mathbb{1}) + \|a\|$ for the spectral norm $\|\cdot\|$.

From the bounded sequence $(\frac{F(x_i)}{\|x_i\|})_{i\in\mathbb{N}}$ we select another subsequence, such that

$$\lambda := \lim_{i\to\infty} \frac{F(x_i)}{\|x_i\|} \in \mathbb{R}$$

converges. Defining $y_i := x_i - F(x_i)\mathbb{1}$ gives $e^{y_i} = R(x_i) \stackrel{i\to\infty}{\to} \rho$ and

$$\frac{y_i}{\|y_i\|} = \left(\frac{x_i}{\|x_i\|} - \frac{F(x_i)}{\|x_i\|}\mathbb{1}\right) / \left\| \frac{x_i}{\|x_i\|} - \frac{F(x_i)}{\|x_i\|}\mathbb{1} \right\| \stackrel{i\to\infty}{\to} \frac{u-\lambda\mathbb{1}}{\|u-\lambda\mathbb{1}\|}\,.$$

Since $(u-\lambda)/\|u-\lambda\|$ has the same maximal projection as u, the claim follows from Lemma 4.27.1. □

The following statement is an idea from Theorem 2 b) in [63].

Lemma 6.4. *Let $f : V \to W$ be a continuous map between two finite-dimensional real vector spaces. Let $K \subset V$ be non-empty and bounded, $L \subset W$ be connected and $f(K) \subset L$. If $f(K)$ is open and $f(\overline{K} \setminus K) \cap L = \emptyset$, then $f(K) = L$.*

Proof. Since $f(\overline{K} \setminus K) \cap L = \emptyset$ we have $L \setminus f(K) = L \setminus f(\overline{K}) = (W \setminus f(\overline{K})) \cap L$ and $f(\overline{K}) \cap L = f(K) \cap L$, hence

$$L = \big(f(\overline{K}) \cap L\big) \cup \big(L \setminus f(\overline{K})\big) = (f(K) \cap L) \cup \big((W \setminus f(\overline{K})) \cap L\big)\,. \tag{68}$$

The set $f(K)$ is open in W by assumption and since $f(\overline{K})$ is compact $W \setminus f(\overline{K})$ is open in W. Since $f(K) \cap L \neq \emptyset$ by assumption, (68) is a disconnection of L unless $L \setminus f(K) = \emptyset$. Since L is connected by assumption, $f(K) \supset L$ follows. □

We have collected all arguments needed to compute $\pi_U(\mathcal{E})$. The mean value set $\mathbb{M}(U)$ plays a crucial role (29).

Theorem 6.5. *Let $\mathbb{1} \notin U$. Then $\operatorname{ri}\mathbb{M}_{\mathcal{A}}(U)$ is open in the norm topology of U and the chart change $\pi_U \circ R_{\mathcal{A}}|_\Theta : \Theta \to \operatorname{ri}\mathbb{M}_{\mathcal{A}}(U)$ is a real analytic diffeomorphism. We have $\pi_U(\overline{\mathcal{E}} \setminus \mathcal{E}) = \operatorname{rb}\mathbb{M}_{\mathcal{A}}(U)$.*

Proof. The map $\pi_U \circ R|_\Theta : \Theta \to \pi_U(\mathcal{E})$ is a real analytic diffeomorphism by Proposition 6.1.1 and $\pi_U(\mathcal{E})$ is open relative to U. We shall first show

$$\pi_U(\overline{\mathcal{E}} \setminus \mathcal{E}) \subset \operatorname{rb}\mathbb{M}(U)\,. \tag{69}$$

Let $\rho \in \overline{\mathcal{E}} \setminus \mathcal{E}$. Proposition 6.3 shows that the support projection of ρ satisfies $s(\rho) \preceq p^+(u)$ for a non-zero $u \in U$ and Proposition 4.5 shows that ρ lies in the exposed face $\mathbb{F}(p^+(u)) = F_\perp(\mathcal{S}, u)$ of the state space. Then Lemma 4.9 shows that $\pi_U(\rho)$ lies in the exposed face $F_\perp(\mathbb{M}(U), u)$ of the mean value set. The mean value set $\mathbb{M}(U)$ has non-empty interior because it contains $\pi_U(\mathcal{E})$ and then Theorem 13.1 in [52] proves that the exposed face $F_\perp(\mathbb{M}(U), u)$ is included in the boundary of $\mathbb{M}(U)$. This proves $\pi_U(\rho) \in \operatorname{rb}\mathbb{M}(U)$.

In order to prove that $\pi_U \circ R|_\Theta : \Theta \to \mathrm{ri}\,\mathbb{M}(U)$ is a real analytic diffeomorphism it suffices to prove $\pi_U(\mathcal{E}) = \mathrm{ri}\,\mathbb{M}(U)$. The convex body $\mathbb{M}(U)$ is the projection of the whole state space, so $\pi_U(\mathcal{E}) \subset \mathbb{M}(U)$. But $\mathcal{E} \subset \mathrm{ri}(\mathcal{S})$ holds by (61) and thanks to the equality $\pi_U \circ \mathrm{ri}(\mathcal{S}) = \mathrm{ri} \circ \pi_U(\mathcal{S})$ (see e.g. Theorem 6.6 in [52]) we have $\pi_U(\mathcal{E}) \subset \mathrm{ri}\,\mathbb{M}(U)$. We meet the conditions of Lemma 6.4 with

$$V = \mathcal{A}_{\mathrm{sa}}, \quad W = U, \quad f = \pi_U, \quad K = \mathcal{E} \quad \text{and} \quad L = \mathrm{ri}\,\mathbb{M}(U)\,.$$

Indeed, $K = \mathcal{E} \subset \mathcal{S}$ is non-empty and bounded, the convex set $L = \mathrm{ri}\,\mathbb{M}(U)$ is connected. We have proved above that $f(K) = \pi_U(\mathcal{E})$ is included in $L = \mathrm{ri}\,\mathbb{M}(U)$ and $f(K) = \pi_U(\mathcal{E})$ is open relative to $W = U$. Moreover $\pi_U(\overline{\mathcal{E}} \setminus \mathcal{E}) \subset \mathrm{rb}\,\mathbb{M}(U)$ in (69) implies

$$f(\overline{K} \setminus K) \cap L = \pi_U(\overline{\mathcal{E}} \setminus \mathcal{E}) \cap \mathrm{ri}\,\mathbb{M}(U) \subset \mathrm{rb}\,\mathbb{M}(U) \cap \mathrm{ri}\,\mathbb{M}(U) = \emptyset\,.$$

Then $\pi_U(\mathcal{E}) = \mathrm{ri}\,\mathbb{M}(U)$ follows from Lemma 6.4.

Finally we show $\pi_U(\overline{\mathcal{E}} \setminus \mathcal{E}) = \mathrm{rb}\,\mathbb{M}(U)$. Since $\overline{\mathcal{E}} \subset \mathcal{S}$ is compact and $\pi_U(\mathcal{E}) = \mathrm{ri}\,\mathbb{M}(U)$ we have

$$\mathbb{M}(U) = \overline{\pi_U(\mathcal{E})} \subset \pi_U(\overline{\mathcal{E}}) \subset \mathbb{M}(U)$$

hence $\pi_U(\overline{\mathcal{E}}) = \mathbb{M}(U)$. Then $\pi_U(\mathcal{E}) = \mathrm{ri}\,\mathbb{M}(U)$ proves $\pi_U(\overline{\mathcal{E}} \setminus \mathcal{E}) \supset \mathrm{rb}\,\mathbb{M}(U)$. The opposite inclusion is (69). □

In the sequel $\mathbb{1} \in U$ will naturally occur in our constructions. Let us drop the condition $\mathbb{1} \notin U$. We use the spaces $\mathcal{A}_0$ and $\mathcal{A}_1$ from Definition 4.4.

Corollary 6.6. *The map $\pi_U|_{\mathcal{E}} : \mathcal{E} \to \mathrm{ri}\,\mathbb{M}(U)$ is a bijection and the inverse $(\pi_U|_{\mathcal{E}})^{-1} : \mathrm{ri}\,\mathbb{M}(U) \to \mathcal{E}$ is real analytic.*

Proof. If $\mathbb{1} \notin U$ then the claim follows from Theorem 6.5 and Proposition 6.1.3. We assume $\mathbb{1} \in U$ and we define $\Theta_0 := \pi_{\mathcal{A}_0}(\Theta)$ and $U_0 := \pi_{\mathcal{A}_0}(U)$. We have $U_0 = \mathrm{lin}(\Theta_0)$ and since ${\mathcal{A}_0}^\perp = \mathbb{1}\mathbb{R}$ holds, we have $U = U_0 + \mathbb{1}\mathbb{R}$ and $\Theta = \Theta_0 + \mathbb{1}\mathbb{R}$. Clearly $\mathcal{E} = R_{\mathcal{A}}(\Theta) = R_{\mathcal{A}}(\Theta_0)$ and $(\pi_{U_0}|_{\mathcal{E}})^{-1} : \mathrm{ri}\,\mathbb{M}(U_0) \to \mathcal{E}$ is a real analytic bijection because $\mathbb{1} \notin U_0$. We have $\pi_U = \pi_{U_0} + \pi_{\mathbb{1}\mathbb{R}}$ and

$$\pi_U|_{\mathcal{A}_1} = \pi_{U_0}|_{\mathcal{A}_1} + \frac{1}{\mathrm{tr}\,\mathbb{1}}\mathbb{1}\,.$$

So $\mathrm{ri}\,\mathbb{M}(U) = \mathrm{ri}\,\mathbb{M}(U_0) + \frac{1}{\mathrm{tr}\,\mathbb{1}}\mathbb{1}$ holds and for $u \in \mathrm{ri}\,\mathbb{M}(U)$ the equality $(\pi_U|_{\mathcal{E}})^{-1}(u) = (\pi_{U_0}|_{\mathcal{E}})^{-1}(u - \frac{1}{\mathrm{tr}\,\mathbb{1}}\mathbb{1})$ completes the proof. □

Definition 6.7. We call the continuous bijection $\pi_U|_{\mathcal{E}} : \mathcal{E} \to \mathrm{ri}\,\mathbb{M}_{\mathcal{A}}(U)$ in Corollary 6.6 the *mean value chart* of $\mathcal{E}$. The real analytic inverse

$$(\pi_U|_{\mathcal{E}})^{-1} : \ \mathrm{ri}\ \mathbb{M}_{\mathcal{A}}(U) \ \to \ \mathcal{E} \tag{70}$$

is the *mean value parametrization* of $\mathcal{E}$.

6.2. The extension of an exponential family

We define an extension $\mathrm{ext}(\mathcal{E})$ of $\mathcal{E}$ composed of exponential families in compressed algebras $p\mathcal{A}p$ for orthogonal projections p, one for each face of the mean value set $\mathbb{M}(U)$ (of the vector space U). We obtain a bijective mean value parametrization

$$(\pi_U|_{\mathrm{ext}(\mathcal{E})}) : \mathbb{M}(U) \to \mathrm{ext}(\mathcal{E}) .$$

We obtain a projection with linear fibers

$$\pi_{\mathcal{E}} : \mathcal{S}_{\mathcal{A}} \to \mathrm{ext}(\mathcal{E})$$

from the state space $\mathcal{S}_{\mathcal{A}}$ to the extension. We finish by providing analogues of the *natural* and *canonical* parameters known in statistics.

We recall from (48) for $p \in \mathcal{P}$ and $a \in \mathcal{A}_{\mathrm{sa}}$ the projection to the compressed algebra $p\mathcal{A}p$

$$c^p : \mathcal{A}_{\mathrm{sa}} \to (p\mathcal{A}p)_{\mathrm{sa}} , \quad a \mapsto pap .$$

Only the projections in the lattice $\mathcal{P}^U$ are interesting. Through (46) they correspond to the faces the mean value set $\mathbb{M}(U)$ and they can in principle be computed by spectral analysis, see Remark 4.17. See also §3.2 about the construction of $\mathcal{P}^U$.

Let us investigate the extension defined in (35).

Definition 6.8. For orthogonal projections $p \in \mathcal{P}$ we consider the exponential family

$$\mathcal{E}_p := R_{p\mathcal{A}p}(c^p(\Theta)) .$$

The *extension* of $\mathcal{E}$ is defined in terms of the projection lattice $\mathcal{P}^U$

$$\mathrm{ext}(\mathcal{E}) := \bigcup_{p \in \mathcal{P}^U \setminus \{0\}} \mathcal{E}_p .$$

We begin by parametrizing the extension $\mathrm{ext}(\mathcal{E})$ of the exponential family $\mathcal{E} = R_{\mathcal{A}}(\Theta)$ by mean values. Let $\theta_0, u_1, \ldots, u_k \in \mathcal{A}_{\mathrm{sa}}$ such that $\Theta := \theta_0 + \mathrm{span}_{\mathbb{R}}(u_1, \ldots, u_k)$ and denote $\mathbf{u} = (u_1, \ldots, u_k)$. Then in the notation of (24) the translation vector space of Θ is $U = \mathrm{span}_{\mathbb{R}}(u_1, \ldots, u_k)$. The following parametrization by vectors in $\mathbb{M}(U)$ has a formulation in terms of the mean value map $m_{\mathbf{u}}$ and the convex support $\mathrm{cs}(\mathbf{u})$, defined respectively in (30) and (31).

Lemma 6.9. *The projection $\pi_U|_{\mathrm{ext}(\mathcal{E})} : \mathrm{ext}(\mathcal{E}) \to \mathbb{M}(U)$ is a bijection. The mean value map $m_{\mathbf{u}}|_{\mathrm{ext}(\mathcal{E})} : \mathrm{ext}(\mathcal{E}) \to \mathrm{cs}(\mathbf{u})$ is a bijection.*

Proof. For every non-zero projection $p \in \mathcal{P}$, the mean value chart in Corollary 6.6 proves the bijection

$$\pi_{c^p(U)}|_{\mathcal{E}_p} : \mathcal{E}_p \to \mathrm{ri}\, \mathbb{M}_{p\mathcal{A}p}(c^p(U)) .$$

Using the third diagram in Lemma 4.14 we have the bijection

$$\pi_U|_{\mathcal{E}_p} : \mathcal{E}_p \to \operatorname{ri} \pi_U(\mathbb{F}_{\mathcal{A}}(p)) . \tag{71}$$

The map $\mathcal{P}^U \to \mathcal{F}(\mathbb{M}(U))$, $p \mapsto \pi_U(\mathbb{F}_{\mathcal{A}}(p))$ is a lattice isomorphism from the projection lattice $\mathcal{P}^U$ to the face lattice of $\mathbb{M}(U)$, see (46). Then the stratification (44) of $\mathbb{M}(U)$ into the relative interiors of its faces shows that the bijections (71) assemble to a bijection $\operatorname{ext}(\mathcal{E}) \to \mathbb{M}(U)$. The second claim follows from (32). □

We renew and extend the definition of $\pi_{\mathcal{E}}$ from (26).

Definition 6.10. The projection to the extension $\operatorname{ext}(\mathcal{E})$ is well-defined by Lemma 6.9 as the map

$$\pi_{\mathcal{E}} : \mathcal{S}_{\mathcal{A}} \to \operatorname{ext}(\mathcal{E}), \quad \rho \mapsto (\pi_U|_{\operatorname{ext}(\mathcal{E})})^{-1} \circ \pi_U(\rho) . \tag{72}$$

Coordinates can be put on the canonical parametrization as well. They depend on the projection lattice $\mathcal{P}^U$. The free energy F, defined in (39), relates them to the mean value coordinates.

Corollary 6.11. *Let $x \in \operatorname{cs}(\mathbf{u})$. Then there are a unique projection $p \in \mathcal{P}^U \setminus \{0\}$ and some (in general non-unique) $\lambda_1, \dots, \lambda_k \in \mathbb{R}$, such that*

$$\left\{ \frac{\partial}{\partial \lambda_j} F_{p\mathcal{A}p}(p(\theta_0 + \sum_{i=1}^k \lambda_i u_i)p) \right\}_{j=1}^k = x . \tag{73}$$

For every solution $(p, \lambda_1, \dots, \lambda_k)$ of (73) the state $\rho := R_{p\mathcal{A}p}(p(\theta_0 + \sum_{i=1}^k \lambda_i u_i)p)$ is the unique state in $\operatorname{ext}(\mathcal{E})$ such that $m_{\mathbf{u}}(\rho) = x$.

Proof. If $\rho \in \operatorname{ext}(\mathcal{E})$ then by definition of $\operatorname{ext}(\mathcal{E})$ there exist $p \in \mathcal{P}^U$ and $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ such that for $\theta := \theta_0 + \sum_{i=1}^k \lambda_i u_i$ we have $\rho = R_{p\mathcal{A}p}(c^p(\theta))$. Since $s(\rho) = p$, the projection p is unique. The derivative (62) of the free energy is

$$\frac{\partial}{\partial \lambda_j} F_{p\mathcal{A}p}(c^p(\theta)) = \langle c^p(u_j), R_{p\mathcal{A}p}(c^p(\theta)) \rangle = \langle u_j, R_{p\mathcal{A}p}(c^p(\theta)) \rangle, \quad j = 1, \dots, k .$$

The collection of these k equations solves the existence part of the claim:

$$\left\{ \frac{\partial}{\partial \lambda_j} F_{p\mathcal{A}p}(c^p(\theta)) \right\}_{j=1}^k = m_{\mathbf{u}}(R_{p\mathcal{A}p}(c^p(\theta))) = x . \tag{74}$$

Conversely, if $x \in \operatorname{cs}(\mathbf{u})$, $p \in \mathcal{P}^U \setminus \{0\}$ and $(\lambda_1, \dots, \lambda_k) \in \mathbb{R}^k$ solve (73), then the state $\rho := R_{p\mathcal{A}p}(p(\theta_0 + \sum_{i=1}^k \lambda_i u_i)p)$ has, by (74), the mean values $m_{\mathbf{u}}(\rho) = x$. By definition of the extension $\operatorname{ext}(\mathcal{E})$ the state ρ belongs to $\operatorname{ext}(\mathcal{E})$ and Lemma 6.9 shows that ρ is unique in $\operatorname{ext}(\mathcal{E})$ with the mean value $m_{\mathbf{u}}(\rho) = x$. □

The parameters $(\lambda_1, \ldots, \lambda_k) \in \mathbb{R}^k$ resp. $m_{\mathbf{u}}(\rho) \in \mathbb{R}^k$ are analogues of *canonical parameters* resp. *natural parameters* of $\mathrm{ext}(\mathcal{E})$ in statistics, see §20 in [16]. In the restriction to $p = \mathbb{1}$, they put coordinates on the canonical parametrization (24) resp. on the mean value parametrization (70). Notice that the non-uniqueness of canonical parameters $\lambda_1, \ldots, \lambda_k$ can not be resolved by choosing linearly independent observables $\mathbb{1}, u_1, \ldots, u_k$. Even if $(\lambda_1, \ldots, \lambda_k) \mapsto R_{p\mathcal{A}p}(p(\theta_0 + \sum_{i=1}^k \lambda_i u_i)p)$ is a bijection for $p = \mathbb{1}$, this will no longer be true for smaller projections $p \neq \mathbb{1}$ where linear independence of $p, pu_1p, \ldots, pu_kp$ can be lost.

6.3. The Complete Pythagorean theorem

We prove the Complete Pythagorean theorem for an exponential family $\mathcal{E}$. It is applied to the maximization of the von Neumann entropy in §3.4.

We use the projection $\pi_{\mathcal{E}} : \mathcal{S}_{\mathcal{A}} \to \mathrm{ext}(\mathcal{E})$, defined in (72), with linear fibers parallel to $U^{\perp}$.

Theorem 6.12 (Complete Pythagorean theorem). *If $\rho \in \mathcal{S}_{\mathcal{A}}$ and if σ lies in the extension $\mathrm{ext}(\mathcal{E})$, then $S(\rho, \pi_{\mathcal{E}}(\rho)) + S(\pi_{\mathcal{E}}(\rho), \sigma) = S(\rho, \sigma)$ holds.*

Proof. By Definition 6.8 of the extension $\mathrm{ext}(\mathcal{E})$ there exist projections $p, q \in \mathcal{P}^U$, such that $\pi_{\mathcal{E}}(\rho) \in \mathcal{E}_p$ and $\sigma \in \mathcal{E}_q$. If $p \not\preceq q$ then $s(\rho) \not\preceq q$ follows from Lemma 4.18 since $q \in \mathcal{P}^U$. We get

$$S(\pi_{\mathcal{E}}(\rho), \sigma) = S(\rho, \sigma) = \infty$$

and the non-negativity of the relative entropy proves the claim.

Let $p \preceq q$. Then by Lemma 4.18 we have $s(\rho) \preceq p = s(\pi_{\mathcal{E}}(\rho)) \preceq q = s(\sigma)$ and only finite relative entropies appear in the claimed equation. We subtract the trivial equation

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

and continue to show that the resulting difference (see Remark 4.23.3 for notation of functional calculus)

$$\begin{aligned} x &:= S(\rho, \pi_{\mathcal{E}}(\rho)) - S(\rho, \sigma) - [S(\pi_{\mathcal{E}}(\rho), \pi_{\mathcal{E}}(\rho)) - S(\pi_{\mathcal{E}}(\rho), \sigma)] \\ &= \mathrm{tr}[\rho - \pi_{\mathcal{E}}(\rho)][\log^{[q]} \sigma - \log^{[p]} \pi_{\mathcal{E}}(\rho)] \end{aligned}$$

is zero. By definition of $\mathcal{E}_p$ resp. $\mathcal{E}_q$ there exist $\theta, \widetilde{\theta} \in \Theta$ and $y, \widetilde{y} \in \mathbb{R}$, such that $\log^{[p]}(\pi_{\mathcal{E}}(\rho)) = c^p(\theta) + yp$ resp. $\log^{[q]}(\sigma) = c^q(\widetilde{\theta}) + \widetilde{y}q$. Since $s(\rho) \preceq s(\pi_{\mathcal{E}}(\rho)) = p \preceq q$ this gives

$$x = \mathrm{tr}[\rho - \pi_{\mathcal{E}}(\rho)][q\widetilde{\theta}q - p\theta p] = \mathrm{tr}[\rho - \pi_{\mathcal{E}}(\rho)][\widetilde{\theta} - \theta]\,.$$

Since the translation vector space U of Θ is perpendicular to the fibers of the projection $\pi_{\mathcal{E}}$, the difference $\widetilde{\theta} - \theta$ is perpendicular to $\rho - \pi_{\mathcal{E}}(\rho)$, hence $x = 0$. $\square$

6.4. The Complete projection theorem

We prove the Complete projection theorem. In §3.5 we show the corollary that the extension $\mathrm{ext}(\mathcal{E})$ in Definition 6.8 is the rI-closure (20) and we consider the example of the Staffelberg family. Application to quantum correlations are described in §6.5.

We are going to use the full scope of convex geometry introduced in §4.1. A discussion why poonems are essential is given in §3.6. Let us start by citing Lemma 9 and 13 in [61]. The first lemma follows also from Lemma 4.27.2 in this article. The free energy (39) is denoted by F.

Lemma 6.13. *Suppose $\theta, u \in \mathcal{A}_{\mathrm{sa}}$ and $p := p^+_{\mathcal{A}}(u)$ is the maximal projection of u. We have*

$$\lim_{t\to\infty} R_{\mathcal{A}}(\theta + t\,u) = R_{p\mathcal{A}p}(c^p(\theta))\,. \tag{75}$$

and

$$\lim_{t\to\infty} \big(F_{\mathcal{A}}(\theta + t\,u) - t\,\lambda^+_{\mathcal{A}}(u)\big) = F_{p\mathcal{A}p}(c^p(\theta))\,. \tag{76}$$

For technical reasons we denote the relative entropy with the first argument $\rho \in \mathcal{S}$ fixed by $S_\rho : \mathcal{S} \to [0,\infty]$, $S_\rho(\sigma) := S(\rho,\sigma)$.

Lemma 6.14. *Suppose $\theta, u \in \mathcal{A}_{\mathrm{sa}}$ and u is not proportional to the identity $\mathbb{1}$ in $\mathcal{A}$. If the state ρ belongs to the exposed face $F_\perp(\mathcal{S}_{\mathcal{A}}, u)$ of the state space, then $S_\rho(R_{\mathcal{A}}(\theta + t\,u))$ is strictly monotone decreasing in $t \in \mathbb{R}$ and*

$$\inf_{t\in\mathbb{R}} S_\rho\big(R_{\mathcal{A}}(\theta + t\,u)\big) = \lim_{t\to\infty} S_\rho\big(R_{\mathcal{A}}(\theta + t\,u)\big) = S_\rho\big(\lim_{t\to\infty} R_{\mathcal{A}}(\theta + t\,u)\big)\,.$$

The following approximation along e-geodesics is described in §3.6 with the example of the Swallow family. We use the entropy distance $\mathrm{d}_{\mathcal{E}}$ defined in (27).

Proposition 6.15. *Let $\rho \in \mathcal{S}$ and $\sigma \in \mathrm{ext}(\mathcal{E})$, such that $\sigma \neq \pi_{\mathcal{E}}(\rho)$ and $S_\rho(\sigma) < \infty$. There exists a norm continuous curve $\gamma : [0,1] \to \mathrm{ext}(\mathcal{E})$ from $\gamma(0) = \sigma$ to $\gamma(1) = \pi_{\mathcal{E}}(\rho)$, such that*

1. *$S_\rho(\gamma(t))$ is strictly monotone decreasing in t and*
2. *$\mathrm{d}_{\mathcal{E}}(\rho) \leq S_\rho(\gamma(1))$.*

Proof. We construct γ by concatenation of several e-geodesics. Since $S_\rho(\sigma) < \infty$ we have $s(\rho) \preceq s(\sigma)$. By Lemma 4.18 there exists a unique projection $p \in \mathcal{P}^U$, such that $\rho \in \mathrm{ri}\,\mathbb{F}_{\mathcal{A}}(p) + U^\perp$ and then it follows from Lemma 6.9 that $\pi_{\mathcal{E}}(\rho) \in \mathcal{E}_p$. We denote by $q \in \mathcal{P}^U$ the projection such that $\sigma \in \mathcal{E}_q$, i.e. $s(\sigma) = q$. By Lemma 4.18 we have $p \preceq s(\sigma) = q$. By Corollary 4.16 there exists an access sequence of projections for U including both p and q, say

$$\mathbb{1} = p_0 \succ p_1 \succ \cdots \succ p_m = p\,,$$

where $p = p_m$ for $m \geq 0$ and $q = p_l$ for $l \geq 0$ and $l \leq m$.

We define a number of $(m-l)$ e-geodesic rays in $\mathcal{E}_{p_l}, \mathcal{E}_{p_{l+1}}, \dots, \mathcal{E}_{p_{m-1}}$. From the Definition 4.12.3 of an access sequence and by Corollary 4.11, for each $k = l, \dots, m-1$ there exists $u_k \in c^{p_k}(U)$, such that

$$p_{k+1} = p^+_{p_k \mathcal{A} p_k}(u_k)\,. \tag{77}$$

Moreover, u_k is not a multiple of the identity p_k in $p_k \mathcal{A} p_k$ (because $p_{k+1} \neq p_k$). Let $\theta \in \Theta$ such that $\sigma = R_{q\mathcal{A}q}(c^q(\theta))$. We define for $k = l, \dots, m-1$ the e-geodesic

$$g_k : \mathbb{R} \to \mathcal{E}_{p_k}\,, \quad t \mapsto R_{p_k \mathcal{A} p_k}(c^{p_k}(\theta) + t u_k)\,. \tag{78}$$

Using (75) we define

$$\sigma_{k+1} := \lim_{t \to \infty} g_k(t) = R_{p_{k+1} \mathcal{A} p_{k+1}}(c^{p_{k+1}}(\theta)) \in \mathcal{E}_{p_{k+1}}\,. \tag{79}$$

After reparametrization $t = \frac{s}{1-s}$, each e-geodesic ray $g_l|_{[0,\infty)}, g_{l+1}|_{[0,\infty)}, \dots, g_{m-1}|_{[0,\infty)}$ is defined on the segment $[0,1)$.

We concatenate the reparametrized e-geodesic rays to a continuous curve $\widetilde{\gamma} : [0, m-l] \to \operatorname{ext}(\mathcal{E})$. The pieces fit together by (79). If $\sigma_m \neq \pi_{\mathcal{E}}(\rho)$ then we add the e-geodesic segment in $\mathcal{E}_p$ from σ_m to $\pi_{\mathcal{E}}(\rho)$. This is parametrized under $R_{p\mathcal{A}p}$ by a straight line segment in $c^p(\Theta)$, which we parametrize linearly by the unit interval $[0,1]$. Since $\sigma \neq \pi_{\mathcal{E}}(\rho)$, one of the inequalities $m - l > 0$ or $\sigma_m \neq \pi_{\mathcal{E}}(\rho)$ must be true so we obtain a curve $\gamma : [0,1] \to \operatorname{ext}(\mathcal{E})$ from $\widetilde{\gamma}$ by linear reparametrization with the strictly positive factor $m - l$ or $m - l + 1$.

We argue that S_ρ is strictly monotone decreasing along γ. Let us begin with the rays in the canonical parametrization (78) for $k = l, \dots, m-1$. Since $p \preceq p_{k+1} \preceq p_k$ we have by Proposition 4.5 and (77)

$$\rho \in \mathbb{F}_{\mathcal{A}}(p) \subset \mathbb{F}_{\mathcal{A}}(p_{k+1}) = \mathbb{F}_{p_k \mathcal{A} p_k}(p_{k+1}) = F_\perp(\mathcal{S}_{p_k \mathcal{A} p_k}, u_k)\,. \tag{80}$$

Since u_k is not a multiple of the identity p_k, Lemma 6.14 can be invoked and it shows that S_ρ is strictly monotone decreasing along g_k.

The fact that S_ρ is strictly monotone decreasing along the e-geodesic segment from σ_m to $\pi_{\mathcal{E}}(\rho)$ uses the strict convexity of S_ρ on $\mathcal{E}_p$ in the canonical parametrization $c^p(\Theta) \to \mathcal{E}_p$ under $R_{p\mathcal{A}p}$. Let us recall the details. In order to have an injective parametrization we project $c^p(\Theta)$ onto $(p\mathcal{A}p)_0 = \{a \in (p\mathcal{A}p)_{\mathrm{sa}} \mid \operatorname{tr}(a) = 0\}$ so that $\Theta_0 := \pi_{(p\mathcal{A}p)_0}(c^p(\Theta))$ satisfies $\mathcal{E}_p = R_{p\mathcal{A}p}(c^p(\Theta)) = R_{p\mathcal{A}p}(\Theta_0)$. For all $\theta_0 \in \Theta_0$ an elementary calculation shows

$$S_\rho(R_{p\mathcal{A}p}(\theta_0)) = -S(\rho) - \operatorname{tr}(\rho\theta_0) + F_{p\mathcal{A}p}(\theta_0)$$

where $S(\rho)$ is the von Neumann entropy and F is the free energy (39). Hence for $u_0, v_0 \in \pi_{(p\mathcal{A}p)_0}(c^p(U)) = \operatorname{lin}(\Theta_0) = \Theta_0 - \Theta_0$ the second derivative

$$\begin{aligned}
&\frac{\partial^2}{\partial s \partial t}|_{s=t=0} S_\rho(R_{p\mathcal{A}p}(\theta_0 + s u_0 + t v_0)) \\
&= \frac{\partial^2}{\partial s \partial t}|_{s=t=0} F_{p\mathcal{A}p}(\theta_0 + s u_0 + t v_0)) = \langle\!\langle u_0, v_0 \rangle\!\rangle_{\theta_0}
\end{aligned}$$

equals the BKM-metric (64). As discussed in the paragraph following (65) this is a Riemannian metric, so $S_\rho(R_{p\mathcal{A}p}(\theta_0))$ has a positive definite Hessian throughout Θ_0 and is strictly convex there. Since $\pi_\mathcal{E}(\rho) \in \mathcal{E}_p$, the function S_ρ has on $\mathcal{E}_p \cong \Theta_0$ a global minimum at $\pi_\mathcal{E}(\rho)$, this follows from the projection theorem (26). Hence S_ρ is strictly monotone decreasing along the e-geodesic from $\sigma_m \in \mathcal{E}_p$ to $\pi_\mathcal{E}(\rho) \in \mathcal{E}_p$.

Second, we prove $\mathrm{d}_\mathcal{E}(\rho) \leq S_\rho(\pi_\mathcal{E}(\rho))$ by showing for $k = 0, \dots, m-1$ that $\mathrm{d}_{\mathcal{E}_{p_k}}(\rho) \leq \mathrm{d}_{\mathcal{E}_{p_{k+1}}}(\rho)$ holds. Then

$$\mathrm{d}_\mathcal{E}(\rho) = \mathrm{d}_{\mathcal{E}_{p_0}}(\rho) \leq \mathrm{d}_{\mathcal{E}_{p_1}}(\rho) \leq \cdots \leq \mathrm{d}_{\mathcal{E}_{p_m}}(\rho) = \mathrm{d}_{\mathcal{E}_p}(\rho) \leq S_\rho(\pi_\mathcal{E}(\rho))$$

will follow, the last inequality since $\pi_\mathcal{E}(\rho) \in \mathcal{E}_p$. Let $\tau \in \mathcal{E}_{p_{k+1}}$ and let $\theta \in \Theta$ such that $\tau = R_{p_{k+1}\mathcal{A}p_{k+1}}(c^{p_{k+1}}(\theta))$. The construction from (77) to (80) can be extended to $k = 0, \dots, m-1$. Now Lemma 6.14 shows

$$\mathrm{d}_{\mathcal{E}_{p_k}}(\rho) \leq \inf_{t\in\mathbb{R}} S_\rho(R_{p_k\mathcal{A}p_k}(c^{p_k}(\theta) + tu_k)) = S_\rho(R_{p_{k+1}\mathcal{A}p_{k+1}}(c^{p_{k+1}}(\theta))) = S_\rho(\tau)\,.$$

Taking the infimum over all $\tau \in \mathcal{E}_{p_{k+1}}$ the claim follows. □

Local minimizers in the following theorem are understood in the norm topology: If $(X, \mathcal{T})$ is a topological space, $Y \subset X$ and $f : Y \to \mathbb{R}$, then $x_0 \in Y$ is a *local minimizer* (*maximizer*) of f on Y, if there is a $\mathcal{T}$ open subset $V \subset X$ including x_0, such that for all $x \in V \cap Y$ we have $f(x_0) \leq f(x)$ ($f(x_0) \geq f(x)$). We now consider the entropy distance $\mathrm{d}_\mathcal{E}$, the rI-closure $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$ and the extension $\mathrm{ext}(\mathcal{E})$ defined respectively in (27), (20) and (35).

Theorem 6.16 (Complete projection theorem). *For each $\rho \in \mathcal{S}$ the relative entropy S_ρ has a unique local minimizer on the extension $\mathrm{ext}(\mathcal{E})$ at $\pi_\mathcal{E}(\rho)$. The entropy distance is $\mathrm{d}_\mathcal{E}(\rho) = \min_{\sigma\in\mathrm{ext}(\mathcal{E})} S_\rho(\sigma) = S_\rho(\pi_\mathcal{E}(\rho))$.*

Proof. For each $\rho \in \mathcal{S}$ we observe from Proposition 6.15.1 and from the fact that S_ρ is finite on $\mathcal{E}$, that $\pi_\mathcal{E}(\rho)$ is the unique global minimizer of S_ρ on $\mathrm{ext}(\mathcal{E})$. We denote its value by $\mathrm{d}_{\mathrm{ext}(\mathcal{E})}(\rho)$. With Proposition 6.15.2 we have for all $\sigma \in \mathrm{ext}(\mathcal{E})$

$$\mathrm{d}_\mathcal{E}(\rho) \leq S_\rho(\pi_\mathcal{E}(\rho)) \leq S_\rho(\sigma)\,.$$

Taking the infimum over $\sigma \in \mathrm{ext}(\mathcal{E})$ this shows $\mathrm{d}_\mathcal{E}(\rho) \leq \mathrm{d}_{\mathrm{ext}(\mathcal{E})}(\rho)$. Since the converse inequality is trivial, we have proved

$$\mathrm{d}_\mathcal{E}(\rho) = \mathrm{d}_{\mathrm{ext}(\mathcal{E})}(\rho)\,.$$

Now $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) = \mathrm{cl}^{\mathrm{rI}}(\mathrm{ext}(\mathcal{E}))$ follows from the definition of the rI-closure. We show $\mathrm{cl}^{\mathrm{rI}}(\mathrm{ext}(\mathcal{E})) = \mathrm{ext}(\mathcal{E})$ to conclude $\mathrm{ext}(\mathcal{E}) = \mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$. The inclusion "$\supset$" is trivial and the converse "$\subset$" follows from the distance-like properties of the relative entropy (11) and since the minimum $\mathrm{d}_{\mathrm{ext}(\mathcal{E})}(\rho)$ is attained on $\mathrm{ext}(\mathcal{E})$ for each state $\rho \in \mathcal{S}$.

It remains to discuss local minimizers σ of S_ρ on $\mathrm{ext}(\mathcal{E})$. If $S_\rho(\sigma) < \infty$, then Proposition 6.15.1 shows that σ is not a local minimizer unless $\sigma = \pi_\mathcal{E}(\rho)$. If $S_\rho(\sigma) = \infty$ we observe that $\mathcal{E}$ is norm dense in $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$ by the Pinsker-Csizár inequality (15). Since S_ρ has finite values on $\mathcal{E}$, the state σ is not a local minimizer. □

6.5. Maximizers of the entropy distance

The *mutual information* is a measure of the total correlation in a bipartite quantum system [43, 46]. To see how it is related to an exponential family we consider two identical quantum systems, described by the algebra $\mathcal{A} := \mathrm{Mat}(n, \mathbb{C})$. The algebra of the joint system is the tensor product $\mathcal{A} \otimes \mathcal{A}$. If $f : \mathcal{A} \otimes \mathcal{A} \to \mathbb{C}$ is its state, then the state of the subsystems are $f_1(a) := f(a \otimes \mathbb{1}_n)$ and $f_2(a) := f(\mathbb{1}_n \otimes a)$ for $a \in \mathcal{A}$. The density matrices ρ, ρ_1 and ρ_2 associated respectively to f, f_1 and f_2 can be used to define the mutual information

$$I(\rho) := S(\rho_1) + S(\rho_2) - S(\rho)\,.$$

The mutual information is a continuous function on $\mathcal{S}_{\mathcal{A} \otimes \mathcal{A}}$. Using the vector space $L := \{a \otimes \mathbb{1}_n + \mathbb{1}_n \otimes b \mid a, b \in \mathcal{A}_{\mathrm{sa}}\}$ of *local observables*, the exponential family

$$\mathcal{F} := R(L) = \{\rho_1 \otimes \rho_2 \mid \rho_1, \rho_2 \in \mathcal{S}(\mathcal{A}) \text{ invertible } \}$$

contains all invertible product states $\rho_1 \otimes \rho_2$, having no correlation. The mean value chart (40) and the projection theorem (28) show for invertible $\rho \in \mathcal{S}_{\mathcal{A} \otimes \mathcal{A}}$ that mutual information is the entropy distance (27)

$$I(\rho) = \mathrm{d}_{\mathcal{F}}(\rho) = \inf\{S(\rho, \sigma) \mid \sigma \in \mathcal{F}\}\,.$$

Maximization of correlation measures in terms of the entropy distance from an exponential family is proposed in [5] as a structuring principle in natural systems, see also [6, 61] and the references therein. We prove two necessary conditions for a local maximizer of the entropy distance $\mathrm{d}_{\mathcal{E}}$ from an exponential family $\mathcal{E}$ in a C*-subalgebra $\mathcal{A}$ of $\mathrm{Mat}(n, \mathbb{C})$. See the beginning of §6 for notation.

The first condition, an upper bound on the rank, enforces a certain degree of determinism on local maximizers. We shall use the fact that a unique face of $\mathcal{S}_{\mathcal{A}}$ exists, which contains a given state $\rho \in \mathcal{S}_{\mathcal{A}}$ in its relative interior (44).

Proposition 6.17. *Let $\rho \in \mathcal{S}_{\mathcal{A}}$ be a local maximizer of the entropy distance $\mathrm{d}_{\mathcal{E}}$ from $\mathcal{E}$ and assume that F is the face of the state space $\mathcal{S}_{\mathcal{A}}$ which contains ρ in its relative interior. Then* $\dim(F) \leq \dim(\mathcal{E})$.

Proof. We consider the convex body $K := F \cap (\rho + U^{\perp})$. If two convex sets $X, Y \subset \mathbb{E}^n$ in the finite-dimensional Euclidean vector space $(\mathbb{E}, \langle\cdot,\cdot\rangle)$ share a relative interior point, then $\mathrm{ri}(X \cap Y) = \mathrm{ri}(X) \cap \mathrm{ri}(Y)$ follows, see e.g. Theorem 6.5 in [52]. Hence $\rho \in \mathrm{ri}(K)$ follows.

If $\sigma \in K$, then by definition (72) of the projection $\pi_{\mathcal{E}}$ we have $\pi_{\mathcal{E}}(\sigma) = \pi_{\mathcal{E}}(\rho)$ and Theorem 6.16 allows to rewrite the entropy distance

$$\mathrm{d}_{\mathcal{E}}(\sigma) = S(\sigma, \pi_{\mathcal{E}}(\sigma)) = S(\sigma, \pi_{\mathcal{E}}(\rho))\,.$$

We have $p := s(\pi_{\mathcal{E}}(\rho)) \succeq s(\sigma)$ by Lemma 4.18 and, using functional calculus in the algebra $p\mathcal{A}p$ (Definition 4.22.3), we get

$$\mathrm{d}_{\mathcal{E}}(\sigma) = -S(\sigma) - \mathrm{tr}\,\sigma \log^{[p]}(\pi_{\mathcal{E}}(\rho))\,.$$

The von Neumann entropy $S(\sigma)$ is strictly concave, see e.g. §II.B in [57]. Hence $\mathrm{d}_{\mathcal{E}}(\sigma)$ is a sum of a strictly convex function and a linear function, it is strictly convex on K. Since ρ is a local maximizer of $\mathrm{d}_{\mathcal{E}}$ on $\mathcal{S}_{\mathcal{A}}$ it is a local maximizer on K. Since $\rho \in \mathrm{ri}(K)$ holds and $\mathrm{d}_{\mathcal{E}}$ is strictly convex on K, we get $K = \{\rho\}$. Then

$$\dim(F) + \dim(U^{\perp}) \leq \dim(\mathcal{A}_{\mathrm{sa}}) = \dim(U) + \dim(U^{\perp})$$

follows, hence $\dim(F) \leq \dim(U)$. If we choose a parametrization of $\mathcal{E}$, such that $\mathbb{1} \notin U$ (e.g. by replacing Θ by $\pi_{\mathcal{A}_0}(\Theta)$) then Proposition 6.1.1 and 6.1.3 show $\dim(U) = \dim(\mathcal{E})$, completing the proof. □

Let us compare the physically relevant cases of $\mathcal{A} = \mathbb{C}^n$ and $\mathcal{A} = \mathrm{Mat}(n, \mathbb{C})$.

Remark 6.18 (Rank estimates). In a C*-subalgebra $\mathcal{A}$ of $\mathrm{Mat}(n, \mathbb{C})$ let $\rho \in \mathcal{S}_{\mathcal{A}}$ and let F be the face of $\mathcal{S}_{\mathcal{A}}$ containing ρ in its relative interior. Let $p := s(\rho)$ be the support projection of ρ and let $\mathrm{rk}(\rho)$ be the rank of ρ. Then Proposition 4.5 shows $F = \mathcal{S}_{p\mathcal{A}p}$ and $\dim(\mathcal{S}_{p\mathcal{A}p}) = \dim((p\mathcal{A}p)_{\mathrm{sa}}) - 1$ hence

$$\dim(F) = \dim((p\mathcal{A}p)_{\mathrm{sa}}) - 1 = \dim_{\mathbb{C}}(p\mathcal{A}p) - 1\,.$$

If ρ is a local maximizer of the entropy distance $\mathrm{d}_{\mathcal{E}}$ from an exponential family $\mathcal{E}$ then Proposition 6.17 shows

$$\dim_{\mathbb{C}}(p\mathcal{A}p) = \dim(F) + 1 \leq \dim(\mathcal{E}) + 1\,. \tag{81}$$

If $\mathcal{A} \cong \mathbb{C}^n$ is the algebra of diagonal matrices in $\mathrm{Mat}(n, \mathbb{C})$ we have $\dim_{\mathbb{C}}(p\mathcal{A}p) = \mathrm{rk}(\rho)$ hence (81) shows

$$\mathrm{rk}(\rho) \leq \dim(\mathcal{E}) + 1\,. \tag{82}$$

If $\mathcal{A} = \mathrm{Mat}(n, \mathbb{C})$ is the full matrix algebra of size n, then $p\mathcal{A}p$ is unitarily equivalent to the algebra of block diagonal matrices $\mathrm{Mat}(\mathrm{rk}(\rho), \mathbb{C}) \oplus 0_{n-\mathrm{rk}(\rho)}$ and $p\mathcal{A}p$ has dimension $\dim_{\mathbb{C}}(p\mathcal{A}p) = \mathrm{rk}(\rho)^2$. Then (81) proves

$$\mathrm{rk}(\rho) \leq \sqrt{\dim(\mathcal{E}) + 1}\,. \tag{83}$$

The second condition identifies local maximizers as the cutoff of their projection $\pi_{\mathcal{E}}$ to the extension $\mathrm{ext}(\mathcal{E})$. The entropy distance of a local maximizer is a difference of free energies (39). We use functional calculus in compressed algebras (Definition 4.22.3).

Corollary 6.19. *Let $\rho \in \mathcal{S}_{\mathcal{A}}$ be any state. We denote $p := s(\rho)$ the support projection of ρ, $q := s(\pi_{\mathcal{E}}(\rho))$ the support projection of the projection $\pi_{\mathcal{E}}(\rho)$ and we fix a matrix $\theta \in \Theta$ such that $\pi_{\mathcal{E}}(\rho) = R_{q\mathcal{A}q}(q\theta q)$.*

1. *If $u \in (p\mathcal{A}p)_{\mathrm{sa}}$ is a traceless matrix, then $\frac{\partial}{\partial t}\mathrm{d}_{\mathcal{E}}(\rho + tu)|_{t=0} = \langle u, \log^{[p]}(\rho) - p\theta p\rangle$.*
2. *If ρ is a local maximizer of the entropy distance $\mathrm{d}_{\mathcal{E}}$ on the state space $\mathcal{S}_{\mathcal{A}}$, then $\rho = R_{p\mathcal{A}p}(p\theta p)$ and $\mathrm{d}_{\mathcal{E}}(\rho) = F_{q\mathcal{A}q}(q\theta q) - F_{p\mathcal{A}p}(p\theta p)$.*

Proof. By definition (72) of the projection $\pi_{\mathcal{E}}$ there exits a parameter $\theta \in \Theta$ and a projection $q \in \mathcal{P}^U$ such that $\pi_{\mathcal{E}}(\rho) = R_{q\mathcal{A}q}(q\theta q)$. We notice $p \preceq q$ from Lemma 4.18. Corollary 6.19 is proved in Theorem 31 in [61] for $q = \mathbb{1}$ (i.e. $\pi_{\mathcal{E}}(\rho)$ invertible in $\mathcal{A}$), $0 \in \Theta$ (Gibbsian families) and for Θ consisting of traceless matrices. Using the mean value chart (6.7), the proof of Theorem 31 in [61] is valid for affine parameter spaces Θ of trace-less matrices, including $0 \notin \Theta$: All assertions of the theorem are invariant under the substitution of $\theta \mapsto \theta + \lambda q$ for real λ, e.g.

$$\begin{aligned} F_{q\mathcal{A}q}(q(\theta + \lambda\mathbb{1})q) - F_{p\mathcal{A}p}(p(\theta + \lambda\mathbb{1})p) &= F_{q\mathcal{A}q}(q\theta q) + \lambda - [F_{p\mathcal{A}p}(p\theta p) + \lambda] \\ &= F_{q\mathcal{A}q}(q\theta q) - F_{p\mathcal{A}p}(p\theta p)\,. \end{aligned}$$

This proves our claim for arbitrary non-empty affine subspaces $\Theta \subset \mathcal{A}_{\mathrm{sa}}$ if $q = \mathbb{1}$.

Otherwise, if $q \neq \mathbb{1}$, then Theorem 6.16 shows $\mathrm{d}_{\mathcal{E}}(\rho) = \mathrm{d}_{\mathcal{E}_q}(\rho)$. We argue analogously as before but with the algebra $q\mathcal{A}q$ in place of $\mathcal{A}$. This is possible since $\pi_{\mathcal{E}}(\rho)$ is invertible in $q\mathcal{A}q$ and since $\rho \in q\mathcal{A}q$. The latter is true since $s(\rho) = p \preceq q$. □

Remark 6.20 (Earlier results). The idea to Proposition 6.17 goes back to Proposition 3.2 in [5] where (82) is proved on a subset of the probability simplex (6) on a finite set Ω. This inequality was extended in Corollary 2 in [42] to the whole probability simplex. Ay has also proved (82) on a subset of the state space of $\mathrm{Mat}(n,\mathbb{C})$. The extension (81) to the whole state space of a C*-subalgebra of $\mathrm{Mat}(n,\mathbb{C})$ as well as the improvement from (82) to (83) are new.

Corollary 6.19 was first proved in Proposition 3.1 in [5] for a subset of the probability simplex on a finite set Ω and was proved in Theorem 5.1 in [41] on the whole probability simplex.

6.6. Equality conditions for closures

In addition to the rI-closure $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$ and the norm closure $\overline{\mathcal{E}}$ of an exponential family $\mathcal{E}$ let us define the *geodesic closure*

$$\mathrm{cl}^{\mathrm{geo}}(\mathcal{E}) := \{\rho \in \mathcal{S}_{\mathcal{A}} \mid \rho \text{ is the norm limit of an e-geodesic in } \mathcal{E}\} \tag{84}$$

where e-geodesics are one-dimensional exponential families (24). The inclusions

$$\mathrm{cl}^{\mathrm{geo}}(\mathcal{E}) \subset \mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) \subset \overline{\mathcal{E}} \tag{85}$$

are already proved in Corollary 15 in [61]. The first inclusion follows from relative entropy estimates along e-geodesics. The second inclusion follows from the Pinsker-Csiszár inequality. Below we argue that strict inclusions are only possible for a non-commutative algebra $\mathcal{A}$. Examples in $\mathcal{A} = \mathrm{Mat}(2,\mathbb{C}) \oplus \mathbb{C}$ are the *Swallow family* with $\mathrm{cl}^{\mathrm{geo}}(\mathcal{E}) \subsetneq \mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$ and the *Staffelberg family* with $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) \subsetneq \overline{\mathcal{E}}$, see §3.5, §3.6 and also §IV.B and §IV.D in [61].

We recall from Definition 4.2.3 the lattice $\mathcal{F}_{\perp}(\mathbb{M}(U))$ of exposed faces of $\mathbb{M}(U)$.

Proposition 6.21. *The projection $\pi_U(\mathrm{cl}^{\mathrm{geo}}(\mathcal{E}))$ of the geodesic closure can be a proper subset of the mean value set $\mathbb{M}(U)$. In fact,*

$$\pi_U(\mathrm{cl}^{\mathrm{geo}}(\mathcal{E})) = \bigcup_{F \in \mathcal{F}_\perp(\mathbb{M}(U))} \mathrm{ri}(F)\,. \tag{86}$$

In particular, $\mathrm{cl}^{\mathrm{geo}}(\mathcal{E}) = \mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$ *holds if and only if all faces of the mean value set* $\mathbb{M}(U)$ *are exposed faces.*

Proof. Using (75) and Corollary 4.11 we can write[7] the geodesic closure of $\mathcal{E}$ in the form $\mathrm{cl}^{\mathrm{geo}}(\mathcal{E}) = \bigcup_p \mathcal{E}_p$, where the union extends over all non-zero projections p in the exposed projection lattice $\mathcal{P}_{\mathcal{A}}^{U,\perp}$ defined in (45). In (71) we have shown that each of the families $\mathcal{E}_p$ projects to $\pi_U(\mathcal{E}_p) = \mathrm{ri}\,\pi_U(\mathbb{F}_{\mathcal{A}}(p))$. So (86) follows from the lattice isomorphism (46). Theorem 6.16 shows the disjoint union $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) = \bigcup_p \mathcal{E}_p$, where the union extends over all non-zero projections p in the projection lattice $\mathcal{P}_{\mathcal{A}}^{U}$. So the equality $\mathrm{cl}^{\mathrm{geo}}(\mathcal{E}) = \mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$ is equivalent to $\mathcal{P}_{\mathcal{A}}^{U,\perp} = \mathcal{P}_{\mathcal{A}}^{U}$ and this, by the lattice isomorphism (46), mean that all faces of $\mathbb{M}_{\mathcal{A}}(U)$ are exposed faces. $\square$

We prove a condition for $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) = \overline{\mathcal{E}}$ in terms of the entropy distance (27).

Proposition 6.22. *We have* $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) = \overline{\mathcal{E}}$ *if and only if the entropy distance* $\mathrm{d}_{\mathcal{E}}$ *is norm continuous on* $\mathcal{S}_{\mathcal{A}}$.

Proof. If the inclusion $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) \subset \overline{\mathcal{E}}$ in (85) is strict, then there exists a norm convergent sequence $(\rho_i)_{i\in\mathbb{N}} \subset \mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$ with limit $\rho \in \mathcal{S}_{\mathcal{A}} \setminus \mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$ in the compact state space (see Proposition 4.5). By Theorem 6.16 we have $\mathrm{d}_{\mathcal{E}}(\rho) > 0$ while $\mathrm{d}_{\mathcal{E}}(\rho_i) = 0$ for $i \in \mathbb{N}$ hence $\mathrm{d}_{\mathcal{E}}$ is discontinuous at $\rho \in \mathcal{S}_{\mathcal{A}}$.

Conversely, let us prove that $\mathrm{d}_{\mathcal{E}}$ is lower semi-continuous if $\overline{\mathcal{E}} = \mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$. Since $\overline{\mathcal{E}}$ is a compact subset of $\mathcal{A}_{\mathrm{sa}}$, lower semi-continuity of relative entropy (Remark 2.4.1) implies lower semi-continuity of the minimum

$$\mathcal{S}_{\mathcal{A}} \to \mathbb{R}, \qquad \rho \mapsto \min\{S(\rho,\sigma) : \sigma \in \overline{\mathcal{E}}\}\,.$$

The proof given in Theorem 2, p. 116 in [11], uses a covering of $\overline{\mathcal{E}}$ by open balls. This minimum function equals $\mathrm{d}_{\mathcal{E}}$ by Theorem 6.16. Continuity of $\mathrm{d}_{\mathcal{E}}$ follows from the lower semi-continuity of $\mathrm{d}_{\mathcal{E}}$, see e.g. Lemma 4.2 in [5]. $\square$

The statements just proved can be formulated as necessary condition of commutativity of $\mathcal{A}$. If $\mathcal{A} \cong \mathbb{C}^n$ then $\mathcal{S}_{\mathcal{A}}$ is a simplex, hence $\mathbb{M}_{\mathcal{A}}(U)$ is a polytope and all faces of $\mathbb{M}_{\mathcal{A}}(U)$ are exposed faces. Then Proposition 6.21 proves $\mathrm{cl}^{\mathrm{geo}}(\mathcal{E}) = \mathrm{cl}^{\mathrm{rI}}(\mathcal{E})$. Moreover we have $\mathcal{T}^{\mathrm{rI}} = \mathcal{T}^{\|\cdot\|}$ by Corollary 5.19 hence $\mathrm{cl}^{\mathrm{rI}}(\mathcal{E}) = \overline{\mathcal{E}}$. In particular, Proposition 6.22 shows that the entropy distance $\mathrm{d}_{\mathcal{E}}$ is norm continuous in the setting of probability distributions on a finite measurable space (9). This was first proved in Lemma 4.2 in [5].

[7] This is also proved in Proposition 10 in [61].

7. Comments on the representation

We show that our results about the I-/rI-topology hold for an arbitrary finite-dimensional C*-algebra $\mathcal{A}$. In order to define exponential families in $\mathcal{A}$ we chose a representation of $\mathcal{A}$ as a C*-subalgebra of $\mathrm{Mat}(n,\mathbb{C})$. We show that the Complete projection theorem and Pythagorean theorem are independent of this choice.

The first object, the relative entropy, is monotone under C*-morphisms

$$\Phi : \mathcal{B} \to \mathcal{A}$$

between two unital C*-algebras [56], i.e. if $f, g : \mathcal{A} \to \mathbb{C}$ are two states and if $\Phi^*(f) := f \circ \Phi$, then $S(\Phi^*(f), \Phi^*(g)) \leq S(f, g)$ holds. If $\Phi : \mathcal{B} \to \mathcal{A}$ is a C*-isomorphism then Φ^{-1} provides the opposite inequality and $S(\Phi^*(f), \Phi^*(g)) = S(f, g)$ follows. In particular, our results about the I-/rI-topology in §5.2 are valid in any finite-dimensional C*-algebra independent of the representation.

Our definition of exponential family needs normal states on a finite-dimensional von Neumann algebra $\mathcal{A}$, represented by an algebra of linear operators on a Hilbert space. The normal states on $\mathcal{A}$ are represented by positive and normalized trace class operators ρ in $\mathcal{A}$, with associated linear functional

$$a \mapsto \langle a, \rho \rangle = \mathrm{tr}(a\rho)\,, \quad (a \in \mathcal{A})\,,$$

see Theorem 2.4.21 in [15] and § 2.1. However, not all representation are equally suitable. E.g. $\mathbb{C}$ represented as $\{x = (x_i)_{i\in\mathbb{N}} \in l^\infty \mid \exists \lambda \in \mathbb{C}, \forall i \in \mathbb{N} : x_i = \lambda\}$ has no normal state. So we restrict to representation on finite-dimensional Hilbert spaces. We want to see if our results are independent this choice.

Every finite-dimensional C*-subalgebra is, according to Theorem III.1.1 in [21], C*-isomorphic to the direct sum

$$\mathcal{B} := \bigoplus_{i=1}^{N} \mathrm{Mat}(k_i, \mathbb{C})\,, \tag{87}$$

where $N \in \mathbb{N}_0$ and $k \in \mathbb{N}^N$ is a multi-index. Any C*-algebra $\mathcal{A}$ of linear operators on a finite-dimensional Hilbert space, which is C*-isomorphic to $\mathcal{B}$, has the form of

$$\mathcal{A} := \bigoplus_{i=1}^{N} \left\{ \bigoplus_{j=1}^{m_i} a_i \mid a_i \in \mathrm{Mat}(k_i, \mathbb{C}) \right\} \oplus 0_l$$

up to unitary equivalence. Here $m_i \geq 1$ for $i = 1, \ldots, N$ and $l \geq 0$ are integers see Corollary III.2.1 in [21]. Moreover, there is a C*-isomorphism $\Phi : \mathcal{B} \to \mathcal{A}$

$$\Phi\left(\bigoplus_{i=1}^{N} b_i \right) = \bigoplus_{i=1}^{N} \bigoplus_{j=1}^{m_i} b_i \oplus 0_l\,, \quad (b_1, \ldots, b_N) \in \mathcal{B}\,. \tag{88}$$

Let us begin to discuss mean values. While the mean value set (29) pleases with a simple Euclidean geometry, the isomorphic convex support (31) has other advantages. Firstly, it is equivariant under the isomorphism (88): $f(b) = (\Phi^{-1})^* f(\Phi(b))$

holds for all states f on $\mathcal{B}$ and $b \in \mathcal{B}$. The mean value set is not equivariant. Another advantage of the convex support is that its algebraic decomposition into faces becomes a simple inclusion $\mathrm{cs}_{p\mathcal{A}p} \subset \mathrm{cs}_{\mathcal{A}}$, see Lemma 4.19.

Let us discuss exponential families. The adjoint $\Phi^* : \mathcal{A}^* \to \mathcal{B}^*$ of (88) is given for $F_i \in \mathrm{Mat}(k_i, \mathbb{C})$, $i = 1, \ldots, N$, by

$$\Phi^* \left(\bigoplus_{i=1}^{N} \bigoplus_{j=1}^{m_i} F_i \oplus 0_l \right) = \bigoplus_{i=1}^{N} m_i F_i \,. \tag{89}$$

Lemma 7.1. *Let $\Theta \subset \mathcal{B}_{\mathrm{sa}}$ be a non-empty affine subspace and $\mathcal{E} := R_{\mathcal{B}}(\Theta)$. Let $\theta_0 := \bigoplus_{i=1}^{N} \ln(m_i) \mathbb{1}_{k_i} \in \mathcal{B}_{\mathrm{sa}}$. Then the affine space $\widetilde{\Theta} := \Phi(\Theta - \theta_0) \subset \mathcal{A}_{\mathrm{sa}}$ satisfies $(\Phi^*)^{-1}(\mathcal{E}) = R_{\mathcal{A}}(\widetilde{\Theta})$.*

Proof. By (89) we have $\Phi^* \circ R \circ \Phi(\theta) = R(\,\theta + \bigoplus_{i=1}^{N} \ln(m_i)\mathbb{1}_{k_i})$ for $\theta \in \Theta$. $\square$

Lemma 7.1 shows that the class of exponential families is preserved under the isomorphism (88). Clearly $\Phi^*(\rho + U^\perp) = \Phi^*(\rho) + \Phi^{-1}(U)^\perp$ holds for all $\rho \in \mathcal{S}_{\mathcal{A}}$ and $U \subset \mathcal{A}_{\mathrm{sa}}$. So the Complete Pythagorean theorem and the Complete projection theorem (Theorem 6.12 and Theorem 6.16) are valid for C*-algebras represented on finite-dimensional Hilbert spaces independent of the choice of representation. Gibbsian families are not equivariant under the isomorphism (88) because $\widetilde{\Theta}$ in Lemma 7.1 is not necessarily a linear space even though Θ is a linear space.

Acknowledgements. I would like to thank Arleta Szkoła for discussions about quantum statistics and for helpful and critical remarks about this article. I would like to thank František Matúš for an introduction to sequential convergences and Nihat Ay for an introduction to maximizing relative entropy from exponential families. I would like to say my sincere and deep gratitude to my PhD advisor Andreas Knauf who has supported me for so many years already in the Diploma career in 2001. This work was supported by the DFG projects "Geometry and Complexity in Information Theory" and "Quantenstatistik: Entscheidungsprobleme und entropische Funktionale auf Zustandsräumen".

References

[1] E. M. Alfsen, F. W. Shultz: State Spaces of Operator Algebras, Birkhäuser, Boston (2001).

[2] S. Amari: Differential-Geometrical Methods in Statistics, Lecture Notes in Statistics 28, Springer, Berlin (1985).

[3] S. Amari: Information geometry on hierarchy of probability distributions, IEEE Trans. Inf. Theory 47 (2001) 1701–1711.

[4] S. Amari, H. Nagaoka: Methods of Information Geometry, Translations of Mathematical Monographs 191, AMS, Providence (2000).

[5] N. Ay: An information-geometric approach to a theory of pragmatic structuring, Ann. Probab. 30 (2002) 416–436.

[6] N. Ay, A. Knauf: Maximizing multi-information, Kybernetika 42 (2006) 517–538.

[7] O. Barndorff-Nielsen: Information and Exponential Families in Statistical Theory, John Wiley & Sons, New York (1978).

[8] F. Benatti: Dynamics, Information and Complexity in Quantum Systems, Springer, Dordrecht (2009).

[9] I. Bengtsson, S. Weis, K. Życzkowski: Geometry of the set of mixed quantum states: an apophatic approach, in: Geometric Methods in Physics (Białowieża, 2011), P. Kielanowski et al. (ed.), Trends in Mathematics, Birkhäuser, Basel (2013) 175–197.

[10] I. Bengtsson, K. Życzkowski: Geometry of Quantum States. An Introduction to Quantum Entanglement, Cambridge University Press, Cambridge (2006).

[11] C. Berge: Topological Spaces, Oliver & Boyd, Edinburgh (1963).

[12] R. Bhatia: Matrix Analysis, Springer, New York (1997).

[13] G. Birkhoff: Lattice Theory, 3rd. Ed., AMS Colloquium Publications, AMS, Providence (1973).

[14] I. Bjelaković, J.-D. Deuschel, T. Krüger, R. Seiler, R. Siegmund-Schultze, A. Szkoła: A quantum version of Sanov's theorem, Commun. Math. Phys. 260 (2005) 659–671.

[15] O. Bratteli, D. W. Robinson: Operator Algebras and Quantum Statistical Mechanics 1, 2nd Ed., Springer, New York (2002).

[16] N. N. Čencov: Statistical Decision Rules and Optimal Inference, Translations of Mathematical Monographs 53, AMS, Providence (1982).

[17] I. Csiszár: On topological properties of f-divergences, Stud. Sci. Math. Hung. 2 (1967) 329–339.

[18] I. Csiszár, F. Matúš: Information projections revisited, IEEE Trans. Inf. Theory 49 (2003) 1474–1490.

[19] I. Csiszár, F. Matúš: On information closures of exponential families: a counterexample, IEEE Trans. Inf. Theory 50 (2004) 922–924.

[20] I. Csiszár, F. Matúš: Closures of exponential families, Ann. Probab. 33 (2005) 582–600.

[21] K. R. Davidson: C*-Algebras by Example, Fields Institute Monographs 6, AMS, Providence (1996).

[22] R. M. Dudley: On sequential convergence, Trans. Amer. Math. Soc. 112 (1964) 483–507; correction ibid. 148 (1970) 623–624.

[23] R. M. Dudley: Consistency of M-estimators and one-sided bracketing, in: High Dimensional Probability (Oberwolfach, 1996), E. Eberlein et al. (ed.), Progr. Probab. 43, Birkhäuser, Basel (1998) 33–58.

[24] N. Dunford, J. T. Schwartz: Linear Operators. I: General Theory, Interscience, London (1958).

[25] R. Engelking: General Topology, Sigma Series in Pure Mathematics 6, Heldermann, Berlin (1989).

[26] G. L. Gilardoni: On Pinsker's and Vajda's type inequalities for Csiszár's f-divergences, IEEE Trans. Inf. Theory 56 (2010) 5377–5386.

[27] M. R. Grasselli, R. F. Streater: On the uniqueness of the Chentsov metric in quantum information geometry, Infin. Dimens. Anal. Quantum Probab. Relat. Top. 4 (2001) 173–182.

[28] B. Grünbaum: Convex Polytopes, 2nd Ed., Springer, New York (2003).

[29] P. R. Halmos: Finite-Dimensional Vector Spaces, Springer, New York (1987).

[30] P. Harremoës: Information topologies with applications, in: Entropy, Search, Complexity, I. Csiszáe et al. (ed.), Bolyai Society Mathematical Studies 16, Springer, Berlin (2007) 113–150.

[31] M. Hayashi: Quantum Information Theory. An Introduction, Springer, Berlin (2006).

[32] A. S. Holevo: Probabilistic and Statistical Aspects of Quantum Theory, 2nd Ed., Edizioni della Normale, Pisa (2011).

[33] R. S. Ingarden, A. Kossakowski, M. Ohya: Information Dynamics and Open Systems, Kluwer, Dordrecht (1997).

[34] E. T. Jaynes: Information theory and statistical mechanics. I, II, Phys. Rev., II. Ser. 106 (1957) 620–630; 108 (1957) 171–190.

[35] A. Jenčová: Geometry of quantum states: dual connections and divergence functions, Rep. Math. Phys. 47 (2001) 121–138.

[36] T. Kato: Perturbation Theory for Linear Operators, Springer, Berlin (1995).

[37] S. G. Krantz, H. R. Parks: A Primer of Real Analytic Functions, 2nd Ed., Birkhäuser, Boston (2002).

[38] S. Kullback, R. A. Leibler: On information and sufficiency, Ann. Math. Stat. 22 (1951) 79–86.

[39] E. H. Lieb: Convex trace functions and the Wigner-Yanase-Dyson conjecture, Adv. Math. 11 (1973) 267–188.

[40] G. Lindblad: Entropy, information and quantum measurements, Comm. Math. Phys. 33 (1973) 305–322.

[41] F. Matúš: Optimality conditions for maximizers of the information divergence from an exponential family, Kybernetika 43 (2007) 731–746.

[42] F. Matúš, N. Ay: On maximization of the information divergence from an exponential family, in: Proceedings of WUPES'03, J. Vejnarová (ed.), University of Economics Prague (2003) 199–204.

[43] K. Modi, T. Paterek, W. Son, V. Vedral, M. Williamson: Unified view of quantum and classical correlations, Phys. Rev. Lett. 104 (2010) 080501.

[44] G. J. Murphy: C*-Algebras and Operator Theory, Academic Press, Boston (1990).

[45] H. Nagaoka: Differential geometrical aspects of quantum state estimation and relative entropy, in: Quantum Communications and Measurement (Nottingham, 1994), V. P. Belavkin et al. (ed.), Plenum Press, New York (1995) 449–452.

[46] M. A. Nielsen, I. L. Chuang: Quantum Computation and Quantum Information, Cambridge University Press, Cambridge (2000).

[47] M. Ohya, D. Petz: Quantum Entropy and its Use, Springer, Berlin (1993).

[48] D. Petz: Quasi-entropies for finite quantum systems, Rep. Math. Phys. 23 (1986) 57–65.

[49] D. Petz: Geometry of canonical correlation on the state space of a quantum system, J. Math. Phys. 35 (1994) 780–795.

[50] D. Petz: Quantum Information Theory and Quantum Statistics, Springer, Berlin (2008).

[51] M. Reed, B. Simon: Methods of Modern Mathematical Physics I, 2nd Ed., Academic Press, New York (1980).

[52] R. T. Rockafellar: Convex Analysis, Princeton University Press, Princton (1972).

[53] D. Ruelle: Statistical Mechanics. Rigorous Results, World Scientific, Singapore (1999).

[54] R. Sanyal, F. Sottile, B. Sturmfels: Orbitopes, Mathematika 57 (2011) 275–314.

[55] R. Schneider: Convex Bodies: The Brunn-Minkowski Theory, Cambridge University Press, Cambridge (1993).

[56] A. Uhlmann: Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory, Commun. Math. Phys. 54 (1977) 21–32.

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

[58] S. Weis: Exponential Families with Incompatible Statistics and Their Entropy Distance, PhD Thesis, University of Erlangen (2010).

[59] S. Weis: A note on touching cones and faces, J. Convex Analysis 19 (2012) 323–353.

[60] S. Weis: Quantum convex suppport, Linear Algebra Appl. 435 (2011) 3168–3188; correction: Linear Algebra Appl. 436 (2012) p. xvi.

[61] S. Weis, A. Knauf: Entropy distance: new quantum phenomena, J. Math. Phys. 53 (2012) 102206.

[62] D. Werner: Funktionalanalysis, 3rd Ed., Springer, Berlin (2000).

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