Information topologies on non-commutative state spaces

Stephan Weis
Max Planck Institute for Mathematics in the Sciences
Leipzig, Germany
November 24, 2011

Abstract – We investigate the I-topology and the rI-topology on the state spaces of a C*-subalgebra of Mat(n, C) which are defined in terms of convergence with respect to the relative entropy. In quantum information geometry there are a Pythagorean theorem and a projection theorem, valid for an exponential family. We achieve the completion of these theorems to the rI-closure of the exponential family. The completion to the norm closure is not possible since this can be strictly larger than the rI-closure.

The complete projection theorem proves the existence of an rI-projection, a projection with linear fibers from the whole state space onto the rI-closure of the exponential family. The rI-projection allows to study the entropy distance (the infimum of the relative entropy) of a state from the exponential family. We discuss the non-commutative feature of a discontinuous entropy distance and we prove two necessary conditions for local maximizers of the entropy distance. The complete Pythagorean theorem solves a problem in quantum information theory: Maximization of the von Neumann entropy under linear constraints. We provide previously unknown solutions without any support restrictions. The solution set is the rI-closure of the well-known exponential family of Gibbs ensembles.

Index Terms – non-commutative algebra, relative entropy, information topology, exponential family, convex support, Pythagorean theorem, projection theorem, von Neumann entropy.

AMS Subject Classification: 81P45, 81P16, 54D55, 94A17, 90C26.

Contents

1 Introduction .................................................. 2
  1.1 Information convergence in classical probability theory .......... 3
  1.2 Representation of states and mean values ......................... 4
  1.3 The Pythagorean theorem and the projection theorem ............... 6
  1.4 New results on the I- and rI-topology in a matrix algebra .......... 8
  1.5 New results on the rI-closure of an exponential family ............. 10

2 Analysis on the state space of a matrix algebra .................. 12
  2.1 Projections and functional calculus ................................ 13
  2.2 Continuity and convexity of the relative entropy .................... 15
  2.3 Two perturbative statements ..................................... 15
  2.4 Lattices of faces and projections .................................. 19

1sweis@mis.mpg.de
1 Introduction

The Pythagorean theorem and projection theorem in information geometry make statements about the distance of a probability measure from a family of probability measures, see §3 in [AN00] by Amari and Nagaoka and §I.C in [CM03] by Csiszár and Matúš. They provide a geometric frame for 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 [Ba78, Če82, CM03, CM05] using the I-/rI-convergence defined in terms of the relative entropy. This convergence is explained in §1.1. In quantum information theory, see e.g. [AN00, Be09, BZ06, Ho11, IO97, NC00, Pe08], there is also a relative entropy, the Umegaki relative entropy, and one can ask the analogue questions as in classical probability theory.

We report in §1.4 on the I-/rI-convergence on the state space of a C*-subalgebra of Mat(n, C) and about the associated topology, namely the I-/rI-topology. The I-topology includes the rI-topology and both include the norm topology. The I-/rI-topology has properties of a metric topology, e.g. its open sets are unions of disks arising from the relative entropy. But it is quite distinct from the norm topology, e.g. it is not second countable (unless the algebra is commutative) and the state space splits into the connected components of its faces with respect to the I-topology.

The usefulness of the I-/rI-topology in quantum statistics and quantum hypothesis testing is not yet clarified. In contrast to classical probability theory there is no canonical choice of a computational basis and the consequences for measurement and observation will have to be taken into consideration. We have found the general result that the infimum of the relative entropy of a state from a set of states does not decrease if a topological closure operation is applied to the set. This property can be used e.g. in the Sanov theorem of quantum hypothesis testing [BS05].

We have found several applications of the rI-topology in quantum information geometry and quantum information theory. The Pythagorean theorem and projection theorem in information geometry are valid for the relative entropy and for exponential families (we recall them in §1.3). In §1.5 we present their extensions to a complete Pythagorean theorem and to a complete projection theorem.

---

2Here and in the sequel “I” stands for “information” and “rI” for “reverse information”.

---

3 Information topologies on the state space

3.1 The topology of a divergence function

3.2 The I-topology and the rI-topology

4 Exponential families in a matrix algebra

4.1 The mean value chart

4.2 The complete projection theorem

4.3 Local maximizers of the entropy distance

4.4 Non-commutative aspects of exponential families

4.5 The complete Pythagorean theorem

5 Comment on the representation
The complete projection theorem implies an rI-projection with linear fibers, defined for all states, onto the rI-closure of an exponential family. The rI-projection allows to compute the entropy distance from an exponential family. An important example of entropy distance is the stochastic interaction measure of multi-information, see the references in [KW11]. This interaction measure is interesting in statistical physics where phase coexistence generically appears in highly correlated systems. Maximization of the entropy distance from an exponential family is advocated in [Ay02] as a structuring principle in natural systems. In §4.3 we prove two necessary conditions for its local maximizers. One of the condition is an upper bound on the rank, it enforces a certain degree of determinism on a local maximizer.

We present in §1.5 an application of the complete Pythagorean theorem. This allows us to compute on the state space of a C*-subalgebra of Mat(n, C) all maximizers of the von Neumann entropy under linear constraints. To our best knowledge, the non-invertible solutions were unknown previously, partial answers were found by Wichmann [Wi63]. An innovation in our analysis is to study non-exposed faces of projected state spaces using Grünbaum’s notion of poonem. The notion of poonem is equivalent to the notion of face and is defined in §2.4 by sequences of consecutively exposed faces, also known as access sequences [CM05].

In few words, the I-/rI-topology behaves much like a metric topology and at the same time it has still capacity to respect the convex geometry of the state space. The rI-topology has many applications. Further research will show if shorter proofs exist for the complete projection theorem, e.g. using a parallelogram-like identity as in Theorem 1 in [CM03]. Conversely, our results may be helpful to discover new identities or inequalities in quantum information theory.

1.1 Information convergence in classical probability theory

We recall some key properties of information convergence in probability theory. A comprehensive historical account is described in §1.C in [CM03].

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 $\sigma$-finite measure $\lambda$ and $p(x)$ resp. $q(x)$ is the Radon-Nikodym derivative of $P$ resp. $Q$, then the relative entropy is

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

This equals zero if and only if $P = Q$ and otherwise $D(P \| Q)$ is strictly positive or $+\infty$. Given a sequence $(P_n)_{n \in \mathbb{N}} \subseteq \mathcal{M}$ and a probability measure $P \in \mathcal{M}$ we have, according to [CM03], I-convergence resp. rI-convergence of $(P_n)_{n \in \mathbb{N}}$ to $P$ if

$$\lim_{n \to \infty} D(P_n \| P) = 0 \quad \text{resp.} \quad \lim_{n \to \infty} D(P \| P_n) = 0.$$  

(2)

Csiszár has studied these convergences in the much wider context of $f$-divergence. He has proved in Theorem 3 in [Cs67] that the information neighborhoods

$$\{Q \in \mathcal{M} \mid D(Q \| P) < \epsilon\} \quad \text{resp.} \quad \{Q \in \mathcal{M} \mid D(P \| Q) < \epsilon\} \quad (P \in \mathcal{M}, \epsilon > 0)$$  

(3)

Non-exposed faces were ignored in the erroneous statement of Theorem 1 e) in [Wi63].
do not define a base of a topology if \((X, \mathcal{A}) = (\mathbb{N}, 2^{\mathbb{N}})\).

At a later time Dudley and Harremoës [Du98, Ha07] have studied topologies on \(\mathcal{M}\) that are defined by sequential convergence (2). In a different approach, Csiszár [Cs67] considers \(\mathcal{M}\) as a Fréchet (V)-space (a generalization of topological space, see §1 in [Si52]) arising from the family of information neighborhoods (3).

We show in §3 that the two analogous approaches in a \(C^*\)-subalgebra of \(\text{Mat}(n, \mathbb{C})\) are compatible: The Fréchet (V)-space of information neighborhoods is a topological space and its topology corresponds to the convergence.

### 1.2 Representation of states and mean values

We shall work throughout with a \(C^*\)-subalgebra \(\mathcal{A}\) of \(\text{Mat}(n, \mathbb{C})\), i.e. a norm-closed self-adjoint subalgebra \(\mathcal{A}\) of \(\text{Mat}(n, \mathbb{C})\).

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 \(f(1) = 1\). Bratteli and Robinson provide in Theorem 2.4.21 in [Br87] a proof of the one-to-one correspondence between states \(f\) on \(\mathcal{A}\) and matrices \(\rho \succeq 0\) of trace one in \(\mathcal{A}\)

\[
\mathcal{S}_\mathcal{A} = \{ \rho \in \mathcal{A} \mid \rho \succeq 0, \text{tr}(\rho) = 1 \}.
\]

**Definition 1.1.** 1. We denote the identity in \(\text{Mat}(n, \mathbb{C})\) by \(1_n\), the identity in \(\mathcal{A}\) by \(1\). By \(\mathcal{A}_{sa}\) we denote the real vector space of self-adjoint matrices in \(\mathcal{A}\). Let \(a \in \mathcal{A}\). The spectrum of a is

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

its elements are the spectral values of \(a\) in \(\mathcal{A}\). The matrix \(a\) is positive semi-definite if \(a \in \mathcal{A}_{sa}\) 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 [Mu90]. The matrix \(b\) is unique and one defines \(\sqrt{a} := b\). We have \(a^*a \succeq 0\) and put \(|a| := \sqrt{a^*a}\).

2. The standard trace \(\text{tr}\) turns \(\text{Mat}(n, \mathbb{C})\) into a complex Hilbert space with the Hilbert-Schmidt inner product \(\langle a, b \rangle := \text{tr}(ab^*)\) for \(a, b \in \text{Mat}(n, \mathbb{C})\) and we use the two-norm \(\|a\|_2 := \sqrt{\langle a, a \rangle}\). We consider also the spectral norm \(\|a\|\), which is the square root of the largest eigenvalue of \(a^*a\) and the trace norm \(\|a\|_1 := \text{tr}|a|\).

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 \(\mathcal{A}_{sa}\) and we consider \((\mathcal{A}_{sa}, \langle \cdot, \cdot \rangle)\) a Euclidean vector space with the Hilbert-Schmidt inner product.

3. In a 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 \(\perp \) \(\subset \mathbb{E}\) 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 \(A\) is a non-empty affine subspace of \(\mathbb{E}\), then we denote the translation vector space of \(A\) by \(\text{lin}(A)\). We
4. The mean value set of a linear subspace $U \subset \mathcal{A}_{sa}$ is the orthogonal projection of the state space onto $U$

$$
\mathbb{M}(U) := \mathbb{M}_A(U) = \pi_U(\mathcal{S}_A) \subset U.
$$

The mean value mapping is defined for $u_1, \ldots, u_k \in \mathcal{A}_{sa}$ by

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

The convex support of $u_1, \ldots, u_k \in \mathcal{A}_{sa}$ is

$$
\text{cs}(u_1, \ldots, u_k) := \text{cs}_A(u_1, \ldots, u_k) = \{ m_{u_1, \ldots, u_k}(\rho) \mid \rho \in \mathcal{S}_A \} \subset \mathbb{R}^k.
$$

Conditional probability measures and their formal generalization in a C*-subalgebra of $\text{Mat}(n, \mathbb{C})$ are correctly rendered by spectral values (and not by eigenvalues). These concepts will be used in §4.2 and thereafter.

**Remark 1.2.**

1. We denote the *probability simplex* of a non-empty (at most) countable set $\Omega$ by

$$
\mathbb{P}(\Omega) := \{ p = (p_\omega)_{\omega \in \Omega} \mid [0, 1]^\Omega, \sum_{\omega \in \Omega} p_\omega = 1 \}.
$$

The elements of $\mathbb{P}(\Omega)$ are called *probability vectors* on $\Omega$. In the sequel we shall identify the probability vectors $p \in \mathbb{P}(\Omega)$ with the probability measures $P$ on $(\Omega, 2^\Omega)$ using $P(A) := \sum_{\omega \in A} p_\omega$ for $A \subset \Omega$. If $\Omega = \{1, \ldots, N\}$ is finite, we shall write $\mathbb{P}(N) := \mathbb{P}(\{1, \ldots, N\})$.

2. Let $\mathcal{A} \subset \text{Mat}(n, \mathbb{C})$ be C*-isomorphic to $\mathbb{C}^N$, e.g. $\mathcal{A}$ can be the set of diagonal matrices if $N = n$. If $e_1, \ldots, e_N$ denotes the standard ONB of $\mathbb{C}^N$, then every state $\rho \in \mathcal{S}_A \cong \mathcal{S}_{\mathbb{C}^N}$ defines a probability vector by setting $p_\omega := \text{tr}(\rho e_\omega)$ for $\omega \in \{1, \ldots, N\}$ and we have a one-to-one correspondence $\mathbb{P}(N) \cong \mathcal{S}_A$.

3. We use spectral values to describe conditional probability distributions. If $A \subset \{1, \ldots, N\}$ then we have the natural inclusion of $\mathbb{P}(A) \subset \mathbb{P}(N)$. If $P \in \mathbb{P}(N)$ and $P(A) > 0$, then the conditional probability measure $P(\cdot | A) \in \mathbb{P}(A)$ is

$$
P(\{i\} | A) := P(\{i\}) / P(A), \quad (i \in A).
$$

With the identifications in 2 we consider $P(\cdot | A) \in \mathbb{C}^A \subset \mathbb{C}^N \subset \text{Mat}(n, \mathbb{C})$ in three algebras. If $A \subset \{1, \ldots, N\}$, then 0 is a spectral value of $P(\cdot | A)$ in $\mathbb{C}^N$ and a spectral value (and eigenvalue) in $\text{Mat}(n, \mathbb{C})$. On the other hand the spectral values of $P(\cdot | A)$ in $\mathbb{C}^A$ give us the conditional probabilities of $i \in A$.

4. Barndorff-Nielsen [Ba78] has introduced the concept of convex support in probability theory, it was studied in [Če82] and was refined in [CM03, CM05] to investigate mean values and exponential families. For a finite measurable space their definition reduces to ours with $\mathcal{A} \cong \mathbb{C}^N$. 


5. The convex support introduces coordinates on the mean value set $M(U)$. If $u_1, \ldots, u_k \in A_{sa}$ and $U := \text{span}_\mathbb{R}(u_1, \ldots, u_k)$, then the convex bodies $M(U) \cong \text{cs}(u_1, \ldots, u_k)$ are “affinely isomorphic” (see Remark 1.1.1 in [We11]): The mean value mapping restricts to the bijection

$$m_{u_1, \ldots, u_k} |_{M(U)} : M(U) \to \text{cs}(u_1, \ldots, u_k) = \{m_{u_1, \ldots, u_k}(\rho) \mid \rho \in S\}$$

such that $m_{u_1, \ldots, u_k} \circ \pi_U = m_{u_1, \ldots, u_k}$. We prefer the Hilbert-Schmidt geometry in $A_{sa}$ to the coordinates in $\mathbb{R}^k$ and perform most of our analysis in $M(U)$. Exceptions are Corollary 4.16, Theorem 4.26 and Lemma 5.1.

1.3 The Pythagorean theorem and the projection theorem

Two theorems in quantum information geometry will be extended in §4 from the invertible states to the whole state space. Here we recapitulate these theorems with instructions to the easy proofs.

**Definition 1.3.** The relative entropy of $\rho \in S$ from $\sigma \in S$ is $S(\rho, \sigma) = +\infty$ unless $\text{Im}(\rho) \subset \text{Im}(\sigma)$ and then

$$S(\rho, \sigma) := \text{tr}(\rho(\log(\rho) - \log(\sigma))).$$

The logarithm can be defined by functional calculus, see Remark 2.4.4.

The relative entropy satisfies $S(\rho, \sigma) \geq 0$ for all $\rho, \sigma \in S$ with equality if and only if $\rho = \sigma$, see e.g. §11.3 in [Pe08] or §11.3 of [NC00] (This result is sometimes cited as Klein’s inequality). Convexity and continuity properties are recalled in §2.2.

The *Pythagorean theorem of relative entropy*, see e.g. §3.4 in [Pe08], applies to states $\rho, \sigma, \tau \in S$ where $\sigma, \tau$ are invertible and $\rho - \sigma \perp \log(\tau) - \log(\sigma)$ (using the Hilbert-Schmidt inner product). We have

$$S(\rho, \sigma) + S(\sigma, \tau) = S(\rho, \tau). \quad (8)$$

With relative entropy replaced by squared distance, this equation reminds us of the Pythagorean theorem in Euclidean geometry. Classical counterparts of the Pythagorean theorems (8) and (9) and of the projection theorem (11) are discussed e.g. in §3.4 in [AN00].

**Definition 1.4.** We use the real analytic function $^4 R_A : A_{sa} \to A_{sa}$,

$$R(\theta) = R_A(\theta) := \exp_A(\theta)/\text{tr}(\exp_A(\theta)),$$

the exponential $\exp_A$ is defined by functional calculus in the algebra $A$, see Definition 2.3.3. For a non-empty real affine subspace $\Theta \subset A_{sa}$ we define an exponential family in $A$

$$\mathcal{E} := R_A(\Theta) = \{R_A(\theta) \mid \theta \in \Theta\}.$$

We call a one-dimensional exponential family (with or without parametrization) $e$-geodesic. We use the translation vector space $U := \text{lin}(\Theta) = \Theta - \Theta$. 
In the literature, e.g. in §3.4 in [Pe08], the curve $t \mapsto R_A \circ a(t)$ is called e-geodesic, if $a : \mathbb{R} \to \Theta$ is an affine map.

**Example 1.5.** We shall use the Pauli $\sigma$-matrices

$\sigma_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_2 := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ and $\sigma_3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

and for $b = (b_1, b_2, b_3) \in \mathbb{R}^3$ we abbreviate $b\hat{\sigma} := b_1\sigma_1 + b_2\sigma_2 + b_3\sigma_3$.

The *Staffelberg family*, studied in [KW11], is the exponential family

$$R(\text{span}_R(\sigma_1 \oplus 0, \sigma_2 \oplus 1))$$

in the algebra $\text{Mat}(2, \mathbb{C}) \oplus \mathbb{C}$. This exponential family is depicted in Figure 1.

The Pythagorean theorem (8) applies to exponential families. We have for states $\rho \in \mathcal{S}$ and $\sigma, \tau \in \mathcal{E}$, such that $\rho - \sigma \perp U$,

$$S(\rho, \sigma) + S(\sigma, \tau) = S(\rho, \tau). \tag{9}$$

The condition $\rho - \sigma \perp U$ means that the Euclidean straight line from $\sigma$ to $\rho$ is perpendicular to the exponential family $\mathcal{E}$ with respect to the BKM-Riemannian metric, see Remark 4.2. This is indicated by the right angle in Figure 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 $\pi_{\mathcal{E}}(\rho)$ and a projection to $\mathcal{E}$ is defined by

$$\pi_{\mathcal{E}} : (\mathcal{E} + U^\perp) \cap \mathcal{S} \to \mathcal{E}, \quad \rho \mapsto \pi_{\mathcal{E}}(\rho) \tag{10}$$

The entropy distance (12) of $\rho$ from $\mathcal{E}$ is

$$d_{\mathcal{E}}(\rho) = S(\rho, \pi_{\mathcal{E}}(\rho)). \tag{11}$$

Indeed, the intersection has at least one point. If two states $\sigma, \tau$ are in the intersection, then $\rho - \sigma \perp U$ and $\rho - \tau \perp U$. Equality $\sigma = \tau$ follows if we add the two corresponding Pythagorean equations (9).

---

4On first reading this section $\mathcal{A} := \text{Mat}(n, \mathbb{C})$ and $\exp_{\mathcal{A}}(a) := \sum_{i=0}^{\infty} a^i/i!$ for $a \in \mathcal{A}$ can be assumed. If $\mathcal{A}$ is $C^*$-subalgebra of $\text{Mat}(n, \mathbb{C})$ we have $\exp_{\mathcal{A}}(a) = 1 + \sum_{i=1}^{\infty} a^i/i!$ where 1 can differ from the identity $I_n$ of $\text{Mat}(n, \mathbb{C})$, see Definition 1.1.1.
1.4 New results on the I- and rI-topology in a matrix algebra

We discuss similarities of the I-/rI-topology with a metric topology. Other interesting properties of the I-/rI-topology, e.g. its relation to convex geometry, are postponed to §3.2. Since the two topologies have a lot in common, we use a prefix variable $\omega \in \{I, rI\}$ for $\omega$-closure, $\omega$-topology, etc.

Unless otherwise specified we will use the norm topology.

The infimum of the relative entropy of a state from a set of states is important in quantum information theory [CM03, BS05]. We choose this as our starting point.

**Definition 1.6.** Let $\rho, \sigma \in S$. We use the short-hand notations

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

The infimum over $X \subset S$ is denoted by

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

The function

$$d_X(\rho) := S^{rI}(\rho, X) = \inf_{\tau \in X} S(\rho, \tau)$$

is called *entropy distance* in [KW11].

The conditions (13) and (14) imply a large part of our topological results. The *Pinsker-Csiszár inequality*, see e.g. §3.4 in [Pe08], states that for $\rho, \sigma \in S$ we have

$$2S(\rho, \sigma) \geq \|\rho - \sigma\|_1^2. \quad (13)$$

Here the trace norm from Definition 1.1.2 is used. We prove in Proposition 3.18 for all $\rho, \sigma \in S$ and $(\sigma_i)_{i \in \mathbb{N}} \subset S$ the continuity result of

$$\lim_{i \to \infty} S^\omega(\sigma, \sigma_i) = 0 \implies \lim_{i \to \infty} S^\omega(\rho, \sigma_i) = S^\omega(\rho, \sigma). \quad (14)$$

This statement means that the relative entropy is continuous in the first argument for the I-topology and in the second argument for the rI-topology (see Definition 1.9 and Remark 2.6.2).

**Definition 1.7.** The $\omega$-closure of $X \subset S$ is $\text{cl}^\omega(X) := \{\rho \in S \mid S^\omega(\rho, X) = 0\}$.

We will see in Theorem 3.20 that taking $\omega$-closures is allowed under the infimum:

**Theorem.** For all $\rho \in S$ and $X \subset S$ we have $S^\omega(\rho, X) = S^\omega(\rho, \text{cl}^\omega(X))$.

The analogue of the $\omega$-closure is well-known in probability theory [CM05], however it can strictly decrease the infimum as Example 3.11 demonstrates. We turn to topology. 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 $i \geq N$. 
1 INTRODUCTION

Definition 1.8. We define a family of subsets of the state space $S$ by

$$T^\omega := \left\{ U \subset S \mid \rho \in U, (\rho_i)_{i \in \mathbb{N}} \subset S \text{ and } \lim_{i \to \infty} S^\omega(\rho, \rho_i) = 0 \Rightarrow \rho_i \in U \text{ for large } i \right\}.$$ 

The open $\omega$-disk about $\rho \in S$ with radius $\epsilon \in (0, \infty]$ is

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

and the closed $\omega$-disk about $\rho \in S$ with radius $\epsilon \in (0, \infty]$ is

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

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

The family $T^\omega$ is clearly a topology on $S$ and the inclusion $T\|\cdot\| \subset T^\omega$ follows already from the Pinsker-Csiszár inequality (13). We prove in Theorem 3.21.4 the following.

Theorem. The inclusions $T\|\cdot\| \subset T^I \subset T^1$ hold and we have for all sequences $(\rho_i)_{i \in \mathbb{N}} \subset S$ and $\rho \in S$

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

Definition 1.9. We call $T^\omega$ the $\omega$-topology on $S$.

We recall topological concepts used in the following discussion.

Definition 1.10. Let $(X, \mathcal{T})$ be a topological space. An open set $U \in \mathcal{T}$ is a neighborhood of $x \in X$ if $x \in U$ and $\mathcal{T}$ is a Hausdorff topology if distinct points of $X$ have disjoint neighborhoods. A family $\mathcal{B} \subset \mathcal{T}$ is a base for $(X, \mathcal{T})$ if any non-empty open subset of $X$ can be represented as the union of a subfamily of $\mathcal{B}$. A family of neighborhoods $\mathcal{B}(x)$ of $x$ is called a base for $(X, \mathcal{T})$ at $x$ if for any neighborhood $V$ of $x$ there exists $U \in \mathcal{B}(x)$ such that $U \subset V$. The topological space $(X, \mathcal{T})$ is first-countable if at every point $x \in X$ there exists a countable base and $(X, \mathcal{T})$ is second-countable if $(X, \mathcal{T})$ has a countable base.

The above inclusion $T^I \subset T^1$ will be proved using the open $\omega$-disks $\{V^\omega(\rho, \epsilon) \mid \epsilon \in (0, \infty]\}$ which are a base of $T^\omega$ at $\rho \in S$ (this follows from (14) and from general results on convergences). The fact that these $\omega$-disks are a base implies also the equivalence b) below. Equivalence a) is proved in Theorem 3.21.1 (it is equivalent to $C(T^\omega) = C^\omega$).

Theorem. For $\rho \in S$ and sequences $(\rho_i)_{i \in \mathbb{N}} \subset S$ we have

$$\lim_{i \to \infty} S^\omega(\rho, \rho_i) = 0 \iff \forall U \in T^\omega \text{ with } \rho \in U \text{ we have } \rho_i \in U \text{ for large } i$$

$$\iff \forall \epsilon \in (0, \infty] \text{ we have } \rho_i \in V^\omega(\rho, \epsilon) \text{ for large } i.$$
Remark 1.11. Csiszár’s work cited in §1.1 give us a first idea about infinite dimensional algebras. In our context of algebraic formalism, this 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 $l^2 := \{ x \in l^\infty \mid \sum_{i\in\mathbb{N}} |x_i| < \infty \}$ 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

$$
\mathbb{P}(N) = \{ p = (p_i)_{i\in\mathbb{N}} \in [0, 1]^\mathbb{N} \mid \sum_{i\in\mathbb{N}} p_i = 1 \}
$$

defined in (7). The trace is the linear functional $\text{tr} : l^1 \to \mathbb{C}, x \mapsto \sum_{i\in\mathbb{N}} x_i$ and we have the equality of

$$
\mathbb{P}(N) = S_{l^\infty} = \{ x \in l^1 \mid \text{tr}(x) = 1, x \geq 0 \}.
$$

The information neighborhoods (2) do not define a topology on $\mathbb{P}(N) = S_{l^\infty}$ hence the equivalence b) in the above Theorem is wrong for $\mathcal{A} = l^\infty$.

1.5 New results on the rI-closure of an exponential family

We use the rI-topology on the state space $S_A$ of a $C^*$-subalgebra $\mathcal{A}$ of $\text{Mat}(n, \mathbb{C})$ to maximize the von Neumann entropy under linear constraints, including previously unknown solutions. This result follows from extensions of the Pythagorean theorem and the projection theorem in information geometry, which we also present in this section. The complete projection theorem enables us to study local maximizers and non-commutative features of the entropy distance in §4.3 and §4.4.

Definition 1.12. The von Neumann entropy of $\rho \in S$ is

$$
S(\rho) := -\text{tr} \rho \log(\rho),
$$

the logarithm is defined by functional calculus in Remark 2.4.4. The free energy of $\theta \in \mathcal{A}_{sa}$ is

$$
F(\theta) = F_A(\theta) := \log \text{tr} \exp_A(\theta),
$$

the exponential function $\exp_A$ and $R_A(\theta) = \exp_A(\theta)/\text{tr} \exp_A(\theta)$ are introduced in Definition 1.4.

Maximization of the von Neumann entropy under linear constraints is a fundamental problem in quantum statistical mechanics, see e.g. [IO97, Ru99, Pe08, AN00]. Given self-adjoint matrices $u_1, \ldots, u_k \in \mathcal{A}_{sa}$ it is well-known that for suitably chosen $\xi = (\xi_1, \ldots, \xi_k) \in \mathbb{R}^k$ there exist inverse temperatures $\beta_1, \ldots, \beta_k \in \mathbb{R}$, such that

$$
\rho(\xi) = R_A(-\sum_{i=1}^k \beta_i u_i) \tag{17}
$$

5The probability simplex $\mathbb{P}(N) = S_{l^\infty}$ corresponds to the normal states on $l^\infty$ (see e.g. Theorem 2.4.21 in [Br87]) and the dual space of continuous linear maps $l^\infty \to \mathbb{C}$ is strictly larger than $S_{l^\infty}$. It can be represented by bounded additive measures which are not necessarily $\sigma$-additive (see e.g. p. 89 in [We00] and p. 296 in [DS58]).
uniquely maximizes the von Neumann entropy among all states $\rho \in \mathcal{S}_A$ with mean
values $\langle u_i, \rho \rangle = \xi_i$ for $i = 1, \ldots, k$. The inverse temperatures $\beta_1, \ldots, \beta_k \in \mathbb{R}$ can
be computed from the conditions, $j = 1, \ldots, k$
$$\frac{\partial}{\partial \beta_j} F_A(-\sum_{i=1}^k \beta_i u_i) = -\xi_j$$
and the von Neumann entropy is
$$S(\rho(\xi)) = F_A(-\sum_{i=1}^k \beta_i u_i) + \sum_{i=1}^k \beta_i \xi_i.$$ 
All invertible maximizers have the form (17) and they form the exponential family
$$\mathcal{E} := R_A(U)$$ of Gibbs ensembles for the vector space $U := \text{span}_\mathbb{R}(u_1, \ldots, u_k)$. But
there exist maximizers $\rho \in \mathcal{S}$ which are not invertible.

We show in Theorem 4.26 that the complete set of solutions is the rI-closure
$$\text{cl}^1(\mathcal{E}) = \{ \rho \in \mathcal{S}_A \mid \inf_{\sigma \in \mathcal{E}} S(\rho, \sigma) = 0 \}.$$ The explicit solutions depend on a lattice
$\mathcal{P}^U$ of orthogonal projections, which in principle can be computed by spectral anal-
ysis. We use the mean value map $m_{u_1, \ldots, u_k} : \rho \mapsto (\langle u_1, \rho \rangle, \ldots, \langle u_k, \rho \rangle)$ and the convex support
$\text{cs}_A(u_1, \ldots, u_k) = \{ m_{u_1, \ldots, u_k}(\rho) \mid \rho \in \mathcal{S}_A \} \subset \mathbb{R}^k$.

**Theorem.** For every mean value tuple $\xi \in \text{cs}_A(u_1, \ldots, u_k)$ there is a unique maxi-
imizer $\rho(\xi)$ of the von Neumann entropy among all states $\rho \in \mathcal{S}_A$ with mean values
$m_{u_1, \ldots, u_k}(\rho) = \xi$. There exists a unique projection $p \in \mathcal{P}^U$ and there exist inverse
temperatures $\beta_1, \ldots, \beta_k \in \mathbb{R}$ such that
$$(\frac{\partial}{\partial \beta_1}, \ldots, \frac{\partial}{\partial \beta_k}) F_{pA\rho}(-\sum_{i=1}^k \beta_i u_i, p) = -\xi.$$ 
For each solution $(\beta_1, \ldots, \beta_k)$ we have
$$\rho(\xi) = R_{pA\rho}(-\sum_{i=1}^k \beta_i u_i, p)$$
and $\rho(\xi)$ has the von Neumann entropy
$$F_{pA\rho}(-\sum_{i=1}^k \beta_i u_i, p) + \sum_{i=1}^k \beta_i \xi_i.$$ 

The map $\xi \to \rho(\xi)$ is real analytic for invertible solutions $\rho(\xi)$ but can be dis-
tinuous on the whole, if the algebra $\mathcal{A}$ is non-commutative, see Remark 4.22.3.
An example is the Staffelberg family. Its rI-closure is described in Example 1.13 and
drawn in Figure 1. This exponential family is drawn in Figure 4 together with its
mean value set $\mathbb{M}(U) = \pi_U(\mathcal{S}_A)$.

The above theorem will be proved by the following theorems which extend the Pythagorean theorem and the projection theorem in §1.3. Using notation from
Definition 1.4, we consider an exponential family $\mathcal{E}$ and the entropy distance $S_A \to
[0, \infty)$, $d_\mathcal{E}(\rho) = \inf_{\sigma \in \mathcal{E}} S(\rho, \sigma)$.

**Theorem** (Complete projection theorem). Let $\rho \in \mathcal{S}$ be an arbitrary state.

1. The rI-closure $\text{cl}^1(\mathcal{E})$ intersects $\rho + U^\perp$ in a unique state denoted by $\pi_\mathcal{E}(\rho)$.

2. The relative entropy $S(\rho, \cdot)$ has a unique local minimum on $\text{cl}^1(\mathcal{E})$ and the entropy
distance (12) equals $d_\mathcal{E}(\rho) = \min_{\sigma \in \text{cl}^1(\mathcal{E})} S(\rho, \sigma) = S(\rho, \pi_\mathcal{E}(\rho))$. 
The complete projection theorem is proved in Theorem 4.15. The state \( \pi_E(\rho) \) is called the rI-projection of \( \rho \) to \( E \). The theorem implies for \( \rho \in S \) and \( \tau \neq \pi_E(\rho) \) in the rI-closure \( \text{cl}^{\text{rI}}(E) \) that
\[
S(\rho, \tau) - d_E(\rho) > 0.
\]
This inequality is not obvious a priori since the infimum in \( d_E(\rho) \) is over \( E \) (and not over \( \text{cl}^{\text{rI}}(E) \)). The exact difference is proved in Theorem 4.25:

**Theorem** (Complete Pythagorean theorem). If \( \rho \in S_A \) and \( \sigma \in \text{cl}^{\text{rI}}(E) \), then
\[
S(\rho, \pi_E(\rho)) + S(\pi_E(\rho), \sigma) = S(\rho, \sigma).
\]

The main construction towards the complete projection theorem will be an extension \( \text{ext}(E) \) of the exponential family \( E \), defined in terms of the mean value set. We prove in a C*-subalgebra \( A \) of \( \text{Mat}(n, \mathbb{C}) \) that
\[
\text{cl}^{\text{rI}}(E) = \text{ext}(E) \subset E
\]
holds with the rI-closure \( \text{cl}^{\text{rI}}(E) \) and the norm closure \( E \). A strict inclusion is possible:

**Example 1.13.** The Staffelberg family \( E \) from Example 1.5 satisfies \( \text{cl}^{\text{rI}}(E) \subsetneq E \).

This exponential family is depicted in Figure 1: The norm closure \( E \) is the union of \( E \) with the circle about \( E \) (bold) and the closed upright segment (dashed). The rI-closure \( \text{cl}^{\text{rI}}(E) \) is strictly included in \( E \), the upright segment is missing except for its top end (bold point). See [KW11] for this analysis.

An analogue extension \( \text{ext}(B) \) of exponential families \( B \) of Borel probability measure on \( \mathbb{R}^d \) satisfies
\[
\text{cl}^{\text{rI}}(B) \subset B \subset \text{ext}(B),
\]
and contrasts (18). See the introduction and Lemma 6 in [CM05] for these statements. The rI-closure \( \text{cl}^{\text{rI}}(B) \) is defined using the rI-convergence (2), \( B \) is the closure in the variation distance and \( \text{ext}(B) \) is defined using the concept of convex core, which generalizes convex support (6). If \( A \cong \mathbb{C}^N \) is commutative, then
\[
\text{cl}^{\text{rI}}(E) = \text{ext}(E) = E \cong \text{cl}^{\text{rI}}(B) = \text{ext}(B) = B
\]
holds, the correspondence being Remark 1.2.2. Further conditions for a commutative algebra are discussed in §4.4.

## 2 Analysis on the state space of a matrix algebra

This section contains preliminary material mainly by citation from the literature with the exception of perturbation theoretic proofs in §2.3. In §2.4 we cite an algebraic formulation of the convex geometry of the state space and of its projections to linear subspaces.
2.1 Projections and functional calculus

We define functional calculus for normal matrices. This section is a bit technical because we have to work in subalgebras of Mat($n$, $\mathbb{C}$) not containing $I_n$ in order to treat conditional probability measures and their formal matrix analogs. The partial ordering on $A$, and its restriction to the lattice of projections will be needed in §2.3. It plays a major role in §2.4 and §4.2–§4.4. A general reference on lattice theory is [Bi73].

**Definition 2.1.** 1. A map $f : X \rightarrow Y$ 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 $(L, \leq)$ where the infimum $x \land y$ and supremum $x \lor y$ of each two elements $x, y \in L$ exist. A lattice isomorphism is a bijection between two lattices that preserves the lattice structure. A lattice $L$ is complete if for an arbitrary subset $S \subseteq L$ the infimum $\inf S$ and the supremum $\sup S$ exist. The least element $\inf L$ and the greatest element $\sup L$ in a complete lattice $L$ are improper elements of $L$, all other elements of $L$ are proper elements. An atom of a complete lattice $L$ is an element $x \in L$, $x \neq \inf L$, such that $y \leq x$ and $y \neq x$ implies $y = \inf L$ for all $y \in L$.

2. An element $p \in A$ is a projection if $p^* = p = p^2$. The projection lattice of the algebra $A$ is

$$P = P_A := \{ p \in A \mid p^2 = p^* = p \}.$$  \hspace{1cm} (20)

We use the partial ordering $\leq$ on $A$ defined by $a \preceq b$ for $a, b \in A$ if and only if $b - a$ is positive semi-definite. We use the partial order on the projection lattice $P$, which is the restriction of $\leq$. For every projection $p \in P$ the compressed algebra $pAp$ is defined.

3. Let $a \in A$ 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 $\{p_i\}_{i=1}^N \subset P_A$ be a family of non-zero projections such that for $i, j = 1, \ldots, N$ we have $p_ip_j = p_i\delta_{ij}$, where $\delta_{ij} = 0$ unless $i = j$ with $\delta_{ii} = 1$. If $\sum_{i=1}^N p_i = 1$ and

$$a = \sum_{i=1}^N c_i p_i,$$  \hspace{1cm} (21)

then the sum (21) is called spectral form of $a$ in $A$, $\{p_i\}_{i=1}^N$ is a spectral family for $a$ in $A$ and its members are spectral projections of $a$ in $A$. Let us denote the set of eigenvalues of $a$ by $\text{spec}(a) := \text{spec}_{\text{Mat}(n, \mathbb{C})}(a)$.

**Remark 2.2.** 1. It is a classical result of linear algebra, see e.g. §§79–80 in [Ha87], that a normal matrix $a \in \text{Mat}(n, \mathbb{C})$ has a unique spectral form $a = \sum_{\lambda \in \text{spec}(a)} \lambda p_\lambda$ in $a \in \text{Mat}(n, \mathbb{C})$. Moreover, there exist polynomials $\{f_\lambda\}_{\lambda \in \text{spec}(a)}$ in one variable and with complex coefficients, such that $p_\lambda = f_\lambda(a)$ for $\lambda \in \text{spec}(a)$.

2. Let $A \subset \text{Mat}(n, \mathbb{C})$ be a $C^*$-subalgebra of Mat($n$, $\mathbb{C}$) with identity 1 and $a \in A$ a normal matrix. If $a = \sum_{\lambda \in \text{spec}(a)} \lambda p_\lambda$ is the spectral form of $a$ in Mat($n$, $\mathbb{C}$) then it is easy to show that

$$a = \sum_{\lambda \in \text{spec}_A(a)} \lambda (1 p_\lambda).$$
is the unique spectral form of $a$ in $\mathcal{A}$. Either $\text{spec}_\mathcal{A}(a) = \text{spec}(a)$ or $\text{spec}_\mathcal{A}(a) \subseteq\text{spec}_\mathcal{A}(a) \cup \{0\} = \text{spec}(a)$. For all non-zero $\lambda \in \text{spec}_\mathcal{A}(a)$ we have $1p_\lambda = p_\lambda$.

Some special projections and functional calculus will be needed.

**Definition 2.3.** 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 $\lambda \in \text{spec}_\mathcal{A}(a)$. The support projection of $a$, also called support of $a$, is $s(a) := \sum_{\lambda \in \text{spec}_\mathcal{A}(a) \setminus \{0\}} p^\lambda(a)$. The kernel projection of $a$ in $\mathcal{A}$ is $k_\mathcal{A}(a) := 1 - s(a)$.

2. If $a$ is self-adjoint, then the maximum of $\text{spec}_\mathcal{A}(a)$ 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 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 $f(a) = f_\mathcal{A}(a) := \sum_{\lambda \in \text{spec}_\mathcal{A}(a)} f(\lambda)p^\lambda_\mathcal{A}(a)$ is said to be defined by functional calculus in $\mathcal{A}$. If $p \in \mathcal{P}$ we abbreviate functional calculus in $\mathcal{P}$ $\mathcal{A}$ by $f(p)(a) = f_\mathcal{P}(p)(a) := f_p(a)$ provided that the complex valued function $f$ is defined on the spectrum $\text{spec}_{\mathcal{P}}(a)$ of a normal matrix $a \in \mathcal{P}$.

**Remark 2.4.** 1. If $a \in \mathcal{A}$ is a normal matrix and $f : \mathbb{C} \rightarrow \mathbb{C}$ is defined on $\text{spec}_\mathcal{A}(a)$ and on $\text{spec}_{\text{Mat}(n, \mathbb{C})}(a)$, then we have $f_\mathcal{A}(a) = f_{\text{Mat}(n, \mathbb{C})}(a)$. For example, we get for $a, u \in \mathcal{A}$ (the integral is defined in components of matrix entries)

$$\frac{\partial}{\partial t}|_{t=0} \exp_\mathcal{A}(a + tu) = \int_0^1 \exp_\mathcal{A}((1-y)a)u \exp_\mathcal{A}(ya)dy; \tag{22}$$

by multiplication with $1 \in \mathcal{A}$ from the analogue equation in $\text{Mat}(n, \mathbb{C})$. The latter can be proved by polynomial expansion $[\text{Li73}]$.

2. By Remark 2.2.2, the support projection $s(a)$ does not depend on the algebra $\mathcal{A} \subset \text{Mat}(n, \mathbb{C})$ that contains a normal matrix $a$. But $k_\mathcal{A}(a)$ and $p^+_{\mathcal{A}}(a)$ (if $a$ is self-adjoint) do depend on $\mathcal{A}$.

3. The term $\log^{[(1,0)]}(1, 0) = (0, 0)$ is an example of functional calculus in the compressed algebra $\mathbb{C} \oplus \{0\}$ of $\mathbb{C}^2$ while $\log(1, 0)$ is undefined.

4. The relative entropy introduced in Definition 1.3 is understood as the function such that for $\rho, \sigma \in \mathcal{S}$ we have $S(\rho, \sigma) := \infty$ unless $s(\rho) \preceq s(\sigma)$ where

$$S(\rho, \sigma) := \text{tr} \rho (\log^{s(\rho)}(\rho) - \log^{s(\sigma)}(\sigma)).$$

Similarly, the von Neumann entropy of $\rho \in \mathcal{S}$ introduced in Definition 1.12 is $S(\rho) := -\text{tr} \rho \log^{s(\rho)}(\rho)$. By part 2 these definitions restrict from $\text{Mat}(n, \mathbb{C})$ to any $\text{C}^*$-subalgebra of $\text{Mat}(n, \mathbb{C})$.

5. The projection lattice $\mathcal{P}$ with the partial ordering $\preceq$ is a complete lattice. For this and the following two statements see e.g. Remark 2.6 in [We11] or [AS01]. For a self-adjoint matrix $a \in \mathcal{A}_{sa}$ we have

$$a = pap \iff pa = a \iff s(a) \preceq p. \tag{23}$$

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

2.2 Continuity and convexity of the relative entropy

We recall convexity and continuity properties (in the norm topology) of the relative entropy, introduced in Definition 1.3.

**Definition 2.5.** 1. A function \( f : X \to (-\infty, \infty] \) defined on a convex subset \( X \) of a finite-dimensional Euclidean vector space \( E \) is **convex** if for \( x, y \in X \) and \( \lambda \in [0, 1] \) we have

\[
  f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y).
\]

A finite function \( f : X \to \mathbb{R} \) is **strictly convex** if for \( x, y \in X, x \neq y \) and \( \lambda \in (0, 1) \) we have

\[
  f((1 - \lambda)x + \lambda y) < (1 - \lambda)f(x) + \lambda f(y).
\]

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

\[
  \liminf_{i \to \infty} f(x_i) \geq f(x).
\]

3. If \((X, d)\) is a metric space and \( f : X \to (-\infty, \infty] \) then \( f \) is **lower continuous** if for all \( x \in X \) we have

\[
  \lim_{\epsilon \downarrow 0} \inf \{f(y) \mid d(x, y) < \epsilon\} = f(x).
\]

**Remark 2.6.** 1. The relative entropy \( S : \mathcal{S} \times \mathcal{S} \to [0, \infty] \) is a (norm) lower semi-continuous and convex function, see e.g. §III.B in [We78]. Under these assumptions the Corollary of Lemma 17.4 in [Če82] proves that the relative entropy is (norm) lower continuous on \( \mathcal{S} \times \mathcal{S} \).

2. The relative entropy is discontinuous in the norm topology in its first argument already for the algebra \( \mathcal{A} = \mathbb{C}^2 \) of a bit and in the second argument for the algebra \( \mathcal{A} = \text{Mat}(2, \mathbb{C}) \) of a qbit, see Example 3.19. However, in Proposition 3.18 we show that the relative entropy is continuous in the I-topology in its first argument and continuous in the rI-topology in its second argument.

The lower continuity of the relative entropy will be used in Theorem 3.21.5 to study the rI-topology.

2.3 Two perturbative statements

We provide arguments from the perturbation theory of \( \text{Mat}(n, \mathbb{C}) \), which will be used to characterize the rI-convergence and to study exponential families.

We denote the set of eigenvalues of \( a \in \text{Mat}(n, \mathbb{C}) \) by \( \text{spec}(a) = \text{spec}_{\text{Mat}(n, \mathbb{C})}(a) \) and we shall write \( \zeta \) in place of \( \zeta \mathbb{I}_n \) for scalars \( \zeta \in \mathbb{C} \).

**Definition 2.7.** 1. The **resolvent set** of a matrix \( a \in \text{Mat}(n, \mathbb{C}) \) is the complement of the spectrum \( \text{res}(a) := \mathbb{C} \setminus \text{spec}(a) \).
2. The resolvent of \( a \in \text{Mat}(n, \mathbb{C}) \) is defined for \( \zeta \in \text{res}(a) \) by \( (a - \zeta)^{-1} \).

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

\[
(a - \zeta)^{-1} - (b - \zeta)^{-1} = (a - \zeta)^{-1}(b - a)(b - \zeta)^{-1}.
\]

**Remark 2.8.** 1. If \( a, b \in \text{Mat}(n, \mathbb{C}) \) are self-adjoint matrices, let \( \lambda_1^a, \ldots, \lambda_n^a \) denote the eigenvalues of \( a \) arranged in decreasing order and counting multiplicities. Weyl’s perturbation theorem, proved e.g. in §III.2 in [Bl97], states that

\[
\max_{k=1}^{n} |\lambda_k^a - \lambda_k^b| \leq \|a - b\|.
\]

Here the spectral norm from Definition 1.1.2 is used.

2. According to Problem 5.7 on page 40 in [Ka95], if \( \zeta \) belongs to \( \text{res}(a) \) for a normal matrix \( a \in \text{Mat}(n, \mathbb{C}) \), then the resolvent of \( a \) is bounded by

\[
\|(a - \zeta)^{-1}\| \leq \text{dist}(\zeta, \text{spec}(a))^{-1}
\]

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

3. Given a normal matrix \( a \in \text{Mat}(n, \mathbb{C}) \), let \( \Gamma \subset \text{res}(a) \) be a positively oriented circular curve of radius \( r > 0 \). It is well-known, see e.g. Chapter 2 §1.4 in [Ka95], that

\[
P_{\Gamma}(a) := -\frac{1}{2\pi i} \int_{\Gamma} (a - \zeta)^{-1} d\zeta
\]

is the sum of all spectral projections \( p_{\lambda}(a) \) of \( a \) in \( \text{Mat}(n, \mathbb{C}) \), such that \( \lambda \) lies inside \( \Gamma \).

4. Let \( a, b \in \text{Mat}(n, \mathbb{C}) \) be self-adjoint matrices and let \( \Gamma_\lambda \) be disjoint circular curves of radius \( r > 0 \) centered at \( \lambda \in \text{spec}(b) \). If \( \|b - a\| < r \), then by Weyl’s perturbation theorem (25) every eigenvalue of \( a \) lies in exactly one of the circles \( \{\Gamma_\lambda\}_{\lambda \in \text{spec}(b)} \). The projections \( Q_{\lambda}(a) := P_{\Gamma_\lambda}(a) \) in (27) are defined and

\[I_n = \sum_{\lambda} Q_{\lambda}(a) \] holds (with summation over the eigenvalues \( \lambda \in \text{spec}(b) \)). The second resolvent equation (24) and the inequality (26) imply for \( \lambda \in \text{spec}(b) \)

\[
\|Q_{\lambda}(a) - p_{\lambda}(b)\| \leq \frac{1}{2\pi} \int_{\Gamma_\lambda} \|(b - \zeta)^{-1}(b - a)(a - \zeta)^{-1}\| d\zeta \leq \frac{\|b - a\|}{r(r - \|b - a\|)}.
\]

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 3.18.

**Lemma 2.9.** Let \( \rho, \sigma \in S \) and \( (\tau_i)_{i \in \mathbb{N}} \subset S \) 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-Csiszar inequality (13) the sequence \( (\tau_i)_{i \in \mathbb{N}} \) converges to \( \sigma \) in norm. We view \( \tau_i \) as a perturbation of \( \sigma \) and take a sufficiently small circle \( \Gamma \) of radius \( r > 0 \) about \( 0 \in \mathbb{C} \). Then, for large \( i \in \mathbb{N} \) the projection \( P_{\Gamma}(\tau_i) \) in (27) is defined and satisfies \( k(\tau_i) \preceq P_{\Gamma}(\tau_i) \) where \( k(\tau_i) = k_{\text{Mat}(n, \mathbb{C})}(\tau_i) \) is the kernel projection. Then two projections \( p_i, q_i \in A \) are defined by \( p_i := P_{\Gamma}(\tau_i) - k(\tau_i) \) and
$q_i := 1_n - P_{\Gamma}(\tau_i)$, 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 2.3.3 we split the functional calculus into two compressed algebras $p_iA p_i$ and $q_iA q_i$,

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

We have $\tau_j \xrightarrow{j \to \infty} \sigma$, by (28) we have $q_j \tau_j q_j$ and the spectral values of $q_j \tau_j q_j$ in $q_jA q_j$ are strictly larger than $r > 0$, hence the term $\log^{[q]}(q_i \tau_i) \xrightarrow{\infty} \log^{|s(\sigma)|}(\sigma)$ converges. Using the assumption $S(\sigma, \tau_i) \xrightarrow{i \to \infty} 0$ gives $\lim_{i \to \infty} \text{tr} \sigma \log^{[p]}(p_i \tau_i) = 0$.

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 $\sigma$ then $\lambda s(\sigma) \preceq \sigma$. Hence $0 \preceq \frac{1}{\lambda^+ (\rho)} \rho \preceq \sigma$. For all $i \in \mathbb{N}$ we have $\lim_{i \to \infty} \text{tr} \rho \log^{[p]}(p_i \tau_i) \leq 0$ hence

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

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

$$S(\rho, \tau_i) = -S(\rho) - \text{tr} \rho \log^{[p]}(p_i \tau_i) - \text{tr} \rho \log^{[q]}(q_i \tau_i) \xrightarrow{i \to \infty} -S(\rho) - 0 - \text{tr} \rho \log^{[s(\sigma)]}(\sigma) = S(\rho, \sigma)$$

completes the proof. \hfill $\square$

The following statement is used in Proposition 4.3 to set up the mean value chart of an exponential family and in Lemma 4.12 to study rI-closures of exponential families. Part 1 is used implicitly in Lemma 7 in [Wi63].

**Lemma 2.10.** 1. Let $(x_j)_{j \in \mathbb{N}} \subset A_{sa} \setminus \{0\}$ such that $\lim_{j \to \infty} \|x_j\| = \infty$. We assume there exist $u, a \in A_{sa}$ such that $\lim_{j \to \infty} \frac{x_j}{\|x_j\|} = u$ and $\lim_{j \to \infty} \exp_A(x_j) = a$. Then $\text{spec}_{A}(u) \subset (-\infty, 0]$ and $s(a) \preceq k_A(u)$.

2. Let $\theta, u \in A_{sa}$ such that $\text{spec}_{A}(u) \subset (-\infty, 0]$. Then

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

**Proof:** 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 $e^{\lambda y_j}$ in suitable compressed subalgebras. We choose disjoint circular curves $\Gamma_{\lambda}$ in the complex plane about the eigenvalues $\lambda \in \text{spec}(u)$. Using Weyl’s perturbation theorem (25), the projections in (27)

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

are defined for large $j$. Let $\lambda \in \text{spec}(u)$ and $\lambda \neq 0$. The projection $Q^\lambda(x_j)$ is a sum of spectral projections of $x_j$ in Mat$(n, \mathbb{C})$ for non-zero eigenvalues of $x_j$, so $Q^\lambda(x_j) \in A$ by Remark 2.2.2. We consider functional calculus in the compressed algebra $Q^\lambda(x_j)AQ^\lambda(x_j)$,

$$h^\lambda(x_j) := \exp_A^{[Q^\lambda(x_j)]}(Q^\lambda(x_j)x_j) = Q^\lambda(x_j) \exp(x_j).$$
The spectral values of the self-adjoint matrix $Q^\lambda(x_j)y_j$ in $Q^\lambda(x_j)AQ^\lambda(x_j)$ converge for $j \to \infty$ to $\lambda \neq 0$ because there is only one eigenvalue of $u$ in the circle $\Gamma_\lambda$. 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
\[
\|x_j\| \xrightarrow{j \to \infty} \infty \quad \text{implies} \quad h^\lambda(x_j) \xrightarrow{j \to \infty} 0. \tag{29}
\]
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 (28). Hence with summation over $\lambda \in \text{spec}(u) \setminus \{0\}$ we have $s(u) = \lim_{j \to \infty} \sum_{\lambda \neq 0} Q^\lambda(x_j)$. Now the assumed convergence of $\exp_A(x_j) \xrightarrow{j \to \infty} a$ 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 $\text{spec}(u) \subset (-\infty, 0]$. Then (23) and the equation
\[
k_{\mathcal{A}}(u)a = (\mathbbm{1} - s(u))a = a
\]
show $s(a) \leq k_{\mathcal{A}}(u)$.

We prove convergence and calculate the limit in the second statement. For small real parameter $c > 0$ let $x_c := u + c\theta$, then $x_c \xrightarrow{c \to 0} u$. For $\lambda \in \text{spec}(u) \cup \{0\}$ we choose disjoint circular curves $\Gamma_\lambda$ in the complex plane about each such $\lambda$ and we define
\[
Q^\lambda(x_c) := P_{\Gamma_\lambda}(x_c).
\]
For all $\lambda < 0$ the argument in (29) shows $Q^\lambda(x_c) \exp(x_c) \xrightarrow{c \to 0} 0$. Since $\mathbbm{1}_n = \sum_{\lambda \in \text{spec}(u) \cup \{0\}} Q^\lambda(x_c)$ holds for large $j$ we have
\[
\lim_{t \to +\infty} \exp(\theta + tu) = \lim_{c \to 0} \exp\left(\frac{1}{c}x_c\right) = \lim_{c \to 0} Q^0(x_c) \exp(Q^0(x_c)\frac{1}{c}x_c).
\]
By (28) we have $Q^0(x_c) \xrightarrow{c \to 0} k(u) \in \text{Mat}(n, \mathbb{C})$. The first order expansion is calculated in Chapter II §1 equation (1.17) in [Ka95]: With $\tilde{Q} := \frac{1}{2\pi i} \int_{\Gamma_0} (u - \zeta)^{-1}\theta(u - \zeta)^{-1}d\zeta$ we have$^6$
\[
Q^0(x_c) = k(u) + c\tilde{Q} + o(c).
\]
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
\[
\lim_{t \to +\infty} \exp(\theta + tu) = \lim_{c \to 0} \exp(Q^0(x_c)\frac{1}{c}x_c) = k(u) \exp(k(u)\theta k(u)).
\]
Multiplication of this formula with the identity $\mathbbm{1}$ of $\mathcal{A}$ completes the proof. \hfill \Box

$^6$If $g$ is a positive function and $f$ is any function (here with values in $\text{Mat}(n, \mathbb{C})$), then $f = o(g)$ means $\frac{f}{g} \to 0$ and $o(g)$ is called Landau symbol.
2.4 Lattices of faces and projections

In this section we settle definitions of convex geometry. For convenience we provide a selected overview of the algebraic description of the convex geometry of the mean value set \( \mathbb{M}(U) = \pi_U(S) \), defined in (4) as the orthogonal projection of the state space \( S \) onto a linear subspace \( U \subset \mathcal{A}_\text{sa} \).

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 2.11.** 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. If an extreme point of \( C \) defines an exposed face then it is an exposed point. Otherwise the extreme point is a non-exposed point.

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

**Remark 2.12.** 1. As observed e.g. in [KW11, We12], the mean value set \( \mathbb{M}(U) \) can have non-exposed faces even though all faces of \( S \) are exposed. An example is shown in Figure 2.
Figure 2: This clove shape is the mean value set of the *Swallow family*. The supporting hyperplane to the left defines a one-dimensional exposed face. Two non-exposed points are indicated by small circles. The supporting hyperplane to the right defines an exposed point.

2. Let \( C \subset E \) be a convex subset. Different to Rockafellar or Schneider [Ro72, Sc93] we always include \( \emptyset \) and \( C \) to \( F(C) \) so that this set is a lattice. The inclusion \( F_\perp(C) \subset F(C) \) is easy to show and there are various ways to see that \( F_\perp(C) \) and \( F(C) \) are complete lattices ordered by inclusion where the infimum is the intersection, see e.g. §1.1 in [We12] or §2.1 in [We11]. The convex set \( C \) admits by Theorem 18.2 in [Ro72] a partition into relative interiors of its faces

\[
C = \bigcup_{F \in F(C)} \text{ri}(F) .
\]  

(30)

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 \( F(S_A) = F_\perp(S_A) \) of the state space \( S_A \).

**Definition 2.13.** Extreme points of \( S \) are called pure states. For every orthogonal projection \( p \in \mathcal{P}_A \) we set

\[
F(p) = F_A(p) := S_{pA^p}
\]

and we denote the face lattice of the state space by \( F = F_A := F(S_A) \).

**Proposition 2.14** (Proposition 2.9 in [We11]). The state space \( S \) is a convex body of dimension \( \dim(A_{sa}) - 1 \), the affine hull is \( \text{aff}(S) = A_1 \), the translation vector space is \( \text{lin}(S) = A_0 \) and the relative interior consists of all invertible states. The support function at \( a \in A_{sa} \) is the maximal spectral value \( h(S, a) = \lambda^+(a) \) of \( a \). If \( a \in A_{sa} \) is non-zero, then the exposed face of \( a \) is the state space \( F_\perp(S, a) = F(p) \) of the compressed algebra \( pA^p \), where \( p = p^+(a) \) is the maximal projection of \( a \).

**Corollary 2.15** (Corollary 2.10 in [We11]). All faces of the state space \( S \) are exposed. The mapping \( F : \mathcal{P} \to F, p \mapsto F(p) \) is an isomorphism of complete lattices.

**Remark 2.16.** It follows from Corollary 2.15, Proposition 2.14 and (23) that every face of \( S \) can be written as \( F(p) = \{ \rho \in S \mid s(\rho) \preceq p \} \) for some \( p \in \mathcal{P} \) and the relative interior is \( \text{ri}F(p) = \{ \rho \in S \mid s(\rho) = p \} \).

Let us turn to the mean value set \( M(U) = \pi_U(S) \) defined in (4), where \( U \subset A_{sa} \) is a linear subspace. A lifting construction connects to the isomorphism \( F : \mathcal{P} \to F \). This leads to algebraic descriptions of the two face lattices \( F_\perp(M(U)) \subset F(M(U)) \).
Definition 2.17. We define for subsets \( C \subset A \) the (set-valued) lift by
\[
L^U(C) = L^U_A(C) := S_A \cap (C + U^\perp).
\]

We define the \textit{lifted face lattice}
\[
\mathcal{L}^U = \mathcal{L}^U_A := \{ L^U(F) \mid F \in \mathcal{F}(\mathcal{M}(U)) \}
\]
and the \textit{lifted exposed face lattice}
\[
\mathcal{L}^{U,\perp} = \mathcal{L}^{U,\perp}_A := \{ L^U(F) \mid F \in \mathcal{F}_\perp(\mathcal{M}(U)) \}.
\]

Lemma 2.18 (§5 in [We12]). The lift \( L \) restricts to the bijection \( \mathcal{F}(\mathcal{M}(U)) \xrightarrow{L} \mathcal{L}^U \) and to the bijection \( \mathcal{F}_\perp(\mathcal{M}(U)) \xrightarrow{L} \mathcal{L}^{U,\perp} \). These are isomorphisms of complete lattices with inverse \( \pi_U \). For \( u \in U \) we have \( \pi_U[F_\perp(S,u)] = F_\perp(\mathcal{M}(U),u) \) and \( L^U[F_\perp(\mathcal{M}(U),u)] = F_\perp(S,u) \).

The results give rise to useful lattice isomorphisms, if we use appropriately defined lattices of projections.

Definition 2.19. The \textit{projection lattice} resp. \textit{exposed projection lattice} of \( U \) is
\[
P^U = P^U_A := \mathbb{F}^{-1}(\mathcal{L}^U_A) \quad \text{resp.} \quad P^{U,\perp} = P^{U,\perp}_A := \mathbb{F}^{-1}(\mathcal{L}^{U,\perp}_A).
\]

Corollary 2.15 and Lemma 2.18 imply two lattice isomorphisms defined for suitable projections \( p \) by \( p \mapsto \pi_U(\mathbb{F}(p)) \):
\[
P^U \rightarrow \mathcal{F}(\mathcal{M}(U)) \quad \text{resp.} \quad P^{U,\perp} \rightarrow \mathcal{F}_\perp(\mathcal{M}(U)) \quad (32)
\]

between \( P^U \) and the face lattice of the mean value set resp. between \( P^{U,\perp} \) and the exposed face lattice. Lemma 2.18 characterizes the lifted exposed face lattice by
\[
\mathcal{L}^{U,\perp} = \{ F_\perp(S,u) \mid u \in U \} \cup \{\emptyset\}.
\]

The algebraic description in Proposition 2.14 of faces \( F_\perp(S,u) \) of the state space \( S \) translates therefore to the exposed faces of the mean value set \( \mathcal{M}(U) \):

Corollary 2.20. The exposed projection lattice is
\[
P^{U,\perp}_A = \{ p^+_A(u) \mid u \in U \} \cup \{0\}.
\]

In order to understand the non-exposed faces of \( \mathcal{M}(U) \) algebraically, we have to look at sequences of faces.

Definition 2.21. 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 \( F_0, \ldots, F_m \subset C \) an \textit{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 \( i = 0, \ldots, m - 1 \),
\[
F_0 \supseteq F_1 \supseteq \cdots \supseteq F_m.
\]

2. For \( p \in \mathcal{P} \) and \( a \in A_{sa} \) the orthogonal projection \( A_{sa} \rightarrow (pA)p_{sa} \) is
\[
c^p(a) := \pi_{(pAp)_{sa}}(a) = pap.
\]
3. We call a finite sequence $p_0, \ldots, p_m \subset \mathcal{P}^U$ an access sequence (of projections) for $U$ if $p_0 = 1$ and if $p_{i+1}$ belongs to the exposed projection lattice $\mathcal{P}^{P^i(U), \perp}$ for $i = 0, \ldots, m - 1$ and such that $(p_i \succ p_{i+1} : \iff p_i \succeq p_{i+1}$ and $p_i \neq p_{i+1})$

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

Grüenbaum [Gr03] defines a poonem as an element of an access sequence of faces. An example is depicted in Figure 3. In finite dimensions the notion of poonem is equivalent to the notion of face, see e.g. §1.2.1 in [We12].

**Theorem 2.22** (§3.2 in [We11]). The lattice isomorphism $\mathcal{P}^U \rightarrow \mathcal{F}(\mathbb{M}(U))$ in (32) 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

$$(p_0, \ldots, p_m) \mapsto (\pi_U(\mathcal{F}(p_0)), \ldots, \pi_U(\mathcal{F}(p_m))).$$

For convenience we cite further result from §3.2 in [We11].

**Lemma 2.23.** If $p \in \mathcal{P}$ is a projection, then $\mathcal{c}^p(U) \xrightarrow[\pi_U]{} \pi_U(\mathcal{F}(p))$ is a real linear isomorphism and the following diagrams commute.

$$
\begin{array}{ccc}
(pAp)_{sa} & \xrightarrow{\pi_U} & \pi_U((pAp)_{sa}) \\
\pi_{cP(U)} & \downarrow{\pi_U} & \pi_{cP(U)} \\
\mathcal{c}^p(U) & \xrightarrow{\pi_U} & \mathcal{c}^p(U)
\end{array}
\quad
\begin{array}{ccc}
\mathcal{F}(p) & \xrightarrow{\pi_U} & \pi_U(\mathcal{F}(p)) \\
\pi_{cP(U)} & \downarrow{\pi_U} & \pi_{cP(U)} \\
\mathbb{M}_{pAp}(\mathcal{c}^p(U)) & \xrightarrow{\pi_U} & \mathbb{M}_{pAp}(\mathcal{c}^p(U))
\end{array}
$$

**Corollary 2.24.** 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 2.25.** 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 2.26.** Corollary 2.24 implies a computation method for $\mathcal{P}^U$: One has to compute the maximal projection (see Definition 2.3.2) of all elements of $U$, then the maximal projections of elements of $\mathcal{c}^p(U)$ for each previously calculated projection $p$ and so on (see Remark 3.10 and §3.3 in [We11]).

**Lemma 2.27.** If $\rho \in \mathcal{S}$, then $\rho \in \text{ri}(\mathcal{F}(p)) + U^\perp$ holds for a unique projection $p \in \mathcal{P}^U$. We have $p = \bigwedge\{q \in \mathcal{P}^U \mid s(p) \preceq q\}$. 

Figure 3: A poonem constructed by repeated inclusion of exposed faces.
3 Information topologies on the state space

We study the \(1/\text{r}\)-topology on the state space of a \(C^*\)-subalgebra of \(\text{Mat}(n, \mathbb{C})\). Our analysis in §3.2 is based on the idea of divergence function and \(L^*\)-convergence that we recall and customize in §3.1. Unless otherwise specified we will use the norm topology.

3.1 The topology of a divergence function

We generalize the idea of metric space to the notion of divergence function on a set. A topology is associated in the abstract setting of an axiomatic notion of convergence of countable sequences. Conditions on the divergence function, available for the relative entropy on \(\text{Mat}(n, \mathbb{C})\), imply quite strong results. Let \(X\) be any set.

Definition 3.1 (\(L^*\)-convergence\). A relation \(C \subset X^N \times X\) 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 \((x_n)_{n \in \mathbb{N}}\). The convergence \(C\) is a sequential convergence on \(X\), if

\[\begin{align*}
\text{a) } & \text{ } x_n = x \text{ for all } n \text{ implies } x_n \xrightarrow{C} x, \\
\text{b) } & \text{ } x_n \xrightarrow{C} x \text{ and } (y_n)_{n \in \mathbb{N}} \text{ is a subsequence of } (x_n)_{n \in \mathbb{N}} \text{ then } y_n \xrightarrow{C} x.
\end{align*}\]

If \(C\) is a sequential convergence on \(X\), then \((X, C)\) is a sequential space. A sequential convergence \(C\) on \(X\) is an \(L\)-convergence if

\[\begin{align*}
\text{c) } & \text{ } x_n \xrightarrow{C} x \text{ and } x_n \xrightarrow{C} y \text{ implies } x = y.
\end{align*}\]

The \(L\)-convergence \(C\) on \(X\) is an \(L^*\)-convergence and \((X, C)\) is an \(L^*\)-space if

\[\begin{align*}
\text{d) } & \text{ } x_n \xrightarrow{C} x \text{ (i.e. it is false that } x_n \xrightarrow{C} x) \text{ implies the existence of a subsequence } (y_n)_{n \in \mathbb{N}} \text{ of } (x_n)_{n \in \mathbb{N}}, \text{ such that for any subsequence } (z_n)_{n \in \mathbb{N}} \text{ of } (y_n)_{n \in \mathbb{N}} \text{ we have } z_n \xrightarrow{C} x.
\end{align*}\]

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

Remark 3.2 (The topology of a convergence). It is well-known [Du64] 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 \(y \in Y\).

The important example of a metric space will be generalized in Definition 3.14.

Example 3.3 (Metric spaces). Let \((X, d)\) be a metric space for \(d : X \times X \to \mathbb{R}\). Then \(x_n \xrightarrow{C_d} x : \iff \lim_{i \to \infty} d(x, x_i) = 0\) defines an \(L^*\)-convergence \(C_d\) on \(X\) and the disks \(B(x, \epsilon) := \{y \in X \mid d(x, y) < \epsilon\}\) for \(\epsilon > 0\) define a base for \(\mathcal{T}(C_d)\) at \(x \in X\). The topology \(\mathcal{T}(C_d)\) is known as the metric topology.

\text{\footnote{The following definition of an }L^*\text{-space is used in [En89, Du64]. An }L^*\text{-space in the sense of [Be63] uses only the axioms a), b) and d).}}
We can also go the opposite way from a topology to a convergence.

**Definition 3.4** (The convergence of a topology). If \((X, T)\) is a topological space then the convergence \(C(T)\) is defined for sequences \((x_i)_{i \in \mathbb{N}} \subset X\) and \(x \in X\) by

\[
(x_i)_{i \in \mathbb{N}} \xrightarrow{C(T)} x \iff \text{if } x \in U \in T \text{ then } x_i \in U \text{ for large } i.
\]

For any topological space \((X, T)\) the reader should check that \(T \subset T(C(T))\) holds. Similarly, if \(C\) is a convergence on \(X\), then \(C \subset C(T(C))\) holds. An equality condition was proved by Kisyński, see e.g. Theorem 2.1 in [Du64]:

**Theorem 3.5.** If \((X, C)\) is an \(L^*\)-space, then \(C(T(C)) = C\).

Continuity can be expressed in terms of convergence or of topology.

**Definition 3.6.** Let \(f : X \to X'\) be a function between arbitrary sets. If \((X, C)\), \((X', C')\) are sequential spaces, 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 \(T\) resp. \(T'\) is a topology on \(X\) resp. \(X'\), then \(f\) is continuous for \(T\) and \(T'\) if \(f^{-1}(U')\) is \(T\) open for every \(T'\) open set \(U' \subset X'\).

The following proposition is an excerpt of Theorem 2.2 in [Du64].

**Theorem 3.7.** Let \((X, C)\) and \((X', C')\) be sequential spaces and \(f : X \to X'\).

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

We need to study subspaces in some detail.

**Definition 3.8.** Let \(B \subset X\). If \(T\) is a topology on \(X\), then the **subspace topology**

\[
T|_B := \{B \cap U \mid U \in T\}
\]

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

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

**Remark 3.9** (Subspaces). If \(C\) is a convergence on \(X\) and \(B \subset X\), then the inclusion \(T(C)|_B \subset T(C|_B)\) holds (this is easy to prove). If \((X, d)\) is a metric space with convergence \(C_d\) defined in Example 3.3, then \(T(C_d)|_B = T(C_d|_B)\) holds for arbitrary subsets \(B \subset X\). This follows from the fact that the open disks \(B(x, \epsilon)\) for \(\epsilon > 0\) are a base at \(x \in X\). We will generalize this idea in Lemma 3.16.2.

We consider closures in a sequential space.
Definition 3.10. Let \((X, C)\) be a sequential space. The \textit{sequential closure} of \(Y \subset X\) is
\[
\cl_C(Y) := \{x \in X \mid (x_n)_{n \in \N} \xrightarrow{C} x \text{ for a sequence } (x_n)_{n \in \N} \subset Y\}. \tag{35}
\]
The following property is suggested in [Du64]:
\begin{itemize}
  \item[e)] \(x_n \xrightarrow{C} x\) and \((x^{(m)})_n \xrightarrow{C} x_m\) for all \(m \in \N\) implies that there exists a function \(n : \N \to \N\), such that \((x^{(m)})_{n(m)} \xrightarrow{C} x\).
\end{itemize}

A weaker property is proposed in Problem 1.7.18 in [En89]:
\begin{itemize}
  \item[e')] if \(x_n \xrightarrow{C} x\) and for \(n \in \N\) we have \((x^{(n)})_m \xrightarrow{C} x_n\), then there exist sequences of positive integers \(n_1, n_2, \ldots\) and \(m_1, m_2, \ldots\), such that \((x^{(m_k)})_{m_k} \xrightarrow{C} x\).
\end{itemize}

The next examples show that sequential closures in \(L^*\)-spaces of probability measures need not be topological closures.

Example 3.11. The I-/rI-convergence of probability measures in (2) is an \(L^*\)-convergence, see [Ha07, Du98]. Harremoës gives the example of a triangle \(D\) in \(\mathbb{P}(\N)\), the probability simplex (7) of \(\N\), where \(\cl_C(D) \subset \cl_C(\cl_C(D))\) holds for the I-convergence \(C\). Csiszár and Matúš [CM04] discuss an exponential family \(\mathcal{E}\) of Borel probability measures in \(\mathbb{R}^3\) where \(\cl_C(\mathcal{E}) \subset \cl_C(\cl_C(\mathcal{E}))\) holds for the rI-convergence \(C\).

Topological spaces for which sequential closure is the same as ordinary closure are known as \textit{Fréchet spaces} [En89]. Sequential spaces with this property are characterized by e):

Remark 3.12. If \((X, C)\) is a sequential space satisfying e) in Definition 3.10 and \(Y \subset X\) then the sequential closure \(\cl_C(Y)\) is the \(\mathcal{T}(C)\) closure of \(Y\).

In detail, a subset \(Y \subset X\) is \(\mathcal{T}(C)\) closed if and only if \(\cl_C(Y) = Y\) by Remark 3.2. Hence \(\cl_C(Y)\) is the \(\mathcal{T}(C)\) closure of \(Y\) if and only if \(\cl_C(\cl_C(Y)) = \cl_C(Y)\). The latter equation being true for all \(Y \subset X\) is easily shown to be equivalent\(^8\) to e'). The argument is complete since e') follows from e).

We generalize metric spaces, one aspect is to allow infinite “distances”.

Example 3.13. The positive augmented half-line \([0, \infty] := [0, \infty) \cup \{\infty\}\) is considered a topological space with the \textit{Alexandroff compactification} \(\mathcal{T}^c\) of the positive half-line \([0, \infty)\). Open sets are of the form \([0, \infty) \setminus F\), where \(F\) is a norm compact subset of \([0, \infty)\), together with all norm open subsets of \([0, \infty)\). Then \(([0, \infty], \mathcal{T}^c)\) is a compact Hausdorff space, see e.g. Theorem 3.5.11 in [En89]. The convergence \(C^c := C(\mathcal{T}^c)\) clearly equals
\[
\{((x_i), x) \in [0, \infty]^\N \times [0, \infty) \mid x_i < \infty \text{ for large } i \text{ and } \lim_{i \to \infty} x_i = x\} \cup \{((x_i), \infty) \mid (x_i) \subset [0, \infty) \text{ such that } \forall R \in [0, \infty) \text{ we have } x_i \geq R \text{ for large } i\},
\]
where \(\lim_{i \to \infty} x_i = x\) in the first term means that the finite valued sequence members of \((x_i)_{i \in \N}\) converge to \(x\) in norm.

\(^8\)This equivalence holds also for convergences that are not sequential.
It is easy to show $T(C(T^c)) \subset T^c$ (every $U \in T(C(T^c))$ including $\infty$ 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

$$T(C^c) = T^c.$$ 

It is easy to show that $C^c$ is an L*-convergence, hence we obtain from Theorem 3.7.2 for any sequential space $(X, C)$ and any function $f : X \rightarrow [0, \infty]$ that $f$ is continuous for $C$ and $C^c$ if and only if $f$ is continuous for $T(C)$ and $T^c$.

We shall frequently write $\lim_{i \rightarrow \infty} x_i = x$ in place of $x_i \xrightarrow{C^c} x$ for $(x_i)_{i \in \mathbb{N}} \subset [0, \infty]$ and $x \in [0, \infty]$. Finally we study the example that will clear some questions about the I-/rI-topology.

**Definition 3.14 (Divergence functions).** 1. A divergence function on a set $X$ is a function $f : X \times X \rightarrow [0, \infty]$, such that for all $x \in X$ we have $f(x, x) = 0$. Let $C_f$ be the convergence on $X$ defined by

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

2. Two assumptions will suffice for our purposes to analyze the I-/rI-convergence:

A) $(X, d)$ is a metric space and there is a function $g : [0, \infty] \rightarrow [0, \infty]$ such that for all $x, y \in X$ we have $d(x, y) \leq g(f(x, y))$. The function $g$ is continuous for $C^c$ and $C^c$ at $0$ and $g(0) = 0$.

B) For all $x \in X$ the function $X \rightarrow [0, \infty], y \mapsto f(x, y)$ is continuous for $C_f$ and $C^c$.

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

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

and the closed $f$-disk

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

**Remark 3.15.** Property B) fails if $f$ is the relative entropy discussed in the introduction (3) on the probability space $(\mathbb{N}, 2^{\mathbb{N}})$. Property A) holds for that case due to the Pinsker-Csiszár inequality.

We study divergence functions satisfying A) or B).

**Lemma 3.16 (Divergence functions).** Let $f$ be a divergence function on a set $X$. The convergence $C_f$ is a sequential convergence with property d) in Definition 3.1. The sequential closure of $Y \subset X$ is

$$\text{cl}_{C_f}(Y) = \{x \in X \mid \lim_{n \rightarrow \infty} f(x, y_n) = 0 \text{ for a sequence } (y_n)_{n \in \mathbb{N}} \subset Y\}$$

$$= \{x \in X \mid \inf_{n \in \mathbb{N}} f(x, y_n) = 0 \text{ for a sequence } (y_n)_{n \in \mathbb{N}} \subset Y\}$$

$$= \{x \in X \mid \inf_{y \in Y} f(x, y) = 0\}.$$
1. Let $f$ satisfy property $A)$ in Definition 3.14.2 for the metric $d : X \times X \to \mathbb{R}$. Then $C_f$ is an $L^*$-convergence, in particular $C(T(C_f)) = C_f$. We have $C_f \subset C_d$ and $T(C_f) \supset T(C_d)$, in particular $T(C_f)$ is a Hausdorff topology.

2. Let $f$ satisfy property $B)$ in Definition 3.14.2. Then for all $x \in X$ the function $X \to [0, \infty], y \mapsto f(x, y)$ is continuous for $T(C_f)$ and $T^c$. For each $x \in X$ and $\epsilon \in (0, \infty]$ the open $f$-disk $V^f(x, \epsilon)$ is $T(C_f)$ open and the closed $f$-disk $W^f(x, \epsilon)$ is $T(C_f)$ closed. The open $f$-disks $\{V^f(x, \epsilon) \mid \epsilon > 0\}$ are a base for $(X, T(C_f))$ at $x$. In particular $T(C_f)$ is first countable and for any subset $Y \subset X$ we have $T(C_f)|_Y = T(C_f|_Y)$. The sequential convergence $C_f$ has property e) in Definition 3.14, in particular for any subset $Y \subset X$ the sequential closure $cl_{C_f}(Y)$ is the $T(C_f)$ closure of $Y$.

Proof: The statements in the preamble are clear, we show part 1. To prove that $C_f$ is an $L^*$-convergences, it suffices to show condition $c)$ in Definition 3.1. Let $x \in X$ and $(x_i)_{i \in \mathbb{N}} \subset X$. Assuming $x_n \overset{C_f}{\to} x$, 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 $d(x, x_i) \leq g \circ f(x, x_i)$. Hence $\lim_{i \to \infty} d(x, x_i) = 0$ or $x_n \overset{C_d}{\to} x$ likewise. We have proved $C_f \subset C_d$. If follows immediately that $T(C_f) \supset T(C_d)$ and since $T(C_d)$ is Hausdorff, so is $T(C_f)$.

If for a second point $y \in X$ we have $x_n \overset{C_f}{\to} y$, then we get

$$d(x, y) \leq \lim_{i \to \infty} d(x, x_i) + \lim_{i \to \infty} d(x_i, y) = 0.$$ 

This shows $x = y$ and proves that $C_f$ is an $L^*$-convergence. Now Theorem 3.5 shows $C(T(C_f)) = C_f$.

We prove part two. For all $x \in X$ the function $X \to [0, \infty], y \mapsto f(x, y)$ is continuous for $C_f$ and $C^c$ by assumption B). The discussion in the last paragraph of Example 3.13 shows that this function is continuous for $T(C_f)$ and $T^c$. Hence the preimage of every $T^c$ open resp. closed subset of $[0, \infty]$ is $T(C_f)$ open resp. closed. In particular, every open resp. closed $f$-disk is $T(C_f)$ open resp. closed. The open $f$-disks $\{V^f(x, \epsilon) \mid \epsilon > 0\}$ define a base for $(X, T(C_f))$ at $x \in X$: By contradiction, if $U$ is a $T(C_f)$ neighborhood of $x$ and $U$ contains no open $f$-disk about $x$, then there exists a sequence $(x_i)_{i \in \mathbb{N}} \subset X \setminus U$ with $(x_i)_{i \in \mathbb{N}} \overset{C_f}{\to} x$. But $X \setminus U$ is $T(C_f)$ closed and hence contains all $C_f$-limits of sequences in $X \setminus U$. So $x \in X \setminus U$ and $U$ is not a $T(C_f)$ neighborhood of $x$.

The space $(X, T(C_f))$ is first countable, e.g. $\{V^f(x, 1/n) \mid n \in \mathbb{N}\}$ is a base at $x \in X$. If $Y \subset X$, then $T(C)|_Y \subset T(C|_Y)$ holds, see Remark 3.9. 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 $U \in T(C|_Y)$ equals

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

where $I$ is the index set of basis elements $V^f(x, \epsilon)$. The set $U$ consists of all open $f$-disks in $Y$ with center $y_\alpha$ and radius $\epsilon_\alpha$.
for some \( y_\alpha \in Y \) and \( \epsilon_\alpha > 0, \alpha \in I \). We have proved \( U \in \mathcal{T}(C_f)|_Y \).

We prove property e). If \((x_i)_i \in \mathbb{N} \xrightarrow{i \to \infty} x\) 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 large \( i \). For every \( i \in \mathbb{N} \) let us choose an arbitrary sequence \((x'_j)_j \in \mathbb{N} \xrightarrow{j \to \infty} x_i \). By B) \( f(x, \cdot) \) is continuous for \( C_f \) and \( C^\infty \), hence for large \( i \) there exists \( m_i \in \mathbb{N} \) such that \( f(x, x'_j) < \epsilon_i \) for all \( j \geq m_i \). Then \( f(x, x'_{m_i}) \leq \epsilon_i \) for large \( i \) implies
\[
\lim_{i \to \infty} f(x, x'_{m_i}) \leq \lim_{i \to \infty} \epsilon_i = 0 .
\]
This proves property e) of \( C_f \). A consequence for any \( Y \subset X \) is that \( \text{cl}_{C_f}(Y) \) is the \( \mathcal{T}(C_f) \) closure of \( Y \) (see Remark 3.12). \( \square \)

### 3.2 The I-topology and the rI-topology

The relative entropy \( S : \mathcal{S} \times \mathcal{S} \to [0, \infty] \) defines two divergence functions. The associated topologies recognize the convex geometry of the state space \( \mathcal{S} \): The relative interiors of faces of \( \mathcal{S} \) are connected components of the I-topology and on each component the I-topology is the norm topology. The rI-topology fits exactly the needs of the complete projection theorem Theorem 4.15. Of course \( \mathcal{S} \) is homeomorphic to a unit ball in the norm topology, a result known as Theorem of Sz. Nagy, see e.g. §VIII.1 in [Be63]. Corollary 3.22 collects conditions for a commutative algebra.

In the sequel let \( \omega \in \{I, rI\} \). Several important definitions and a summary of similarities between the \( \omega \)-topology and a norm topology are given in §1.4. Definition 1.6 introduces for \( \rho, \sigma \in \mathcal{S}_A \) and \( X \subset \mathcal{S}_A \) the functions \( S^I(\rho, \sigma) = S(\sigma, \rho) \), \( S^{rI}(\rho, \sigma) = S(\rho, \sigma) \) and \( S^\omega(\rho, X) = \inf_{\sigma \in X} S^\omega(\rho, \sigma) \).

**Definition 3.17.** Let \((\rho_i)_i \in \mathbb{N} \subset \mathcal{S} \) be a sequence and let \( \rho \in \mathcal{S} \). We define the \( \omega \)-convergence \( C^\omega \) on \( \mathcal{S} \) by
\[
\rho_i \xrightarrow{C^\omega} \rho : \iff \lim_{i \to \infty} S^\omega(\rho, \rho_i) = 0
\]

We begin with continuity of the relative entropy using on the positive augmented half-line \([0, \infty] \) the \( L^\alpha \)-convergence \( C^\infty \) of the Alexandroff compactification in Example 3.13.

**Proposition 3.18.** For every state \( \rho \in \mathcal{S} \) the mapping \( \mathcal{S} \to [0, \infty], \sigma \mapsto S^\omega(\rho, \sigma) \) is continuous for \( C^\omega \) and \( C^\infty \).

**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 \( s(\rho) \geq s(\sigma) \) holds, i.e. \( S(\sigma, \rho) < \infty \). Since \( \lim_{i \to \infty} S(\tau_i, \sigma) = 0 \) we have \( s(\sigma) \geq s(\tau_i) \) for large \( i \) and hence \( s(\rho) \geq s(\tau_i) \) holds for large \( i \). By the Pinsker-Csiszár inequality (13) the sequence \((\tau_i)_i \in \mathbb{N} \) converges to \( \sigma \) in norm. Hence the continuity of the von Neumann entropy, see e.g. §II.A in [We78], proves
\[
S(\tau_i, \rho) = -S(\tau_i) - \text{tr} \tau_i \log(\rho) \xrightarrow{i \to \infty} -S(\sigma) - \text{tr} \sigma \log(\rho) = S(\sigma, \rho).
\]
Second, we consider $s(\rho) \geq s(\sigma)$, i.e. $S(\sigma, \rho) = \infty$. By Remark 2.6.1 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 $s(\rho) \geq s(\sigma)$ 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 $s(\rho) \leq s(\sigma)$ with $S(\rho, \sigma) < \infty$. Since $S(\sigma, \tau_i) \to 0$ we have $s(\sigma) \leq s(\tau_i)$ for large $i$. Perturbation theory used in Lemma 2.9 completes the proof. □

Norm topology is too coarse for a continuity result similar to Proposition 3.18.

**Example 3.19.** If $A = \mathbb{C}^2$ then $S( (\frac{n-1}{n}, \frac{1}{n}), (1,0)) = \infty$ for all $n \in \mathbb{N}$ while $S((1,0), (1,0)) = 0$. If $A = \text{Mat}(2, \mathbb{C})$, then for real $\alpha$ we have

$$S \left( \frac{1}{2}(I_2 + \sigma_1), \frac{1}{2}(I_2 + \cos(\alpha)\sigma_1 + \sin(\alpha)\sigma_2) \right) = \begin{cases} 0 & \text{if } \alpha = 0 \mod 2\pi, \\ \infty & \text{else.} \end{cases}$$

A less trivial example is Example 9 in [KW11]: Any non-negative limit of $S(\rho, \sigma)$ (or divergence) can be achieved for smooth paths $\sigma_n$ converging in norm to an arbitrary point $\rho$ in the boundary of the Bloch ball $\mathcal{S}_{\text{Mat}(2, \mathbb{C})}$.

Taking the $\omega$-closure of a set does not decrease the relative entropy.

**Theorem 3.20.** Let $\rho \in \mathcal{S}$ and $X \subset \mathcal{S}$. Then $S^\omega(\rho, X) = S^\omega(\rho, \text{cl}^\omega(X))$ holds.

**Proof:** For every state $\sigma \in \text{cl}^\omega(X)$ there exists by Definition 1.7 a sequence $(\sigma_i)_{i \in \mathbb{N}} \subset X$, such that $\sigma_i \xrightarrow{\omega} \sigma$. Proposition 3.18 shows that the relative entropies $\lim_{i \to \infty} S(\rho, \sigma_i) = S(\rho, \sigma)$ converge. Hence

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

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

We now investigate the $\omega$-topology of the state space $\mathcal{S}$. The face lattice $\mathcal{F}$ of the state space $\mathcal{S}$ was introduced in §2.4. Several concepts around the $\omega$-topology were already introduced in §1.4. For example, the topology $\mathcal{T}^\omega = \mathcal{T}(C^\omega)$ is the $\omega$-topology on $\mathcal{S}$ according to Definition 1.8 and Definition 1.9. We denote the norm convergence on $\mathcal{S}$ by $C^{\|\|}$ and the norm topology on $\mathcal{S}$ by $\mathcal{T}^{\|\|} = \mathcal{T}(C^{\|\|})$. The preamble of Lemma 3.16 proves for subsets $X \subset \mathcal{S}$ that the sequential closure (35) equals the $\omega$-closure defined in Definition 1.7 and denoted by $\text{cl}^\omega(X)$. We have

$$\text{cl}^\omega(X) = \{ \rho \in \mathcal{S} \mid \lim_{i \to \infty} S(\rho, \rho_i) = 0 \text{ for a sequence } (\rho_i)_{i \in \mathbb{N}} \subset X \}$$

$$= \{ \rho \in \mathcal{S} \mid \inf_{i \in \mathbb{N}} S(\rho, \rho_i) = 0 \text{ for a sequence } (\rho_i)_{i \in \mathbb{N}} \subset X \}$$

$$= \{ \rho \in \mathcal{S} \mid S(\rho, X) = 0 \}.$$
Theorem 3.21. 1. The convergence $C^\omega$ is an $L^*$-convergence, in particular $C(T^\omega) = C^\omega$. We have $C^\omega \subset C_{\|\|}$ and $T^\omega \supset T_{\|\|}$, in particular $T^\omega$ is a Hausdorff topology.

2. For every $\rho \in S$ the mapping $S \to [0, \infty]$, $\sigma \mapsto S^\omega(\rho, \sigma)$ is continuous for $C^\omega$ and $C^e$ and for $T^\omega$ and $T^e$. For each $\rho, \sigma \in S$ and $\epsilon \in (0, \infty]$ the open $\omega$-disk $V^\omega(\rho, \epsilon)$ is $T^\omega$ open and the closed $\omega$-disk $W^\omega(\rho, \epsilon)$ is $T^\omega$ closed. The open $\omega$-disks $\{V^\omega(\rho, \epsilon) \mid \epsilon \in (0, \infty]\}$ are a base for $(S, T^\omega)$ at $\rho$. In particular $T^\omega$ is first countable and for any subset $X \subset S$ we have $T^\omega|_X = T(C^\omega|_X)$. For any subset $X \subset S$ the sequential closure $\text{cl}^e(X)$ is the $T^\omega$ closure of $X$.

3. Every term in the partition $S = \bigcup_{F \in F} \text{ri} F$ is a $T^1$ connected component of $S$. For all faces $F \in F$ we have $C^1_{|\text{ri} F} = C_{\|\|}_{|\text{ri} F}$ and $T^1_{|\text{ri} F} = T_{\|\|}_{|\text{ri} F}$.

4. We have $T_{\|\|} \subset T^1 \subset T^\omega$ and $C_{\|\|} \supset C^1 \supset C^\omega$.

5. We have $\text{cl}^1(\text{ri} S) = S$, in particular the topological space $(S, T^1)$ is connected.

Proof: The divergence functions $S^1(\rho, \sigma) := S(\sigma, \rho)$ and $S^1(\rho, \sigma) := S(\rho, \sigma)$ defined for $\rho, \sigma \in S$ satisfy the condition A) in Definition 3.14.2 by the Pinsker-Csiszár inequality (13) and they satisfy the condition B) by Proposition 3.18. Hence Lemma 3.16 part 1 and 2 prove part 1 and 2.

We show part 3. According to part 2, for every $\rho \in S$ the open I-disk of infinite radius is $T^1$ open and has by Remark 2.16 the form

$$V^1(\rho, \infty) = \{\sigma \in S \mid S(\sigma, \rho) < \infty\} = \{\sigma \in S \mid s(\sigma) \leq s(\rho)\} = F(s(\rho)).$$

By the lattice isomorphism $F : P \to F$ in Corollary 2.15 we obtain that every face $F$ of $S$ is $T^1$ open. Let us show that riF is $T^1$ open. The complement $S \setminus F$ is $T^1$ closed and the relative boundary $\text{rb} F$ of $F$ (in the norm topology) is norm closed. By part 2 we have $T_{\|\|} \subset T^1$ hence $\text{rb} F$ is $T^1$ closed and

$$\text{ri} F = S \setminus (\text{rb} F \cup (S \setminus F))$$

is $T^1$ open. Now by the stratification (30) the relative interior $\text{ri} F = S \setminus \bigcup_{G \in F} \text{ri} G$ is $T^1$ closed.

Let $F \in F$ be an arbitrary face. Since the relative entropy is norm continuous on $\text{ri} F \times \text{ri} F$ we have $C^1_{\|\|}_{|\text{ri} F} \subset C^1_{|\text{ri} F}$ and the converse inclusion follows from the Pinsker-Csiszár inequality. Hence $C^1_{\|\|}_{|\text{ri} F} = C^1_{|\text{ri} F}$ follows. With part 2 we have

$$T^1_{|\text{ri} F} = T(C^1_{|\text{ri} F}) = T(C^1_{\|\|}_{|\text{ri} F}) = T_{\|\|}_{|\text{ri} F}.$$

To show part 4 we begin with a proof of $T^1 \subset T^\omega$. We first notice $C^1_{|\text{ri} F} = C^1_{|\text{ri} F}$ for every face $F$ of $S$. This follows from $C^1_{\|\|}_{|\text{ri} F} = C^1_{|\text{ri} F}$ proved in part 3 and from $C^1_{\|\|}_{|\text{ri} F} = C^1_{\|\|}_{|\text{ri} F}$, which can be proved analogously. By part 2 we have

$$T(C^1_{|\text{ri} F}) = T(C^1_{\|\|}_{|\text{ri} F}) = T(C^1_{\|\|}_{|\text{ri} F}).$$

Let $U \in T^1$. Then $U \cap \text{ri} F \in T^1_{|\text{ri} F} = T^1_{|\text{ri} F}$ and since riF is $T^1$ open by part 3, this shows $U \cap \text{ri} F \in T^1$. Now $U = \bigcup_{F \in F}(U \cap \text{ri} F) \in T^1$ and we have proved $T^1 \subset T^1$. 

Part 2 adds the inequality $\mathcal{T}^\|\!\!\cdot\| \subset \mathcal{T}^1 \subset \mathcal{T}^1$. By part 1, these topologies arise from $L^*$-convergences hence we get $C^\|\!\!\cdot\| \supset C^1 \supset C^1$ from Theorem 3.5.

We prove part 5 and we first show that any non-empty $\mathcal{T}^1$ open set $U \subset S$ must intersect $\text{ri} \mathcal{S}$. Let $\rho \in U$, then by part 2 there is $\epsilon > 0$ such that $U$ contains an open $\text{ri}$-disk $V^1(\rho, \epsilon) = \{ \sigma \in \mathcal{S} \mid S(\rho, \sigma) < \epsilon \}$. We show that $V^1(\rho, \epsilon)$ intersects $\text{ri} \mathcal{S}$. The norm closure $\mathcal{S} = \overline{\text{ri} \mathcal{S}}$ contains $\rho$ hence there is a sequence $(\rho_i)_{i \in \mathbb{N}} \subset \text{ri} \mathcal{S}$ such that $\rho = \lim_{i \to \infty} \rho_i$. The relative entropy is lower continuous by Remark 2.6.1, so we have

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

Hence $(\rho_i)_{i \in \mathbb{N}} \cap V^1(\rho, \epsilon) \neq \emptyset$ proves $\text{ri} \mathcal{S} \cap U \neq \emptyset$.

As shown in the previous paragraph, the relative boundary $\text{ri} \mathcal{S}$ does not contain a $\mathcal{T}^1$ open set so the $\mathcal{T}^1$ closure of $\text{ri} \mathcal{S}$ equals $\mathcal{S}$. The claim $\text{cl}^1(\text{ri} \mathcal{S}) = \mathcal{S}$ follows because the $\mathcal{T}^1$ closure equals the sequential closure $\text{cl}^1(\text{ri} \mathcal{S})$ by part 2.

We show that $\mathcal{S}$ is $\mathcal{T}^1$ connected. By part 3 and 4 we have $\mathcal{T}^1|_{\text{ri} \mathcal{S}} = \mathcal{T}^\|\!\!\cdot\||_{\text{ri} \mathcal{S}}$. The convex set $\text{ri} \mathcal{S}$ is connected in the norm topology hence in the $\mathcal{T}^1$ topology. This shows that $\mathcal{S} = \text{cl}^1(\text{ri} \mathcal{S})$ is $\mathcal{T}^1$ connected because the closure of a connected set is connected, see e.g. §IV.7 in [Be63].

We formulate conditions for a commutative algebra.

**Corollary 3.22.** If $\dim C(A) > 1$, then $C^1 \subset C^\|\!\!\cdot\|$, $\mathcal{T}^1 \supset \mathcal{T}^1$ and $\mathcal{S}$ is not $\mathcal{T}^1$ compact. The following assertions are equivalent.

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

**Proof:** Item 1 implies 4. If $\mathcal{A}$ is commutative, then by (65) it is isomorphic to $C^N$. We can argue by convergence in components of $C^N$ and find $C^\|\!\!\cdot\| = C^\|\!\!\cdot\|$. We prove the statements in the headline. If $\dim C(A) > 1$, then by (65) $\mathcal{A}$ contains a $C^*$-subalgebra $B \cong C^2$ and by Example 3.19 we have $C^1|_{\mathcal{S}_B} \subset C^\|\!\!\cdot\||_{\mathcal{S}_B}$ while $C^\|\!\!\cdot\| |_{\mathcal{S}_B} = C^\|\!\!\cdot\||_{\mathcal{S}_B}$ was shown in the previous paragraph. So $C^1 \subset C^\|\!\!\cdot\|$ follows and we have also $\mathcal{T}^1 \subset \mathcal{T}^1$ because $C^\omega = C(\mathcal{T}^\omega)$ holds for $\omega \in \{1, 1, \mathcal{I} \}$ by Theorem 3.21.1.

We show that $\mathcal{S}$ is not $\mathcal{T}^1$ compact if $\dim C(A) > 1$. Theorem 3.21.3 shows $(\text{ri}\mathcal{S}, \mathcal{T}^\|\!\!\cdot\||_{\text{ri} \mathcal{S}})$ is not a compact topological space since $\text{ri} \mathcal{S}$ is the relative interior of a convex set of dimension $> 0$. Then $\mathcal{S}$ is not $\mathcal{T}^1$ compact because $\text{ri} \mathcal{S}$ is its $\mathcal{T}^1$ connected component.

Item 5 implies 3 and 6. By Proposition 2.14 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 [Be63]. Since $\mathcal{T}^1 = \mathcal{T}^\|\!\!\cdot\|$ is assumed, the state space is $\mathcal{T}^1$ compact and $\mathcal{T}^1$ second countable.

Item 1 implies 2. For every face $F \in \mathcal{F}$ we have $\mathcal{T}^1|_{\text{ri} F} = \mathcal{T}^\|\!\!\cdot\||_{\text{ri} F}$ by Theorem 3.21.3. As shown in the previous paragraph, $\mathcal{T}^\|\!\!\cdot\||_F$ is second countable. Since $\text{ri} F$ is an $\mathcal{T}^\|\!\!\cdot\||_F$ open subset of $F$, the topology $\mathcal{T}^1|_{\text{ri} F} = \mathcal{T}^\|\!\!\cdot\||_{\text{ri} F}$ is second countable. The simplex $\mathcal{S}$ is partitioned into finitely many relative interiors $\text{ri} F$ of faces $F$ by (30). Since each of these sets is a $\mathcal{T}^1$ connected component of $\mathcal{S}$, the proof is complete.
We prepare an argument to show that each of items 2, 3 or 6 implies 1. By Remark 2.16 we can write for any state \( \rho \in \mathcal{S} \) the open rI-disk of infinite radius in the form
\[
V^{rI}(\rho, \infty) = \{ \sigma \in \mathcal{S} \mid s(\rho) \preceq s(\sigma) \} = \bigcup_{p \in \mathcal{P}} rI\mathbb{F}(p). \tag{37}
\]
Here \( \mathcal{P} \) denotes the projection lattice of \( \mathcal{A} \). The open rI-disks are \( rI \)-open by Theorem 3.21.2. For pure states \( p, q \in \mathcal{P} \cap \mathcal{S} \) we have
\[
p \not\in V^{rI}(q, \infty) \text{ if } p \neq q.
\]
If \( \mathcal{A} \) is non-commutative then \( \mathcal{A} \) contains a C*-subalgebra isomorphic to \( \text{Mat}(2, \mathbb{C}) \), see (65), hence \( \mathcal{P} \cap \mathcal{S} \) is uncountable infinite.

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

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

Item 2 implies 1. Theorem 3.21.4 shows \( T^{rI} \subset T^I \) so the arguments in the previous paragraph apply unmodified.

Item 4 and 5 are equivalent. By Theorem 3.21.1 we have \( C^{rI} = C(T^{rI}) \) and \( C^{\|\|} = C(T^{\|\|}) \) holds because norm convergence is an \( L^* \)-convergence. On the other hand \( T^{rI} = T(C^{rI}) \) and \( T^{\|\|} = T(C^{\|\|}) \) hold by definition. \( \square \)

4 Exponential families in a matrix algebra

The complete projection theorem for an exponential family \( \mathcal{E} \) in a C*-subalgebra \( \mathcal{A} \) of \( \text{Mat}(n, \mathbb{C}) \) is proved in §4.2. It is connected to the rI-topology and will be formulated in terms of the rI-closure \( \text{cl}^{rI}(\mathcal{E}) \) defined in §1.4. The resulting rI-projection onto \( \text{cl}^{rI}(\mathcal{E}) \) is used in §4.3 to study local maximizers of the entropy distance \( d_{\mathcal{E}} \) defined in (12). In §4.4 we discuss non-commutative phenomena of the entropy distance and of the mean value parametrization of \( \text{cl}^{rI}(\mathcal{E}) \). We also make a continuity conjecture for \( d_{\mathcal{E}} \). In §4.5 we prove the complete Pythagorean theorem and we maximize the von Neumann entropy under linear constraints providing previously unknown solutions.

The analysis is based on the mean value parametrization of \( \mathcal{E} \) developed in §4.1. From Definition 1.4 we recall the real analytic function
\[
R_{\mathcal{A}} : \mathcal{A}_{sa} \to \mathcal{A}_{sa}, \quad R(\theta) = R_{\mathcal{A}}(\theta) = \exp_{\mathcal{A}}(a)/\text{tr}(\exp_{\mathcal{A}}(a)).
\]
Throughout this section we consider a non-empty affine subspace \( \Theta \subset \mathcal{A}_{sa} \), its translation vector space
\[
U := \text{lin}(\Theta) = \Theta - \Theta
\]
and the exponential family \( \mathcal{E} := R_{\mathcal{A}}(\Theta) \). Using orthogonal projection \( \pi_U : \mathcal{A}_{sa} \to U \), the complete projection theorem implies the bijection \( \pi_U|_{\text{cl}^{rI}(\mathcal{E})} : \text{cl}^{rI}(\mathcal{E}) \to \mathbb{M}(U) \) to the mean value set \( \mathbb{M}(U) = \pi_U(\mathcal{S}) \) defined in (4).
4.1 The mean value chart

We generalize the identity chart of the manifold of invertible states \( \text{ri} S \) in a C*-subalgebra \( \mathcal{A} \) of \( \text{Mat}(n, \mathbb{C}) \) to the mean value chart of an exponential family. Its inverse is the real analytic mean value parametrization of the exponential families.

We recall from §2.3 in [We11] that the relative interior of the state space consists of all invertible states,

\[
\text{ri} S = \{ \rho \in S \mid \rho^{-1} \text{ exists in } \mathcal{A} \}
\]  

and that \( \text{ri} S \) is open in the norm topology of \( \mathcal{A} \).

Restrictions to affine subspaces of \( \mathcal{A} \) are the rule in subsequent arguments, hence we accept relatively open convex subsets of \( \mathcal{A} \) (in place of open subset of \( \mathbb{R}^d \)) as domains of differentiable maps and as ranges of diffeomorphisms and charts.

**Proposition 4.1.** Let \( 1 \not\in U \) for the multiplicative identity \( 1 \) of \( \mathcal{A} \).

1. The projection \( \pi_U(\mathcal{E}) \) of \( \mathcal{E} \) onto \( U \) 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 \( \Theta \) has codimension one in \( \mathcal{A} \), then \( R_\mathcal{A}|_\Theta : \Theta \to \text{ri} S \) is a real analytic diffeomorphism.

3. The bijections \((R_\mathcal{A}|_\Theta)^{-1} : \Theta \to \mathcal{E} \) and \((\pi_U|_\mathcal{E})^{-1} : \pi_U(\mathcal{E}) \to \mathcal{E} \) are global charts for \( \mathcal{E} \) and \( \pi_U(\mathcal{E}) \) is real analytic.

**Proof:** In part 1, the derivative of \( \mathcal{A} \to \mathbb{R} \), \( a \mapsto \text{tr} \exp \mathcal{A}(a) \) can be computed from (22) using cyclic reordering under the trace. For \( a, u \in \mathcal{A} \) we have

\[
\left. \frac{\partial}{\partial t} \right|_{t=0} \text{tr} \exp \mathcal{A}(a + tu) = \langle u, \exp \mathcal{A}(a) \rangle.
\]

Hence the free energy (Definition 1.12) has the derivative for \( \theta, u \in \mathcal{A} \)

\[
\left. \frac{\partial}{\partial t} \right|_{t=0} F_\mathcal{A}(\theta + tu) = \langle u, R_\mathcal{A}(\theta) \rangle.
\]

From the product rule and (22) we get

\[
\left. \frac{\partial}{\partial t} \right|_{t=0} R_\mathcal{A}(\theta + tu) = \int_0^1 R_\mathcal{A}(\theta)^{1-y}uR_\mathcal{A}(\theta)^ydy - \langle u, R_\mathcal{A}(\theta) \rangle R_\mathcal{A}(\theta).
\]

For \( \theta, u, v \in \mathcal{A} \) we consider the real symmetric bilinear form

\[
\langle\langle u, v \rangle\rangle_\theta := \left. \frac{\partial^2}{\partial s \partial t} \right|_{s=t=0} F_\mathcal{A}(\theta + su + tv).
\]

If restricted to \( \theta \in \Theta \) and \( u, v \in U \) this bilinear form is called BKM-metric, see Remark 4.2. We obtain from (39) and (40)

\[
\langle\langle u, v \rangle\rangle_\theta = \langle u, \left. \frac{\partial}{\partial t} \right|_{t=0} R_\mathcal{A}(\theta + tu) \rangle = \int_0^1 \langle \xi(u, y), \xi(v, y) \rangle dy
\]

with the not necessarily self-adjoint matrix

\[
\xi(u, y) := R_\mathcal{A}(\theta)^{\frac{y}{2}} [u - \langle u, R_\mathcal{A}(\theta) \rangle 1] R_\mathcal{A}(\theta)^{\frac{1-y}{2}}.
\]
We have $\langle u, u \rangle_\theta > 0$ unless $u \in A_{sa}$ is a (real) scalar multiple of $\mathbb{1}$. Hence $\langle \cdot, \cdot \rangle_\theta$ is a non-degenerate bilinear form on every linear subspace of $A_{sa}$ not containing $\mathbb{1}$, in particular 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 (42)

$$\frac{\partial}{\partial t}|_{t=0} \pi_U \circ R(\theta + tu) = \pi_U \left( \frac{\partial}{\partial t}|_{t=0} R(\theta + tu) \right) = \sum_{i=1}^k \langle u, u_i \rangle_\theta u_i. \tag{43}$$

Since $\langle \cdot, \cdot \rangle_\theta$ 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 [KP02]. This implies that the image $\pi_U \circ R(\Theta)$ is an open subset relative to $U$.

On the other hand, the global injectivity of $\pi_U \circ R|_\Theta$ follows immediately from the projection theorem (11): 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 $\mathbb{1} \not\in U$. This completes the proof that $\pi_U \circ R|_\Theta : \Theta \to \pi_U(\mathcal{E})$ is a real analytic diffeomorphism.

In part 2, if $\Theta = A_0$ is the space of traceless matrices, then

$$\log_0 : \mathfrak{ri} \mathcal{S} \to A_0, \quad \rho \mapsto \log(\rho) - \frac{\text{tr} \log(\rho)}{\text{tr} \rho} \mathbb{1} \tag{44}$$

is inverse to $R|_{A_0}$ and this shows $R(\Theta) = \mathfrak{ri} \mathcal{S}$. Since $R(\theta + \mathbb{1}) = R(\theta)$ for all $\theta \in A_{sa}$, we have $R(\Theta) = \mathfrak{ri} \mathcal{S}$ for every affine subspace $\Theta \subset A_{sa}$ of codimension one and with $\mathbb{1} \not\in \text{lin}(\Theta)$.

In part 3, by virtue of the real analytic diffeomorphism in 1 it is sufficient to prove that $R|_{A_0} : \Theta \to \mathcal{E}$ is a real analytic bijection. The function $R_A$ is real analytic by definition and $R_A|_{A_0}$ is invertible on $\mathcal{E}$ by 2. \hfill\Box

**Remark 4.2.** If $\Theta$ has codimension one in $A_{sa}$ and if $\mathbb{1} \not\in U$ then the scalar product (42) defined at $\theta \in \Theta$ for $u, v \in U$ is the *BKM-metric*, a Riemannian metric on $\Theta \cong \mathfrak{ri} \mathcal{S}$ named after Bogoliubov, Kubo and Mori, see e.g. [Pe08]. Indeed,

$$v^{(-1)} := \frac{\partial}{\partial t}|_{t=0} R_A(\theta + tv)$$

is the $(-1)$-representation of a tangent vector in the identity chart $\text{id} : \mathfrak{ri} \mathcal{S} \to \mathfrak{ri} \mathcal{S}$. The $(+1)$-representation of $u^{(-1)} = \frac{\partial}{\partial t}|_{t=0} R_A(\theta + tu)$ equals

$$u^{(+1)} := D \log_A(u^{(-1)}) = \frac{\partial}{\partial t}|_{t=0} \log_A \circ R_A(\theta + tu) = u + \lambda \mathbb{1}$$

for some $\lambda \in \mathbb{R}$. Since $v^{(-1)}$ has trace zero, we arrive at the *mixed representation*

$$\langle u, v \rangle_\theta = \text{tr}(u^{(+1)} v^{(-1)})$$

of the BKM-metric, see e.g. [GS01].

Let us calculate the range of the chart $\pi_U|_{\mathcal{E}}$. The following statement gives us also an upper bound on the norm closure $\overline{\mathcal{E}}$. It is used implicitly for linear $\Theta$ in Lemma 7 in [Wi63].
Proposition 4.3. We assume \((x_i)_{i \in \mathbb{N}} \subset \Theta\) and that the states \(\rho_i := R_A(x_i)\), \(i \in \mathbb{N}\), converge in norm to \(\rho = \lim_{i \to \infty} \rho_i\). Then \(\rho \in \mathcal{E}\) or \(\lim_{i \to \infty} \|x_i\| = \infty\). In the second case, the set \(M \subset U\) of accumulation points of \(\left(\frac{x_i}{\|x_i\|}\right)_{i \in \mathbb{N}}\) is non-empty and we have \(s(\rho) \leq p^+_A(u)\) for every \(u \in M\).

Proof: If the sequence \((x_i)_{i \in \mathbb{N}}\) has a bounded subsequence then it has an accumulation point in \(\Theta\). Descending to a subsequence we assume the convergence of \(x := \lim_{i \to \infty} x_i \in \Theta\). By continuity of \(R_A\) we have

\[
\rho = \lim_{i \to \infty} R_A(x_i) = R_A(x) \in \mathcal{E}.
\]

Otherwise, if \(\rho \notin \mathcal{E}\) we have \(\lim_{i \to \infty} \|x_i\| = \infty\). We fix for all following arguments an accumulation point \(u\) of \(\left(\frac{x_i}{\|x_i\|}\right)_{i \in \mathbb{N}}\) (clearly \(u \in U\)). Let \(\Phi : A \to \text{Mat}(N, \mathbb{C})\) be a C*-algebra embedding with \(\Phi(1) = 1_N\). Weyl’s perturbation theorem (25) proves that the largest eigenvalue of \(\Phi(\frac{x_i}{\|x_i\|})\) converges to the largest eigenvalue of \(\Phi(u)\). Since the largest spectral value of \(a \in A\) is the largest eigenvalue of \(\Phi(a)\), we obtain

\[
\lim_{i \to \infty} \lambda^+_A(\frac{x_i}{\|x_i\|}) = \lambda^+_A(u).
\]

(45)

Let us assume \(u\) is not a real multiple of \(1\) (otherwise \(p^+_A(u) = 1\) and the statement \(s(\rho) \leq p^+_A(u)\) is trivial). We put

\[
y_i := x_i - \lambda^+_A(x_i)1
\]

for \(i \in \mathbb{N}\). Using \(\propto\) as proportionality by a positive real number, we have for \(i \in \mathbb{N}\)

\[
y_i = x_i - \lambda^+_A(x_i)1 \propto \frac{x_i}{\|x_i\|} - \lambda^+_A(\frac{x_i}{\|x_i\|})1.
\]

Since \(\left(\frac{x_i}{\|x_i\|}\right)_{i \in \mathbb{N}}\) converges to \(u\) and since \(u\) is not proportional to \(1\) we have \(\|y_i\| > 0\) for large \(i\). Hence (45) implies

\[
\lim_{i \to \infty} \frac{y_i}{\|y_i\|} = \frac{u - \lambda^+_A(u)1}{\|u - \lambda^+_A(u)1\|}.
\]

(46)

Since \(\lambda^+(y_i) = 0\), we have \(\|e^{y_i}\| \leq 1\) for all \(i \in \mathbb{N}\). Hence it is possible to select a subsequence of \((y_i)_{i \in \mathbb{N}}\), which we call by the same name, such that

\[
\tilde{u} := \lim_{i \to \infty} \frac{y_i}{\|y_i\|} \quad \text{and} \quad a := \lim_{i \to \infty} \exp_A(y_i).
\]

The argument in the first paragraph (applied to \(\mathcal{E} = R_A(\Theta + \mathbb{R}1)\)) shows \(\lim_{i \to \infty} \|y_i\| = \infty\) because

\[
\rho = \lim_{i \to \infty} R_A(x_i) = \lim_{i \to \infty} R_A(y_i) \notin \mathcal{E}.
\]

Now Lemma 2.10.1 shows \(s(a) \leq p^+_A(\tilde{u})\). Clearly \(\rho = \frac{a}{\text{tr} a}\) holds and we finish by proving \(p^+_A(u) = p^+_A(\tilde{u})\), which follows from (46).

\(\Box\)

A statement like the following is implicitly used in Theorem 2 b) in [Wi63]. For completeness we provide a proof.
**Lemma 4.4.** 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 \( f(K) \cap L = \emptyset \), then \( f(K) = L \).

**Proof:** Since \( f(K) \cap L = \emptyset \) we have \( L \setminus f(K) = L \setminus f(K) = (W \setminus f(K)) \cap L \) and \( f(K) \cap L = f(K) \cap L \), hence

\[
L = (f(K) \cap L) \cup (L \setminus f(K)) = (f(K) \cap L) \cup ((W \setminus f(K)) \cap L).
\] (47)

The set \( f(K) \) is open in \( W \) by assumption and since \( f(K) \) is compact \( W \setminus f(K) \) is open in \( W \). Since \( f(K) \cap L \neq \emptyset \) by assumption, (47) is a disconnection of \( L \) unless \( L \setminus f(K) = \emptyset \). Since \( L \) is connected by assumption, \( f(K) \supset L \) follows. \( \square \)

We have collected all arguments needed to compute \( \pi_U(\mathcal{E}) \). The mean value set \( \mathcal{M}(U) \) plays a crucial role (4).

**Theorem 4.5.** Let \( U \not\subset U \). Then \( \text{ri} \mathcal{M}_A(U) \) is open in the norm topology of \( U \) and the chart change \( \pi_U \circ R_A|_{\Theta} : \Theta \to \text{ri} \mathcal{M}_A(U) \) is a real analytic diffeomorphism. We have \( \pi_U(\mathcal{E} \setminus \mathcal{E}) = \text{rb} \mathcal{M}_A(U) \).

**Proof:** The map \( \pi_U \circ R|_{\Theta} : \Theta \to \pi_U(\mathcal{E}) \) is a real analytic diffeomorphism by Proposition 4.1.1 and \( \pi_U(\mathcal{E}) \) is open relative to \( U \). We shall first show

\[
\pi_U(\mathcal{E} \setminus \mathcal{E}) \subset \text{rb} \mathcal{M}_A(U).
\] (48)

Let \( \rho \in \mathcal{E} \setminus \mathcal{E} \). Proposition 4.3 shows that the support projection of \( \rho \) satisfies \( s(\rho) \leq p^+(u) \) for a non-zero \( u \in U \) and Proposition 2.14 shows that \( \rho \) lies in the exposed face \( \mathbb{F}(p^+(u)) = F_\perp(S, u) \) of the state space. Then Lemma 2.18 shows that \( \pi_U(\rho) \) lies in the exposed face \( F_\perp(M(U), u) \) of the mean value set. The mean value set \( \mathcal{M}(U) \) has non-empty interior because it contains \( \pi_U(\mathcal{E}) \) and then Theorem 13.1 in [Ro72] proves that the exposed face \( F_\perp(M(U), u) \) is included in the boundary of \( \mathcal{M}(U) \). This proves \( \pi_U(\rho) \in \text{rb} \mathcal{M}_A(U) \).

In order to prove that \( \pi_U \circ R|_{\Theta} : \Theta \to \text{ri} \mathcal{M}(U) \) is a real analytic diffeomorphism it suffices to prove \( \pi_U(\mathcal{E}) = \text{ri} \mathcal{M}(U) \). The convex body \( \mathcal{M}(U) \) is the projection of the whole state space, so \( \pi_U(\mathcal{E}) \subset \mathcal{M}(U) \). But \( \mathcal{E} \subset \text{ri}(S) \) holds by (38) and thanks to the equality \( \pi_U \circ \text{ri}(S) = \pi_U(\mathcal{E}) \) (see e.g. Theorem 6.6 in [Ro72]) we have \( \pi_U(\mathcal{E}) \subset \text{ri} \mathcal{M}(U) \). We meet the conditions of Lemma 4.4 with

\[
V = A_{sa}, \quad W = U, \quad f = \pi_U, \quad K = \mathcal{E} \quad \text{and} \quad L = \text{ri} \mathcal{M}(U).
\]

Indeed, \( K = \mathcal{E} \subset S \) is non-empty and bounded, the convex set \( L = \text{ri} \mathcal{M}(U) \) is connected. We have proved in this paragraph that \( f(K) = \pi_U(\mathcal{E}) \) is included in \( L = \text{ri} \mathcal{M}(U) \) and \( f(K) = \pi_U(\mathcal{E}) \) is open relative to \( W = U \) as the range of the diffeomorphism \( \pi_U \circ R|_{\Theta} \). Moreover \( \pi_U(\mathcal{E} \setminus \mathcal{E}) \subset \text{rb} \mathcal{M}(U) \) in (48) implies

\[
f(K) \cap L = \pi_U(\mathcal{E} \setminus \mathcal{E}) \cap \text{ri} \mathcal{M}(U) \subset \text{rb} \mathcal{M}(U) \cap \text{ri} \mathcal{M}(U) = \emptyset.
\]

Then \( \pi_U(\mathcal{E}) = \text{ri} \mathcal{M}(U) \) follows.
Finally we show $\pi_U(\mathcal{E} \setminus \mathcal{E}) = \text{rb} \mathcal{M}(U)$. Since $\mathcal{E} \subset \mathcal{S}$ is compact and $\pi_U(\mathcal{E}) = \text{ri} \mathcal{M}(U)$ we have
$$\mathcal{M}(U) = \overline{\pi_U(\mathcal{E})} \subset \pi_U(\mathcal{E}) \subset \mathcal{M}(U)$$
hence $\pi_U(\mathcal{E}) = \mathcal{M}(U)$. Then $\pi_U(\mathcal{E}) = \text{ri} \mathcal{M}(U)$ proves $\pi_U(\mathcal{E} \setminus \mathcal{E}) \supset \text{rb} \mathcal{M}(U)$. The opposite inclusion is (48). □

In the sequel $1 \in U$ will naturally occur in our constructions.

**Corollary 4.6.** The map $\pi_U |_{\mathcal{E}} : \mathcal{E} \to \text{ri} \mathcal{M}(U)$ is a bijection and the inverse $(\pi_U |_{\mathcal{E}})^{-1} : \text{ri} \mathcal{M}(U) \to \mathcal{E}$ is real analytic.

**Proof:** If $1 \not\in U$ then the claim follows from Theorem 4.5 and Proposition 4.1.3. We assume $1 \in U$ and we define $\Theta_0 := \pi_{A_0}(\Theta)$ and $U_0 := \pi_{A_0}(U)$. We have $U_0 = \text{lin}(\Theta_0)$ and since $A_0 \perp 1 = 1 \mathbb{R}$ holds, we have $U = U_0 + 1 \mathbb{R}$ and $\Theta = \Theta_0 + 1 \mathbb{R}$. Clearly $\mathcal{E} = R_A(\Theta) = R_A(\Theta_0)$ and $(\pi_{U_0} |_{\mathcal{E}})^{-1} : \text{ri} \mathcal{M}(U_0) \to \mathcal{E}$ is a real analytic bijection because $1 \not\in U_0$. We have $\pi_U = \pi_{U_0} + \pi_{1 \mathbb{R}}$ and
$$\pi_U |_{A_1} = \pi_{U_0} |_{A_1} + \frac{1}{1 \mathbb{I}} 1 .$$
Hence $\text{ri} \mathcal{M}(U) = \text{ri} \mathcal{M}(U_0) + \frac{1}{1 \mathbb{I}} 1$ holds and for $u \in \text{ri} \mathcal{M}(U)$ the equality $(\pi_U |_{\mathcal{E}})^{-1}(u) = (\pi_{U_0} |_{\mathcal{E}})^{-1}(u - \frac{1}{1 \mathbb{I}} 1)$ completes the proof. □

**Definition 4.7.** We call the continuous bijection $\pi_U |_{\mathcal{E}} : \mathcal{E} \to \text{ri} \mathcal{M}(U)$ in Corollary 4.6 the **mean value chart** of $\mathcal{E}$. The real analytic inverse $(\pi_U |_{\mathcal{E}})^{-1} : \text{ri} \mathcal{M}(U) \to \mathcal{E}$ is the **mean value parametrization** of $\mathcal{E}$.

The mean value parametrization in a non-commutative algebra is demonstrated in Figure 4 with the simplest non-trivial examples:

**Example 4.8.** The **Swallow family**, studied in [KW11], is the exponential family
$$R(\text{span}_\mathbb{R}(\sigma_1 \oplus 1, \sigma_2 \oplus 1))$$
in the algebra $\text{Mat}(2, \mathbb{C}) \oplus \mathbb{C}$. The **Staffelberg family** was defined in Example 1.5 as $R(\text{span}_\mathbb{R}(\sigma_1 \oplus 0, \sigma_2 \oplus 1))$. These exponential families are included in the exponential

![Figure 4: The Swallow family (left) and Staffelberg family (right) are sketched by e-geodesics. The cones about the families replace the state space. The mean value sets are projections of the state space, their boundaries are drawn below.](image-url)
family \( R(V) \) for \( V := \text{span}(\sigma_1 \oplus 0, \sigma_2 \oplus 0, -I_2 \oplus 2) \). The norm closure \( \overline{R(V)} \) is the 3D-cone drawn about each family. It is shown in Example 6.3 in [We10] that \( R(V) \) consists of all matrices with real coefficients \( a, b, c \)

\[
M := \begin{pmatrix}
  a & b-ic & 0 \\
  b+ic & a & 0 \\
  0 & 0 & 1-2a
\end{pmatrix}
\]

such that \( M \succeq 0 \). If \( E \) denotes one of the above families, let \( U := \log_0(\mathcal{E}) \) be the linear subspace of \( \mathcal{A}_0 \) such that \( E = R(U) \), see (44). It is known that

\[
\mathbb{M}(U) = \pi_U(\overline{R(V)}) = \pi_U(S_{\text{Mat}(2, \mathbb{C})}) = \pi_U(S_{\text{Mat}(3, \mathbb{C})}).
\]

In fact, to apply Lemma 3.13 in [We11] we should first use an isometry that permutes \( \sigma_2 \oplus 0 \) with \( \sigma_3 \oplus 0 \) and makes all entries of matrices in \( U \) and \( V \) real.

### 4.2 The complete projection theorem

We extend the projection theorem for an exponential family \( \mathcal{E} \), explained in §1.3, to the entire state space \( \mathcal{S}_A \). This will be done by defining an extension \( \text{ext}(\mathcal{E}) \) of \( \mathcal{E} \) composed of exponential families in compressed algebras \( pA_{p} \) for orthogonal projections \( p \), one for each face of the mean value set \( \mathbb{M}(U) \) (of the vector space \( U \)). As turns out, the extension \( \text{ext}(\mathcal{E}) \) equals the \( rI \)-closure \( \text{cl}^{rI}(\mathcal{E}) \).

We obtain a bijective mean value parametrization

\[
(\pi_U|_{\text{cl}^{rI}(\mathcal{E})}) : \mathbb{M}(U) \to \text{cl}^{rI}(\mathcal{E}),
\]

whose continuity will be discussed in §4.4. We obtain an \( rI \)-projection with linear fibers

\[
\pi_\mathcal{E} : \mathcal{S}_A \to \text{cl}^{rI}(\mathcal{E})
\]

from the state space \( \mathcal{S}_A \) to the \( rI \)-closure. We finish by providing analogues of the natural and canonical parameters known in statistics, both for the exponential family \( \mathcal{E} \) and for its \( rI \)-closure.

For orthogonal projections \( p \in \mathcal{P} \) the orthogonal projection of \( \mathcal{A}_{sa} \) onto \( (pA_{p})_{sa} \) was denoted in (34) by

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

Only the projections in the lattice \( \mathcal{P}^U \) are interesting. They correspond to the faces the mean value set \( \mathbb{M}(U) \) by (32) and can in principle be computed by spectral analysis, see Remark 2.26.

**Definition 4.9.** For orthogonal projections \( p \in \mathcal{P} \) we consider the exponential family

\[
\mathcal{E}_p := R_{pA_{p}}(c^p(\Theta)).
\]

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

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

We begin by parametrizing the extension \( \text{ext}(\mathcal{E}) \) from the mean value set \( \mathbb{M}(U) \).
**Lemma 4.10.** The projection \( \pi_U|_{\text{ext}(\mathcal{E})} : \text{ext}(\mathcal{E}) \to \mathbb{M}(U) \) is a bijection.

*Proof:* For every non-zero projection \( p \in \mathcal{P} \), the mean value chart in Corollary 4.6 proves the bijection
\[
\pi_{cP(U)}|_{\mathcal{E}_p} : \mathcal{E}_p \to \text{ri} \, \mathbb{M}_{P,p}(c^p(U)).
\]
Using the third diagram in Lemma 2.23 we have the bijection
\[
\pi_U|_{\mathcal{E}_p} : \mathcal{E}_p \to \text{ri} \, \pi_U(\mathcal{F}_A(p)).
\] (49)

The map \( \mathcal{P}^U \to \mathcal{F}(\mathbb{M}(U)) \), \( p \mapsto \pi_U(\mathcal{F}_A(p)) \) is a lattice isomorphism from the projection lattice \( \mathcal{P}^U \) to the face lattice of the mean value set, see (32). Then the stratification (30) of \( \mathbb{M}(U) \) into the relative interiors of its faces shows that the bijections (49) assemble to a bijection \( \text{ext}(\mathcal{E}) \to \mathbb{M}(U) \). \( \square \)

We will denote the following extension of the projection \( \pi_\mathcal{E} \) from (10) by the same symbol \( \pi_\mathcal{E} \) and we call it \( rI\)-projection according to the result of Theorem 4.15.

**Definition 4.11.** The \( rI\)-projection to \( \mathcal{E} \) is well-defined by Lemma 4.10 as the map
\[
\pi_\mathcal{E} : \mathcal{S}_A \to \text{ext}(\mathcal{E}), \quad \rho \mapsto (\pi_U|_{\text{ext}(\mathcal{E})})^{-1} \circ \pi_U(\rho).
\]
For technical reasons we denote the relative entropy with the first argument \( \rho \in \mathcal{S} \) fixed by \( S_\rho : \mathcal{S} \to [0, \infty], S_\rho(\sigma) := S(\rho, \sigma) \).

For convenience we cite Lemma 7 and Lemma 10 in [KW11]. The first proposition follows immediately from Lemma 2.10.2, the second requires some computation.

**Lemma 4.12.** Suppose \( \theta, u \in \mathcal{A}_{sa} \) and \( p := p_A^+(u) \) is the maximal projection of \( u \). We have
\[
\lim_{t \to \infty} R_A(\theta + tu) = R_{pA}(c^p(\theta)).
\] (50)
and
\[
\lim_{t \to \infty} (F_A(\theta + tu) - t \lambda_A^+(u)) = F_{pA}(c^p(\theta)).
\] (51)

**Lemma 4.13.** Suppose \( \theta, u \in \mathcal{A}_{sa} \) and \( u \) is not proportional to the multiplicative identity \( 1 \) in \( \mathcal{A} \). If the state \( \rho \) belongs to the exposed face \( F^+(\mathcal{S}_A, u) \) of the state space, then \( S_\rho(R_A(\theta + tu)) \) is strictly monotone decreasing with \( t \in \mathbb{R} \) and
\[
\inf_{t \in \mathbb{R}} S_\rho(R_A(\theta + tu)) = \lim_{t \to -\infty} S_\rho(R_A(\theta + tu)) = S_\rho(\lim_{t \to -\infty} R_A(\theta + tu)).
\]

The next proposition is an intriguing interplay between convex geometry and matrix calculus making a statement about the relative entropy \( S_\rho(\sigma) = S(\rho, \sigma) \) along a single curve.

**Proposition 4.14.** Let \( \rho \in \mathcal{S}_A \) and \( \sigma \in \text{ext}(\mathcal{E}) \), such that \( \sigma \neq \pi_\mathcal{E}(\rho) \) and \( S_\rho(\sigma) < \infty \). There exists a continuous path \( \gamma : [0, 1] \to \text{ext}(\mathcal{E}) \) 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. \(\mathcal{d}_\varepsilon(\rho) \leq S_\rho(\gamma(1))\).

**Proof:** We construct \(\gamma\) by concatenation of several e-geodesics. Since \(S_\rho(\sigma) < \infty\) we have \(s(\rho) \leq s(\sigma)\). By Lemma 2.27 there exists a projection \(p \in \mathcal{P}^U\), such that \(\rho \in \text{ri} F_A(p) + U^\perp\) and then it follows from Lemma 4.10 that \(\pi_\varepsilon(\rho) \in \mathcal{E}_p\). We denote by \(q \in \mathcal{P}^U\) the projection such that \(\sigma \in \mathcal{E}_q\), i.e. \(s(\sigma) = q\). By Lemma 2.27 we have

\[
p = \bigwedge\{r \in \mathcal{P}^U \mid s(\rho) \leq r\} \succeq s(\sigma) = q\,.
\]

By Corollary 2.25 there exists an access sequence of projections for \(U\) including both \(p\) and \(q\), say

\[
1 = p_0 \succ p_1 \succ \cdots \succ p_m = p,
\]

where \(m \geq 0\), \(p = p_m\) and \(q = p_l\) for \(l \leq m\).

We define \((m - l)\) e-geodesic rays in \(\mathcal{E}_{p_1}, \mathcal{E}_{p_{l+1}}, \ldots, \mathcal{E}_{p_{m-1}}\). From the Definition 2.13.1 of an access sequence and by Corollary 2.20, for each \(k = 0, \ldots, m - 1\) there exists \(u_k \in c^{p_k}(U)\), such that

\[
p_{k+1} = p^{+}_{p_k A p_k}(u_k).
\]

Moreover, \(u_k\) is not a multiple of the identity \(p_k\) in \(p_k A p_k\) (because \(p_{k+1} \neq p_k\)). Let \(\theta \in \Theta\) such that \(\sigma = R_{q, A_q}(\varepsilon'(\theta))\). We define for \(k = 0, \ldots, m - 1\) the e-geodesic

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

By (50) we can define

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

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

We concatenate the reparametrized e-geodesic rays to a continuous curve \(\tilde{\gamma} : [0, m - l] \to \text{ext}(\mathcal{E})\). The pieces fit together by (54). If \(\sigma_m \neq \pi_\varepsilon(\rho)\) then we add the e-geodesic segment in \(\mathcal{E}_p\) from \(\sigma_m\) to \(\pi_\varepsilon(\rho)\). This is parametrized under \(R_{p_k A p_k}\) by a straight line segment in \(c^\Theta(\Theta)\), which we parametrize linearly by the unit interval \([0, 1]\). Since \(\sigma \neq \pi_\varepsilon(\rho)\), one of the inequalities \(m - l > 0\) or \(\sigma_m \neq \pi_\varepsilon(\rho)\) must be true so we obtain a curve \(\gamma : [0, 1] \to \text{ext}(\mathcal{E})\) from \(\tilde{\gamma}\) by a parametrization spedup by a factor of \(m - l\) or \(m - l + 1\).

We argue that \(S_\rho\) is strictly monotone decreasing along \(\gamma\). This can be done for the rays in the natural parametrization (53) for \(k = 0, \ldots, m - 1\). Since \(p \preceq p_{k+1} \preceq p_k\) we have by Proposition 2.14 and (52)

\[
\rho \in F_A(p) \subset F_A(p_{k+1}) = F_{p_k A p_k}(p_{k+1}) = F_{\perp}(S_{p_k A p_k}; u_k).
\]

Since \(u_k\) is not a multiple of the identity \(p_k\), Lemma 4.13 can be invoked and it shows that \(S_\rho\) is strictly monotone decreasing along \(g_k\).

The fact that \(S_\rho\) is strictly monotone decreasing along the e-geodesic segment from \(\sigma_m\) to \(\pi_\varepsilon(\rho)\) uses the strict convexity of \(S_\rho\) on \(\mathcal{E}_p\) in the parametrization of \(R_{p_k A p_k}\). In order to have an injective parametrization we project \(c^\Theta(\Theta)\) onto \((p A p)\) =
\{a \in (pAp)_{sa} \mid \text{tr}(a) = 0 \} \) so that \( \Theta_0 := \pi_{(pAp)_{0}}(c^p(\Theta)) \) satisfies \( \mathcal{E}_p = R_{pAp}(c^p(\Theta)) = R_{pAp}(\Theta_0) \). For all \( \theta_0 \in \Theta_0 \) an elementary calculation shows
\[ S_p(R_{pAp}(\theta_0)) = -S(\rho) - \text{tr}(\rho \theta_0) + F_{pAp}(\theta_0) \]
where \( S(\rho) \) is the von Neumann entropy and \( F \) is the free energy from Definition 1.12. Hence for \( u_0, v_0 \in \pi_{(pAp)_{0}}(c^p(U)) \) equals the BKM-metric (41). As discussed in the paragraph following (42) the free energy has a positive definite Hessian on \( \Theta_0 \) so \( S_\rho \circ R_{pAp} \) is strictly convex on \( \Theta_0 \).

Since \( \pi_\xi(\rho) \in \mathcal{E}_p \), the function \( S_\rho \) has on \( \mathcal{E}_p \equiv \Theta_0 \) a global minimum at \( \pi_\xi(\rho) \), this follows from the projection theorem (10). Hence \( S_\rho \) is strictly monotone decreasing along the e-geodesic from \( \sigma_m \in \mathcal{E}_p \) to \( \pi_\xi(\rho) \in \mathcal{E}_p \).

Second, we prove \( d_\xi(\rho) \leq S_\rho(\pi_\xi(\rho)) \) by showing for \( k = 0, \ldots, m - 1 \) that \( d_{\xi_{pk}}(\rho) \leq d_{\xi_{pk+1}}(\rho) \) holds. Then
\[ d_\xi(\rho) = d_{\xi_{p0}}(\rho) \leq d_{\xi_{p1}}(\rho) \leq \cdots \leq d_{\xi_{pm}}(\rho) = d_{\xi_{p}}(\rho) \leq S_\rho(\pi_\xi(\rho)) \]
will follow, the last inequality since \( \pi_\xi(\rho) \in \mathcal{E}_p \). Let \( \tau \in \mathcal{E}_{pk+1} \) and let \( \theta \in \Theta \) such that \( \tau = R_{pk+1Ap}(c_{pk+1}(\theta)) \). Then by (52), (55) and Lemma 4.13 we have
\[ d_{\xi_{pk}}(\rho) \leq \inf_{\theta \in \mathbb{R}} S_\rho(R_{pk+1Ap}(c_{pk}(\theta) + tu_k)) = S_\rho(R_{pk+1Ap}(c_{pk+1}(\theta))) = S_\rho(\tau) \]
Taking the infimum over all \( \tau \in \mathcal{E}_{pk+1} \) the claim follows. \( \square \)

Local minimizers in the following theorem are understood in the norm topology: If \((X, \mathcal{T})\) is a topological space, \(Y \subseteq X\) and \(f : Y \to \mathbb{R}\), then \(x_0 \in Y\) is a local minimizer (maximizer) of \(f\) on \(Y\), if there is a \(\mathcal{T}\) neighborhood \(V\) of \(x_0\) in \(X\), 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 rI-closure \(c_{\text{rI}}(\mathcal{E}) = \{\rho \in \mathcal{S} \mid \inf_{\sigma \in \mathcal{E}} S(\rho, \sigma) = 0\}\).

**Theorem 4.15** (Complete projection theorem). We have \(c_{\text{rI}}(\mathcal{E}) = \text{ext}(\mathcal{E})\) and \(\pi_U|_{c_{\text{rI}}(\mathcal{E})} : c_{\text{rI}}(\mathcal{E}) \to M_A(U)\) is a bijection. For each \(\rho \in \mathcal{S}\) the relative entropy \(S_\rho\) has a unique local minimizer on the rI-closure \(c_{\text{rI}}(\mathcal{E})\) at the rI-projection \(\pi_\xi(\rho)\). The entropy distance is \(d_\xi(\rho) = S_\rho(\pi_\xi(\rho)) = \min_{\sigma \in c_{\text{rI}}(\mathcal{E})} S_\rho(\sigma)\).

**Proof:** For each \(\rho \in \mathcal{S}\) we observe from Proposition 4.14.1 and from the fact that \(S_\rho\) is finite on \(\mathcal{E}\), that \(\pi_\xi(\rho)\) is the unique global minimizer of \(S_\rho\) on \(\text{ext}(\mathcal{E})\). By Proposition 4.14.2 all \(\sigma \in \text{ext}(\mathcal{E})\) satisfy \(d_\xi(\rho) \leq S_\rho(\pi_\xi(\rho)) \leq S_\rho(\sigma)\). Taking the infimum over \(\sigma \in \text{ext}(\mathcal{E})\) this shows \(d_\xi(\rho) \leq d_{\text{ext}(\mathcal{E})}(\rho)\) and since the converse inequality is trivial, we have proved
\[ d_\xi(\rho) = d_{\text{ext}(\mathcal{E})}(\rho) . \]
Now \(c_{\text{rI}}(\mathcal{E}) = c_{\text{rI}}(\text{ext}(\mathcal{E}))\) follows immediately, see (12) and Definition 1.7. We show \(c_{\text{rI}}(\text{ext}(\mathcal{E})) = \text{ext}(\mathcal{E})\). The inclusion “\(\subseteq\)” is trivial. Conversely, let \(\rho \in c_{\text{rI}}(\text{ext}(\mathcal{E}))\). Then \(d_{\text{ext}(\mathcal{E})}(\rho) = 0\). Since the relative entropy is non-negative and since \(\pi_\xi(\rho)\)
is a global minimizer of $S_\rho$ on $\text{ext}(\mathcal{E})$ we have $S_\rho(\pi_\epsilon(\rho)) = 0$, i.e. $\rho = \pi_\epsilon(\rho)$. In particular $\rho \in \text{ext}(\mathcal{E})$. We conclude $\text{ext}(\mathcal{E}) = \text{cl}^1(\mathcal{E})$. Now Lemma 4.10 shows that $\pi_U^\epsilon|_{\text{cl}^1(\mathcal{E})} : \text{cl}^1(\mathcal{E}) \to M(A(U))$ is a bijection.

It remains to discuss local minimizers $\sigma \in \text{cl}^1(\mathcal{E})$ of $S_\rho$ on $\text{cl}^1(\mathcal{E})$. If $S_\rho(\sigma) < \infty$, then Proposition 4.14.1 shows that $\sigma$ is not a local minimizer unless $\sigma = \pi_\epsilon(\rho)$. If $S_\rho(\sigma) = \infty$ we observe that $\mathcal{E}$ is norm dense in $\text{cl}^1(\mathcal{E})$ by the Pinsker-Csizár inequality (13). Since $S_\rho$ has finite values on $\mathcal{E}$, the state $\sigma$ is not a local minimizer. □

The following coordinates $\langle \lambda_1, \ldots, \lambda_k \rangle \in \mathbb{R}^k$ depend on the knowledge of the projection lattice $\mathcal{P}^U$, which in principle can be computed by the method of Remark 2.26. The derivative of the free energy $F$ allows us to compute them.

**Corollary 4.16.** Let $\theta_0, u_1, \ldots, u_k \in A_{sa}$ and let $\Theta := \theta_0 + \text{span}_{\mathbb{R}}(u_1, \ldots, u_k)$ (then $U = \text{span}_{\mathbb{R}}(u_1, \ldots, u_k)$ and $\mathcal{E} = R_A(\Theta)$). The mean value map defines a bijection from the $rI$-closure of $\mathcal{E}$ to the convex support

$$m_{u_1, \ldots, u_k}|_{\text{cl}^1(\mathcal{E})} : \text{cl}^1(\mathcal{E}) \to \text{cs}(u_1, \ldots, u_k).$$

For every $\rho \in \text{cl}^1(\mathcal{E})$ exists a unique projection $p \in \mathcal{P}^U$ and some (in general non-unique) $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$, such that

$$\left(\frac{\partial}{\partial \lambda_1}, \ldots, \frac{\partial}{\partial \lambda_k}\right) F_{p, Ap}(c^\rho(\theta_0 + \sum_{i=1}^k \lambda_i u_i)) = m_{u_1, \ldots, u_k}(\rho).$$

(56)

For each solution $[p, (\lambda_1, \ldots, \lambda_k)]$ to (56) we have $R_{p, Ap}(c^\rho(\theta_0 + \sum_{i=1}^k \lambda_i u_i)) = \rho$.

**Proof:** Theorem 4.15 shows $\text{cl}^1(\mathcal{E}) = \text{ext}(\mathcal{E})$. If $\rho \in \text{cl}^1(\mathcal{E})$ then by definition of $\text{ext}(\mathcal{E})$ there exists $p \in \mathcal{P}^U$ and $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ such that for $\theta := \theta_0 + \sum_{i=1}^k \lambda_i u_i$ we have $\rho = R_{p, Ap}(c^\rho(\theta))$. Since $s(\rho) = p$, the projection $p$ is unique. The derivative (39) of the free energy is for $j = 1, \ldots, k$ given by

$$\frac{\partial}{\partial \lambda_j} F_{p, Ap}(c^\rho(\theta)) = \langle c^\rho(u_j), R_{p, Ap}(c^\rho(\theta)) \rangle = \langle u_j, R_{p, Ap}(c^\rho(\theta)) \rangle$$

hence the existence part follows from $\rho = R_{p, Ap}(c^\rho(\theta))$ and the equation

$$\left(\frac{\partial}{\partial \lambda_1}, \ldots, \frac{\partial}{\partial \lambda_k}\right) F_{p, Ap}(c^\rho(\theta)) = m_{u_1, \ldots, u_k}(R_{p, Ap}(c^\rho(\theta))).$$

(57)

On the other hand, if $[p, (\lambda_1, \ldots, \lambda_k)]$ solves (56), then (57) implies

$$m_{u_1, \ldots, u_k}(R_{p, Ap}(c^\rho(\theta))) = m_{u_1, \ldots, u_k}(\rho).$$

This implies $R_{p, Ap}(c^\rho(\theta)) = \rho$ since the mean value map restricts to the bijection

$$m_{u_1, \ldots, u_k}|_{\text{cl}^1(\mathcal{E})} : \text{cl}^1(\mathcal{E}) \to \text{cs}(u_1, \ldots, u_k).$$

Indeed, $\pi_U^\epsilon|_{\text{cl}^1(\mathcal{E})} : \text{cl}^1(\mathcal{E}) \to M(U)$ is a bijection by Theorem 4.15. The claim follows because $m_{u_1, \ldots, u_k} = m_{u_1, \ldots, u_k} \circ \pi_U$ holds and $m_{u_1, \ldots, u_k}|_{M(U)} : M(U) \to \text{cs}(u_1, \ldots, u_k)$ is a bijection by Remark 1.2.5. □

The parameters $m_{u_1, \ldots, u_k}(\rho) \in \mathbb{R}^k$ of $\rho \in \text{cl}^1(\mathcal{E})$ are analogues of *natural parameters*, the parameters $\langle \lambda_1, \ldots, \lambda_k \rangle \in \mathbb{R}^k$ are analogues of *canonical parameters* in statistics, see §20 in [Če82].
4.3 Local maximizers of the entropy distance

Maximization of the entropy distance from an exponential family is proposed in [Ay02] as a structuring principle in natural systems. We prove two necessary conditions for a local maximizer of the entropy distance $d_\mathcal{E}$ from an exponential family $\mathcal{E}$. One condition, an upper bounds on the rank, enforces determinism on the local maximizer. The second condition identifies local maximizers as the cutoff of their rI-projection.

The idea to Theorem 4.17 is from Proposition 3.2 in [Ay02] where the assertion is proved for a subset of the probability simplex (7) on a finite set $\Omega$. The statement was proved in Corollary 2 in [MA03] on the whole probability simplex on $\Omega$.

We shall use the fact that a unique face of $S_A$ exists, which contains a given state $\rho \in S_A$ in its relative interior (30).

**Theorem 4.17.** Let $\rho \in S_A$ be a local maximizer of the entropy distance $d_\mathcal{E}$ from $\mathcal{E}$ and assume that $F$ is the face of the state space $S_A$ which contains $\rho$ in its relative interior. Then $\dim(F) \leq \dim(A)$.

**Proof:** We consider the convex set $K := F \cap (\rho + U^\perp)$. If two convex sets $X, Y \subset \mathbb{E}^n$ in the finite-dimensional Euclidean vector space $(\mathbb{E}, (\cdot, \cdot))$ share a relative interior point, then $\text{ri}(X \cap Y) = \text{ri}(X) \cap \text{ri}(Y)$ follows, see e.g. Theorem 6.5 in [Ro72]. Hence $\rho \in \text{ri}(K)$ follows.

If $\sigma \in K$, then by Definition 4.11 of the rI-projection $\pi_\mathcal{E}$ we have $\pi_\mathcal{E}(\sigma) = \pi_\mathcal{E}(\rho)$ and Theorem 4.15 allows to rewrite the entropy distance

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

We have $p := s(\pi_\mathcal{E}(\rho)) \succeq s(\sigma)$ by Lemma 2.27 and with notation of functional calculus from Definition 2.3.3 we get

$$d_\mathcal{E}(\sigma) = -S(\sigma) - \text{tr} \sigma \log[p](\pi_\mathcal{E}(\rho)).$$

The von Neumann entropy $S(\sigma)$ is strictly concave, see e.g. §II.B in [We78]. Hence $d_\mathcal{E}(\sigma)$ as a sum of a strictly convex function and a linear function is strictly convex on $K$. Since $\rho$ is a local maximizer of $d_\mathcal{E}$ on $S_A$ it is a local maximizer of the strictly convex function $d_\mathcal{E}$ on $K$. Since in addition $\rho \in \text{ri}(K)$ holds, we get $K = \{\rho\}$. Then

$$\dim F + \dim U^\perp \leq \dim A_{sa} = \dim U + \dim U^\perp$$

follows, hence $\dim F \leq \dim U$. If we choose a parametrization of $\mathcal{E}$, such that $1 \notin U$ (e.g. by replacing $\Theta$ by $\pi_{A_0}(\Theta)$) then Proposition 4.1.1 and 4.1.3 show $\dim U = \dim \mathcal{E}$ completing the proof. \hfill $\Box$

The bound in Theorem 4.17 enjoys a quadratic improvement in $n \in \mathbb{N}$ from the commutative algebra $\mathbb{C}^n$ to the non-commutative algebra Mat($n, \mathbb{C}$).

**Remark 4.18.** In a C*-subalgebra $\mathcal{A}$ of Mat($n, \mathbb{C}$) let $\rho \in S_A$ and let $F$ be the face of $S_A$ containing $\rho$ in its relative interior. Let $p := s(\rho)$ be the support projection
of $\rho$ and let $\text{rk}(\rho)$ be the rank of $\rho$. Then Proposition 2.14 shows $F = S_{p\mathcal{A}p}$ and $\dim(S_{p\mathcal{A}p}) = \dim((p\mathcal{A}p)_s) - 1$ hence
\[
\dim(F) = \dim((p\mathcal{A}p)_s) - 1 = \dim_C(p\mathcal{A}p) - 1.
\]
If $\rho$ is a local maximizer of the entropy distance $d_\mathcal{E}$ from an exponential family $\mathcal{E}$ then Theorem 4.17 shows
\[
\dim_C(p\mathcal{A}p) = \dim(F) + 1 \leq \dim(\mathcal{E}) + 1.
\]
(58) shows
\[
\text{rk}(\rho) \leq \dim(\mathcal{E}) - 1.
\]
If $\mathcal{A} \cong \mathbb{C}^n$ is the algebra of diagonal matrices in $\text{Mat}(n, \mathbb{C})$ we have $\dim_C(p\mathcal{A}p) = \text{rk}(\rho)$ hence (58) shows
\[
\text{rk}(\rho) \leq \sqrt{\dim(\mathcal{E}) - 1}.
\]

The following Theorem 4.19 was first proved in Proposition 3.1 in [Ay02] for a subset of the probability simplex on a finite set $\Omega$ and was proved in Theorem 5.1 in [Ma07] on the whole probability simplex on $\Omega$. In this commutative setting the theorem says that a local maximizer $P$ of the entropy distance $d_\mathcal{E}$ equals the conditional probability distribution $P = Q(\cdot|A)$, see Remark 1.2.3, of its rI-projection $Q = \pi_\mathcal{E}(P)$ where $A \subset \Omega$ is the support set of $P$.

The entropy distance of a local maximizer is a difference of free energies (Definition 1.12). We use functional calculus in compressed algebras (Definition 2.3.3).

**Theorem 4.19.** Let $\rho \in S_A$ be any state, $p := s(\rho)$ be the support projection of $\rho$ and let $q := s(\pi_\mathcal{E}(\rho))$ be the support projection of the rI-projection $\pi_\mathcal{E}(\rho)$. Then there exists $\theta \in \Theta$ such that $\pi_\mathcal{E}(\rho) = R_{q\mathcal{A}q}(q\theta q)$.

1. If $u \in (p\mathcal{A}p)_s$ is a traceless matrix, then $\frac{\partial}{\partial t}d_\mathcal{E}(\rho + tu)|_{t=0} = \langle u, \log[\rho](\rho) - p\theta p \rangle$.

2. If $\rho$ is a local maximizer of the entropy distance $d_\mathcal{E}$ on the state space $S_A$, then $\rho = R_{p\mathcal{A}p}(p\theta p)$ and $d_\mathcal{E}(\rho) = F_{q\mathcal{A}q}(q\theta q) - F_{p\mathcal{A}p}(p\theta p)$.

**Proof:** The parameter $\theta \in \Theta$ and a projection $q \in \mathcal{P}^U$ such that $\pi_\mathcal{E}(\rho) = R_{q\mathcal{A}q}(q\theta q)$ exists by Definition 4.11 of the rI-projection to $\mathcal{E}$. We us notice $p \preceq q$ from Lemma 2.27. The theorem is proved in Theorem 22 in [KW11] for $q = 1$, i.e. for $\pi_\mathcal{E}(\rho)$ invertible in $\mathcal{A}$, and for $\Theta$ consisting of traceless matrices. All assertions are invariant under the substitution of $\theta \mapsto \theta + \lambda q$ for real $\lambda$, e.g.

\[
F_{q\mathcal{A}q}(q(\theta + \lambda 1)q) - F_{p\mathcal{A}p}(p(\theta + \lambda 1)p) = F_{q\mathcal{A}q}(q\theta q) + \lambda - [F_{p\mathcal{A}p}(p\theta p) + \lambda]
\]
\[
= F_{q\mathcal{A}q}(q\theta q) - F_{p\mathcal{A}p}(p\theta p).
\]

This proves our claim for arbitrary non-empty affine subspaces $\Theta \subset A_s$ if $q = 1$. Otherwise, if $q \neq 1$, then Theorem 4.15 shows $d_\mathcal{E}(\rho) = d_{\mathcal{E}_q}(\rho)$. 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 $s(\rho) = p \preceq q$. \qed
4.4 Non-commutative aspects of exponential families

Some non-commutative aspects of the I-/rI-topology are collected in Corollary 3.22. We shall now consider an exponential family $\mathcal{E}$. This leads to a continuity conjecture for the entropy distance.

The geodesic closure of $\mathcal{E}$ is

$$\text{cl}^{\text{geo}}(\mathcal{E}) := \{ \rho \in S_A \mid \rho \text{ is the limit of an e-geodesic in } \mathcal{E} \}$$

and the inclusions (with $\overline{\mathcal{E}}$ the norm closure)

$$\text{cl}^{\text{geo}}(\mathcal{E}) \subset \text{cl}^{\text{rI}}(\mathcal{E}) \subset \overline{\mathcal{E}}$$

were already proved in Corollary 12 in [KW11]. The first inclusion follows from the arguments used below in the proof of Theorem 4.20. The second inclusion follows from the Pinsker-Csiszár inequality.

We will see that strict inclusions in (60) are only possible for a non-commutative algebra $A$. Examples in $A = \text{Mat}(2, \mathbb{C}) \oplus \mathbb{C}$ are the Swallow family with $\text{cl}^{\text{geo}}(\mathcal{E}) \subsetneq \text{cl}^{\text{rI}}(\mathcal{E})$ and the Staffelberg family with $\text{cl}^{\text{rI}}(\mathcal{E}) \subsetneq \overline{\mathcal{E}}$, see §3.3 in [KW11].

**Theorem 4.20.** We have $\text{cl}^{\text{geo}}(\mathcal{E}) = \text{cl}^{\text{rI}}(\mathcal{E})$ if and only if all faces of the mean value set $M_A(U)$ are exposed faces.

**Proof:** Using (50) and Corollary 2.20 we can write the geodesic closure of $\mathcal{E}$ in the form

$$\text{cl}^{\text{geo}}(\mathcal{E}) = \bigcup_{\mathcal{P}_A \uparrow \downarrow \{0\}} \mathcal{E}_p,$$

where $\mathcal{P}_A \uparrow \downarrow$ is the exposed projection lattice defined in (31). This is also proved in Proposition 8 in [KW11]. On the other hand we have the disjoint union

$$\text{cl}^{\text{rI}}(\mathcal{E}) = \bigcup_{\mathcal{P}_A \uparrow \downarrow \{0\}} \mathcal{E}_p$$

by Theorem 4.15 and Definition 4.9. By the lattice isomorphism (32) the equality $\mathcal{P}_A \uparrow \downarrow = \mathcal{P}_A \uparrow$ is equivalent to the property that all faces of $M_A(U)$ are exposed faces. □

We discuss the norm continuity of the entropy distance (12).

**Theorem 4.21.** We have $\text{cl}^{\text{rI}}(\mathcal{E}) = \overline{\mathcal{E}}$ if and only if the entropy distance $d_\mathcal{E}$ is norm continuous on $S_A$.

**Proof:** If the inclusion $\text{cl}^{\text{rI}}(\mathcal{E}) \subset \overline{\mathcal{E}}$ in (60) is strict, then there exists a norm convergent sequence $(\rho_i)_{i \in \mathbb{N}} \subset \text{cl}^{\text{rI}}(\mathcal{E})$ with limit $\rho \in S_A \setminus \text{cl}^{\text{rI}}(\mathcal{E})$ in the compact state space (see Proposition 2.14). By Theorem 4.15 we have $d_\mathcal{E}(\rho_i) > 0$ while $d_\mathcal{E}(\rho_i) = 0$ for $i \in \mathbb{N}$ hence $d_\mathcal{E}$ is discontinuous at $\rho \in S_A$.

Conversely, let us prove that $d_\mathcal{E}$ is lower semi-continuous if $\overline{\mathcal{E}} = \text{cl}^{\text{rI}}(\mathcal{E})$. Since $\overline{\mathcal{E}}$ is a compact subset of $A_\text{sa}$, lower semi-continuity of relative entropy (Remark 2.6.1) implies lower semi-continuity of the minimum

$$S_A \to \mathbb{R}, \quad \rho \mapsto \min\{S(\rho, \sigma) : \sigma \in \overline{\mathcal{E}}\}.$$
This can be proved using a covering of $\mathcal{E}$ by open balls, see Theorem 2 on page 116 in [Be63]. This minimum function equals $d_\mathcal{E}$ by Theorem 4.15.

In order to deduce the continuity of $d_\mathcal{E}$ from its lower semi-continuity we reproduce the proof of Lemma 4.2 in [Ay02]. Let $\rho \in \mathcal{S}_A$ and $(\rho_i)_{i \in \mathbb{N}} \subset \mathcal{S}_A$ such that $\rho = \lim_{i \to \infty} \rho_i$. The function $\tau \mapsto S(\tau, \sigma)$ is norm continuous on $\mathcal{S}_A$ for every invertible state $\sigma$. Since $\mathcal{E}$ consists of invertible states we have for all $\sigma \in \mathcal{E}$
\[
\lim_{i \to \infty} S(\rho_i, \sigma) = S(\rho, \sigma)
\]
and as $d_\mathcal{E}(\rho_i) \leq S(\rho_i, \sigma)$ holds for all $i \in \mathbb{N}$ we get
\[
\limsup_{i \to \infty} d_\mathcal{E}(\rho_i) \leq \limsup_{i \to \infty} S(\rho_i, \sigma) = S(\rho, \sigma).
\]
Taking the infimum over all $\sigma \in \mathcal{E}$ and using the lower semi-continuity of $d_\mathcal{E}$ we have
\[
\limsup_{i \to \infty} d_\mathcal{E}(\rho_i) \leq d_\mathcal{E}(\rho) \leq \liminf_{i \to \infty} d_\mathcal{E}(\rho_i).
\]
This shows $\lim_{n \to \infty} S_\mathcal{E}(\rho_n) = S_\mathcal{E}(\rho)$ and proves continuity of $S_\mathcal{E}$. □

Remark 4.22. 1. If $\mathcal{A} \cong \mathbb{C}^N$ then $\mathcal{S}_A$ is a simplex, hence $\mathcal{M}_A(U)$ is a polytope and all faces of $\mathcal{M}_A(U)$ are exposed faces. Then Theorem 4.20 proves $\text{cl}^{\text{geo}}(\mathcal{E}) = \text{cl}^1(\mathcal{E})$. Moreover we have $\mathcal{T}^{\text{cl}} = \mathcal{T}^\|_\text{cl}$ by Corollary 3.22 hence $\text{cl}^1(\mathcal{E}) = \mathcal{E}$. In addition, Theorem 4.21 shows that the entropy distance $d_\mathcal{E}$ is norm continuous, a result first proved in [Ay02].

2. The two non-commutative phenomena in Theorem 4.20 and Theorem 4.21 seem to be connected: A set of two-dimensional families in $\mathcal{A} = \text{Mat}(2, \mathbb{C}) \oplus \mathbb{C}$ has revealed that non-exposed faces of the mean value set $\mathcal{M}_A(U)$ are generic in the sense that they exist in an open subset of the Grassmannian manifold of linear parameter spaces $\Theta = U$. Exponential families with a (norm) discontinuous entropy distance exist only on the boundary of this open set and are associated to the creation of non-exposed face of $\mathcal{M}_A(U)$. Connections between the two theorems may be rigorously studied by methods of convex algebraic geometry, see e.g. [RS10]. Some ideas in this direction have been collected in §2.4 in [We11] and in [We]. We are aware that non-exposed faces of $\mathcal{M}(U)$ are not generic for 2D-planes $U$ in the algebra $\text{Mat}(3, \mathbb{C})$, see [RS05].

3. The projection $\pi_U|_{\text{cl}^1(\mathcal{E})} : \text{cl}^1(\mathcal{E}) \to \mathcal{M}_A(U)$ is a bijection by Theorem 4.15. Since this is a continuous function and since its domain and range is compact, it is clear that the mean value parametrization of $\text{cl}^1(\mathcal{E})$
\[
(\pi_U|_{\text{cl}^1(\mathcal{E})})^{-1} : \mathcal{M}(U) \to \text{cl}^1(\mathcal{E})
\]
is continuous if and only if $\text{cl}^1(\mathcal{E}) = \overline{\mathcal{E}}$. The mean value parametrization
\[
(\pi_U|_\mathcal{E})^{-1} : \text{ri} \mathcal{M}(U) \to \mathcal{E}
\]
is real analytic by Corollary 4.6. Similarly, in Theorem 4.26 the assignment $\text{cs}(u_1, \ldots, u_k) \to \text{cl}^1([\text{span}(u_1, \ldots, u_k)])$, $\xi \mapsto \rho(\xi)$ can be discontinuous while the restriction $\text{ri} \text{cs}(u_1, \ldots, u_k) \to R(\text{span}(u_1, \ldots, u_k))$ is real analytic.
The norm continuity of the entropy distance in $\mathbb{C}^n$ may have a counterpart in a non-commutative algebra.

**Conjecture 4.23.** The entropy distance $d_{E}$ from $E$ is continuous for $T_{rI}$ and $T_{c}$.

**Remark 4.24** (Continuity of the entropy distance). 1. Conjecture 4.23 is not impossible because $T_{rI} \supset T_{\|\cdot\|}$ holds by Theorem 3.21.1. The finer topology $T_{rI}$ may balance the norm discontinuities found in Theorem 4.21. Secondly, in a commutative algebra the norm continuity of $d_{E}$ and the equality $T_{rI} = T_{\|\cdot\|}$, see Remark 4.22.1, is a sign of evidence for the $rI$-continuity.

2. The method of Theorem 4.21 is useless for any proof of the continuity of $d_{E}$ for $T_{rI}$ and $T_{c}$. This can be said because $cl_{rI}(E)$ is not $T_{rI}$ compact unless the mean value set $M(U)$ is a polytope. Otherwise $M(U)$ has infinitely many extreme points $\pi_{U}(F_{A}(p))$ for certain $p \in P^{U}$. For such $p$ the open $rI$-disks $V_{rI}(\frac{p}{\text{tr}(p)}, \infty)$, defined in (37), are an infinite open cover of $cl_{rI}(E)$ and each $\frac{p}{\text{tr}(p)}$ is only contained in one set of the cover.

### 4.5 The complete Pythagorean theorem

We prove the complete Pythagorean theorem for an exponential family $E$. We apply it to the maximization of the von Neumann entropy under linear constraints providing previously unknown solutions.

The entropy distance $d_{E}(\rho)$ of a state $\rho \in S_{A}$ from the exponential family $E$ equals the relative entropy from its $rI$-projection $\pi_{E}(\rho) \in cl_{rI}(E)$ to the $rI$-closure $cl_{rI}(E)$ of $E$. In addition, $d_{E}(\rho)$ is less than the relative entropy from any state $\sigma \neq \pi_{E}(\rho)$ in the $rI$-closure of $E$,

$$d_{E}(\rho) = S(\rho, \pi_{E}(\rho)) < S(\rho, \sigma).$$

This is proved in Theorem 4.15. Now we compute the value of that difference, extending the Pythagorean theorem in §1.3.

**Theorem 4.25** (Complete Pythagorean theorem). If $\rho \in S_{A}$ and if $\sigma$ lies in the $rI$-closure $cl_{rI}(E)$ of $E$, then $S(\rho, \pi_{E}(\rho)) + S(\pi_{E}(\rho), \sigma) = S(\rho, \sigma)$.

**Proof:** Theorem 4.15 shows $cl_{rI}(E) = \text{ext}(E)$. Then by Definition 4.9 of the extension $\text{ext}(E)$ there exist projections $p, q \in P^{U}$, such that $\pi_{E}(\rho) \in E_{p}$ and $\sigma \in E_{q}$. If $p \not\preceq q$ then $s(\rho) \not\preceq q$ follows from Lemma 2.27 since $q \in P^{U}$. We get

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

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

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

$$S(\pi_{E}(\rho), \pi_{E}(\rho)) + S(\pi_{E}(\rho), \sigma) = S(\pi_{E}(\rho), \sigma)$$
and continue to show that the resulting difference
\[ x := S(\rho, \pi_\mathcal{E}(\rho)) - S(\rho, \sigma) - [S(\pi_\mathcal{E}(\rho), \pi_\mathcal{E}(\rho)) - S(\pi_\mathcal{E}(\rho), \sigma)] \]
\[ = \text{tr}[\rho - \pi_\mathcal{E}(\rho)][\log[p] \sigma - \log[p] \pi_\mathcal{E}(\rho)] \]

is zero. By Definition 4.9 of \( \mathcal{E}_p \) resp. \( \mathcal{E}_q \) there exist \( \theta, \tilde{\theta} \in \Theta \) and \( y, \tilde{y} \in \mathbb{R} \), such that
\[ \log[p] \pi_\mathcal{E}(\rho) = c^p(\theta) + yp \text{ resp. } \log[q] \sigma = c^q(\tilde{\theta}) + \tilde{y}q. \]
Since \( s(\rho) \leq s(\pi_\mathcal{E}(\rho)) = p \leq q \) this gives
\[ x = \text{tr}[\rho - \pi_\mathcal{E}(\rho)][c^q(\tilde{\theta}) - c^p(\theta)] = \text{tr}[\rho - \pi_\mathcal{E}(\rho)][\tilde{\theta} - \theta]. \]

By Definition 4.11 of the projection \( \pi_\mathcal{E} \) the difference \( \rho - \pi_\mathcal{E}(\rho) \) is perpendicular to \( U \), hence \( x = 0 \). \( \square \)

We maximize the von Neumann entropy under linear constraints. The invertible solutions for \( p = 1 \) are well-known, see e.g. p. 125 in [IO97]. The non-invertible solutions depend on a projection lattice \( \mathcal{P}^U \), where the linear space \( U \) imposes the linear constraints. The lattice \( \mathcal{P}^U \) can in principle be computed by spectral analysis, see Remark 2.26. We formulate the result in terms of the mean value map \( m_{u_1,\ldots,u_k} \) and the convex support \( \text{cs}(u_1,\ldots,u_k) \), defined in (5) and (6).

**Theorem 4.26.** 1. Let \( U \subset \mathcal{A}_{sa} \) be a linear subspace of self-adjoint matrices. We consider the exponential family \( \mathcal{E} := R_\mathcal{A}(U) \). For every \( u \) in the mean value set \( \mathcal{M}_\mathcal{A}(U) \) the equality
\[ \arg\max\{S(\rho) \mid \rho \in (u + U^\perp) \cap S_\mathcal{A}\} = (\pi_\mathcal{E}|_{\mathcal{P}^U})^{-1}(u) \]
holds and \( (\pi_\mathcal{E}|_{\mathcal{P}^U})^{-1}(u) \) has cardinality one. In particular, the set of solutions to (61) is the rI-closure of \( \mathcal{E} \).

2. Let \( U := \text{span}_\mathbb{R}(u_1,\ldots,u_k) \) for some \( u_1,\ldots,u_k \in \mathcal{A}_{sa} \) and let \( \xi = (\xi_1,\ldots,\xi_k) \in \text{cs}_\mathcal{A}(u_1,\ldots,u_k) \) belong to the convex support. Then
\[ \arg\max\{S(\rho) \mid \rho \in S_\mathcal{A} \text{ such that } m_{u_1,\ldots,u_k}(\rho) = \xi\} \]
has cardinality one. There exists a unique projection \( p \in \mathcal{P}^U \) and there exist real numbers \( \beta_1,\ldots,\beta_k \) such that
\[ \left(\frac{\partial}{\partial \beta_1},\ldots,\frac{\partial}{\partial \beta_k}\right) F_{p_\mathcal{P} \mathcal{A}}(c^p(-\sum_{i=1}^k \beta_i u_i)) = -\xi \]
holds for the free energy \( F_{p_\mathcal{P} \mathcal{A}} \) on the algebra \( p_\mathcal{P} \mathcal{A} \). If \( \rho(\xi) \) denotes the maximizer in (62) then each solution \( [p, (\beta_1,\ldots,\beta_k)] \) of (63) satisfies
\[ \rho(\xi) = R_{p_\mathcal{P} \mathcal{A}}(c^p(-\sum_{i=1}^k \beta_i u_i)) \]
and its von Neumann entropy is \( S(\rho(\xi)) = F_{p_\mathcal{P} \mathcal{A}}(c^p(-\sum_{i=1}^k \beta_i u_i)) + \sum_{i=1}^k \beta_i \xi_i \).
Proof: There is a bijection \( \pi_U|_{cl^I(E)} : cl^I(E) \to M_\mathcal{A}(U) \) by Theorem 4.15. Hence we can consider the inverse \((\pi_U|_{cl^I(E)})^{-1} : M_\mathcal{A}(U) \to cl^I(E)\). If \( u \in \mathcal{M}(U) \) then in order to prove (61), it is sufficient to show for all \( \rho \in (u + U^\perp) \cap \mathcal{S}_\mathcal{A} \) with \( \rho \neq (\pi_U|_{cl^I(E)})^{-1}(u) \), that

\[
S\left((\pi_U|_{cl^I(E)})^{-1}(u)\right) > S(\rho).
\]

Since \( \pi_U(\rho) = u \) we have \((\pi_U|_{cl^I(E)})^{-1}(u) = \pi_E(\rho)\). Since \( \frac{1}{t+1} \in E = R_\mathcal{A}(U) \), Theorem 4.25 shows

\[
S(\rho, \pi_E(\rho)) + S(\pi_E(\rho), \frac{1}{t+1}) = S(\rho, \frac{1}{t+1}).
\]

Hence \( S(\sigma, \frac{1}{t+1}) = \log \text{tr} \mathbb{1} - S(\sigma) \) for \( \sigma \in \mathcal{S}_\mathcal{A} \) shows

\[
S(\rho) = S(\pi_E(\rho)) - S(\rho, \pi_E(\rho)).
\]

The state \( \pi_E(\rho) \) is the maximizer since the non-negative relative entropy \( S(\rho, \pi_E(\rho)) \) is only zero for \( \rho = \pi_E(\rho) \). It is clear from the construction that the set of maximizers is the \( rI \)-closure \( cl^I(E) \).

The coordinate formulation (62) follows from Remark 1.2.5 stating \( m_{u_1,\ldots,u_k} = m_{u_1,\ldots,u_k} \circ \pi_U \) and that \( m_{u_1,\ldots,u_k}|_{\mathcal{M}(U)} : \mathcal{M}(U) \to cs(u_1,\ldots,u_k) \) is a bijection.

The differential condition (63) on the coefficients \([\beta_1,\ldots,\beta_k]\) of the maximizer \( \rho := \rho(\xi) \) and its exponential expression (64) are proved in Corollary 4.16 (up to the sign). Let \( \rho = R_{\nu A}(c^p(-\sum_{i=1}^k \beta_i u_i)) \). Then we have

\[
-\text{tr} \rho[c^p(-\sum_{i=1}^k \beta_i u_i)] = \sum_{i=1}^k \beta_i \langle u_i, \rho \rangle = \sum_{i=1}^k \beta_i \xi_i.
\]

The von Neumann entropy \( S(\rho) = -\text{tr} \rho \log(\rho) \) simplifies to

\[
S(\rho) = -\text{tr} \rho[c^p(-\sum_{i=1}^k \beta_i u_i)] - F_{\nu A}(c^p(-\sum_{i=1}^k \beta_i u_i))
\]

\[
= F_{\nu A}(c^p(-\sum_{i=1}^k \beta_i u_i)) + \sum_{i=1}^k \beta_i \xi_i
\]

by an expansion of the exponential expression (64) under the logarithm \( \log[\rho] \). \( \square \)

5 Comment on the representation

We now consider an arbitrary finite-dimensional C*-algebra \( \mathcal{A} \). We show that the \( I/ri \)-topology is independence of the representation. In order that exponential families can be defined in \( \mathcal{A} \) we have to use a representation of \( \mathcal{A} \) as a C*-subalgebra of \( \text{Mat}(n, \mathbb{C}) \) but the complete projection theorem and Pythagorean theorem are independent of this choice. The maximization of the von Neumann entropy does depend on the representation of \( \mathcal{A} \) as a C*-subalgebra of \( \text{Mat}(n, \mathbb{C}) \).

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

\[
\Phi : \mathcal{B} \to \mathcal{A}
\]

between two unital C*-algebras [Uh77], 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
A C*-isomorphism then \( \Phi^{-1} \) provides the opposite inequality and \( S(\Phi^*(f), \Phi^*(g)) = S(f, g) \) follows. In particular, our results about the \( I/rI \)-topology in §3.2 are valid in any finite-dimensional C*-algebra independent of the representation.

The second object, an exponential family, can be defined if the finite-dimensional C*-algebra \( A \) is represented by an algebra of linear operators on a Hilbert space. The normal state space of \( A \) consists of all positive and normalized trace class operators \( \rho \) in \( A \), with associated functional (see Theorem 2.4.21 in [Br87], also Remark 1.11)

\[
\rho \mapsto \langle a, \rho \rangle = \text{tr}(a\rho), \quad (a \in A).
\]

Elements \( \rho \) in an exponential family are defined by functional calculus but the normalization condition restricts the possible representations of \( A \). E.g. the one-dimensional algebra \( \mathbb{C} \) 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.

Every finite-dimensional C*-subalgebra is, according to Theorem III.1.1 in [Da96], C*-isomorphic to the direct sum

\[
B := \bigoplus_{i=1}^N \text{Mat}(k_i, \mathbb{C}),
\]

where \( N \in \mathbb{N}_0 \) and \( k \in \mathbb{N}^N \) is a multi-index. Up to unitary equivalence, any C*-algebra \( A \) of linear operators which is C*-isomorphic to \( B \), has the form

\[
A := \bigoplus_{i=1}^N \{ \bigoplus_{j=1}^{m_i} a_i \mid a_i \in \text{Mat}(k_i, \mathbb{C}) \} \oplus 0_l
\]

with cardinalities \( m_i \geq 1 \) for \( i = 1, \ldots, N \) and \( l \geq 0 \), see Corollary III.2.1 in [Da96], and there is a C*-isomorphism \( \Phi : B \to A \)

\[
\Phi(\bigoplus_{i=1}^N b_i) = \bigoplus_{i=1}^N \bigoplus_{j=1}^{m_i} b_i \oplus 0_l, \quad (b_1, \ldots, b_N) \in B.
\]

Only matrices that vanish on each summand of infinite multiplicity \( m_i \) can be normalized so exponential families can only be defined in an algebra \( A \) with \( k_1 m_1 + \cdots + k_N m_N < \infty \). Hence we shall assume \( A \subset \text{Mat}(n, \mathbb{C}) \) for \( n \in \mathbb{N} \) in the following.

Still, we are interested how different representations of \( A \) as a C*-subalgebra of \( \text{Mat}(n, \mathbb{C}) \) affect our results about exponential families. Let us begin with mean values. We had preferred the mean value set (4) to the isomorphic convex support support (6) because of the simplicity of the Hilbert-Schmidt Euclidean geometry, see Remark 1.2.5. Convex support has other advantages. Firstly, it is equivariant under the isomorphism (66): \( f(b) = (\Phi^{-1})^* f(\Phi(b)) \) holds for all states \( f \) on \( B \) and \( b \in B \). The mean value set is not equivariant. Another advantage of the convex support is that the decomposition of Lemma 2.23 becomes a simple inclusion \( cs_{pA^p} \subset cs_A \).

**Lemma 5.1.** Let \( u_1, \ldots, u_k \in A_{sa} \), put \( U := \text{span}_a(u_1, \ldots, u_k) \) and let \( p \in \mathcal{P}^U \). Then the following diagram commutes.

\[
\begin{array}{cccccc}
\mathbb{F}(p) & \xrightarrow{\pi_U} & \pi_U(\mathbb{F}(p)) & \xrightarrow{m_{u_1,\ldots,u_k}} & m_{u_1,\ldots,u_k}(\mathbb{F}(p)) & \subset & cs_A(u_1, \ldots, u_k) \\
\downarrow \pi_{\mathcal{P}(U)} \quad & & & & & & \\
M_{pA^p}(\mathcal{P}(U)) & \xrightarrow{m_{\mathcal{P}(u_1),\ldots,\mathcal{P}(u_k)}} & \mathcal{C}_{pA^p}(\mathcal{P}(u_1), \ldots, \mathcal{P}(u_k)) & \quad & \end{array}
\]
**Proof:** We extend the second diagram in Lemma 2.23. By Remark 1.2.5 we have $m_{u_1,\ldots,u_k} \circ \pi_U = m_{u_1,\ldots,u_k}$ and $m_{u_1,\ldots,u_k}|M(U) : M(U) \to cs_{A}(u_1,\ldots,u_k)$ is a bijection. In the algebra $pA\rho$ this means $m_{c\rho(u_1),\ldots,c\rho(u_k)} \circ \pi_{c\rho(U)} = m_{c\rho(u_1),\ldots,c\rho(u_k)}$ and

$$m_{c\rho(u_1),\ldots,c\rho(u_k)}|M(c\rho(U)) : M(c\rho(U)) \to cs_{pA\rho}(c\rho(u_1),\ldots,c\rho(u_k))$$

is a bijection. For all $a \in pA\rho$ and $i = 1,\ldots,k$ we have

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

hence $m_{u_1,\ldots,u_k}(a) = m_{c\rho(u_1),\ldots,c\rho(u_k)}(a)$ holds and completes the proof. \hfill $\Box$

Next we consider exponential families. The adjoint $\Phi^* : A^* \to B^*$ of (66) is given for $F_i \in \text{Mat}(k_i, \mathbb{C})$, $i = 1,\ldots,N$, by

$$\Phi^*(\bigoplus_{i=1}^{N} m_i F_i \oplus 0_i) = \bigoplus_{i=1}^{N} m_i F_i.$$

(67)

**Lemma 5.2.** Let $\Theta \subset B_{sa}$ be a non-empty affine subspace and $E := R_{B}(\Theta)$. Let $\theta_0 := \bigoplus_{i=1}^{N} \ln(m_i)1_{k_i} \in B_{sa}$. Then the affine space $\Theta := \Phi(\Theta - \theta_0) \subset A_{sa}$ satisfies $(\Phi^*)^{-1}(E) = R_{A}(\Theta)$.

**Proof:** By (67) we have $\Phi^* \circ R \circ \Phi(\theta) = R(\theta + \bigoplus_{i=1}^{N} \ln(m_i)1_{k_i})$ for $\theta \in \Theta$. \hfill $\Box$

Lemma 5.2 shows that the class of exponential families is preserved under the isomorphism (66). Clearly $\Phi^*(\rho + U^\perp) = \Phi^*(\rho) + \Phi^{-1}(U)^\perp$ holds for all $\rho \in S_{A}$ and $U \subset A_{sa}$. We conclude that the complete Pythagorean theorem and the complete projection theorem (Theorem 4.15 and Theorem 4.25) are valid for finite-dimensional C*-algebras independent of the representation.

We finish with the negative result that the von Neumann entropy maximization in Theorem 4.26 is not equivariant under the isomorphism (66). The reason is that $\Theta$ in Lemma 5.2 is not necessarily a linear space even though $\Theta$ is a linear space.

**Example 5.3.** The unconstrained maximum of the von Neumann entropy in the algebra $B = \mathbb{C}^2$ is $\log(2)$ while for $n \geq 3$ and $A = \mathbb{C} \oplus \{(\lambda,\ldots,\lambda) \in \mathbb{C}^{n-1} | \lambda \in \mathbb{C}\}$ the maximum is $\log(n)$. The second maximum is assumed uniquely at $(\frac{1}{n},\ldots,\frac{1}{n})$ and $\Phi^*(\frac{1}{n},\ldots,\frac{1}{n}) = (\frac{1}{n},\frac{n-1}{n})$ has the von Neumann entropy

$$S(\frac{1}{n},\frac{n-1}{n}) = \log(2) - \frac{1}{n} \log(\frac{2(n-1)}{n}(n-1)) < \log(2)$$

by the Bernoulli inequality. Moreover $S(\frac{1}{n},\frac{n-1}{n}) \xrightarrow{n \to \infty} 0$ holds by the continuity of the von Neumann entropy.

**Acknowledgment:** The complete projection theorem has grown in the years 2006–2009 in numerous discussions with Andreas Knauf who was the advisor of the PhD-thesis [We10]. Nihat Ay has brought [Ay02] to our attention. František Matuš has introduced us to sequential convergences. We thank Arleta Szkoła for many helpful and critical remarks about the final form of this article.
References

[AS01] E. M. Alfsen and F. W. Shultz, *State Spaces of Operator Algebras*, Birkhäuser (2001).

[Am01] S. Amari, *Information Geometry on Hierarchy of Probability Distributions*, IEEE Trans. Inf. Theory vol. 47 no. 5 1701–1711 (2001).

[AN00] S. Amari and H. Nagaoka, *Methods of Information Geometry*, AMS Translations of Mathematical Monographs vol. 191 (2000).

[Ay02] N. Ay, *An Information-Geometric Approach to a Theory of Pragmatic Structuring*, Ann. Probab. vol. 30 416–436 (2002).

[Ba78] O. Barndorff-Nielsen, *Information and Exponential Families in Statistical Theory*, John Wiley & Sons New York (1978).

[Be09] F. Benatti, *Dynamics, Information and Complexity in Quantum Systems*, Springer-Verlag (2009).

[BZ06] I. Bengtsson and K. Życzkowski, *Geometry of Quantum states. An Introduction to Quantum Entanglement*, Cambridge University Press (2006).

[Be63] C. Berge, *Topological Spaces*, Oliver and Boyd Ltd (1963).

[Bh97] R. Bhatia, *Matrix Analysis*, Springer-Verlag (1997).

[Bi73] G. Birkhoff, *Lattice Theory*, AMS Colloquium Publications 3rd. ed. (1973).

[Br87] O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics 1*, 2nd ed. Springer (2002).

[BS05] I. Bjelaković, J.-D. Deuschel, T. Krüger, R. Seiler, R. Siegmund-Schultze and A. Szkoła, *A Quantum Version of Sanov’s Theorem*, Comm. Math. Phys. vol. 260 659–671 (2005).

[Cs67] I. Csiszár, *On topological properties of f-divergences*, Stud. Sci. Math. Hungar. vol. 2 329–339 (1967).

[CM03] I. Csiszár and F. Matúš, *Information Projections Revisited*, IEEE Trans. Inf. Theory vol. 49 1474–1490 (2003).

[CM04] I. Csiszár and F. Matúš, *On Information Closures of Exponential Families: A Counterexample*, IEEE Trans. Inform. Theory vol. 50 no. 5 922–924 (2004).

[CM05] I. Csiszár and F. Matúš, *Closures of Exponential Families*, Ann. Probab. vol. 33 582–600 (2005).

[Če82] N. N. Čencov, *Statistical Decision Rules and Optimal Inference*, AMS Translations of Mathematical Monographs vol. 53 (1982).
REFERENCES

[Da96] K. R. Davidson, *C*-Algebras by Example*, Fields Institute Monographs vol. 6 (1996).

[Du64] R. M. Dudley, *On Sequential Convergence*, Trans. Amer. Soc. vol. 112 483–507 (1964); correction ibid. vol. 148 (1970).

[Du98] R. M. Dudley, *Consistency of M-Estimators and One-Sided Bracketing*, Progr. Probab. vol. 43 33–58, Birkhäuser (1998).

[DS58] N. Dunford and J. T. Schwartz, *Linear Operators*, I General Theory, Interscience Publishers London (1958).

[En89] R. Engelking, *General Topology*, Sigma Series in Pure Mathematics vol. 6, Heldermann (1989).

[Gr03] B. Grünbaum, *Convex Polytopes*, Springer-Verlag 2nd. ed. (2003).

[Ha87] P. R. Halmos, *Finite-Dimensional Vector Spaces*, Springer-Verlag (1987).

[Ha07] P. Harremoës, *Information Topologies With Applications*, Entropy, Search, Complexity vol. 16 of Bolyai Society Mathematical Studies (2007).

[Ho11] A. S. Holevo, *Probabilistic and statistical aspects of quantum theory*, 2nd ed. Edizioni della Normale (2011).

[IO97] R. S. Ingarden, A. Kossakowski and M. Ohya, *Information Dynamics and Open Systems*, Kluwer Academic Publishers Group (1997).

[Ka95] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag (1995).

[KP02] S. G. Krantz and H. R. Parks, *A Primer of Real Analytic Functions*, Birkhäuser (2002).

[Li73] E. H. Lieb, *Convex Trace Functions and the Wigner-Yanase-Dyson Conjecture*, Adv. Math. vol. 11 267–188 (1973).

[MA03] F. Matúš and N. Ay, *On maximization of the information divergence from an exponential family*, Proceedings of WUPES'03, University of Economics Prague 199–204 (2003).

[Ma07] F. Matúš, *Optimality conditions for maximizers of the information divergence from an exponential family*, Kybernetika vol. 43 731–746 (2007).

[Mu90] G. J. Murphy, *C*-Algebras and Operator Theory*, Academic Press (1990).
[NC00] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press (2000).

[Pe08] D. Petz, *Quantum Information Theory and Quantum Statistics*, Springer-Verlag (2008).

[RS80] M. Reed and B. Simon, *Methods of modern mathematical physics I*, 2nd ed., Academic Press, Inc. (1980).

[Ro72] R.T. Rockafellar, *Convex Analysis*, Princeton University Press (1972).

[RS05] L. Rodman and I.M. Spitkovsky, 3 × 3-matrices with a flat portion on the boundary of the numerical range, Lin. Alg. Appl. vol. 397 193–207 (2005).

[RS10] P. Rostalski and B. Sturmfels, *Dualities in Convex Algebraic Geometry*, http://arxiv.org/abs/1006.4894

[Ru99] D. Ruelle, *Statistical mechanics. Rigorous results*, World Scientific Publishing (1999).

[Sc93] R. Schneider, *Convex Bodies: The Brunn-Minkowski Theory*, Cambridge University Press (1993).

[Si52] W. Sierpiński, *General topology*, Transl. by C. Cecilia Krieger, Mathematical expositions vol. 7 University of Toronto Press (1952).

[Uh77] A. Uhlmann, *Relative Entropy and the Wigner-Yanase-Dyson-Lieb Concavity in an Interpolation Theory*, Commum. Math. Phys. vol. 54 21–32 (1977).

[We78] A. Wehrl, *General Properties of Entropy*, Rev. Mod. Phys. vol. 50 no. 2 (1978).

[We10] S. Weis, *Exponential Families with Incompatible Statistics and Their Entropy Distance*, PhD Dissertation, University of Erlangen (2010).

[We11] S. Weis, *Quantum Convex Suppprt*, Lin. Alg. Appl. vol. 435 3168–3188 (2011). correction: From the Editor-in-Chief, ibid. vol. 436 no. 1 (2012).

[We12] S. Weis, *A Note on Touching Cones and Faces*, Journal of Convex Analysis vol. 19 no. 2 (2012).

[We] S. Weis, *Duality of non-exposed faces*, http://arxiv.org/abs/1107.2319

[We00] D. Werner, *Funktionalanalysis*, 3rd ed. Springer-Verlag (2000).

[Wi63] E.H. Wichmann, *Density matrices arising from incomplete measurements*, J. Math. Phys. vol. 4 884–896 (1963).