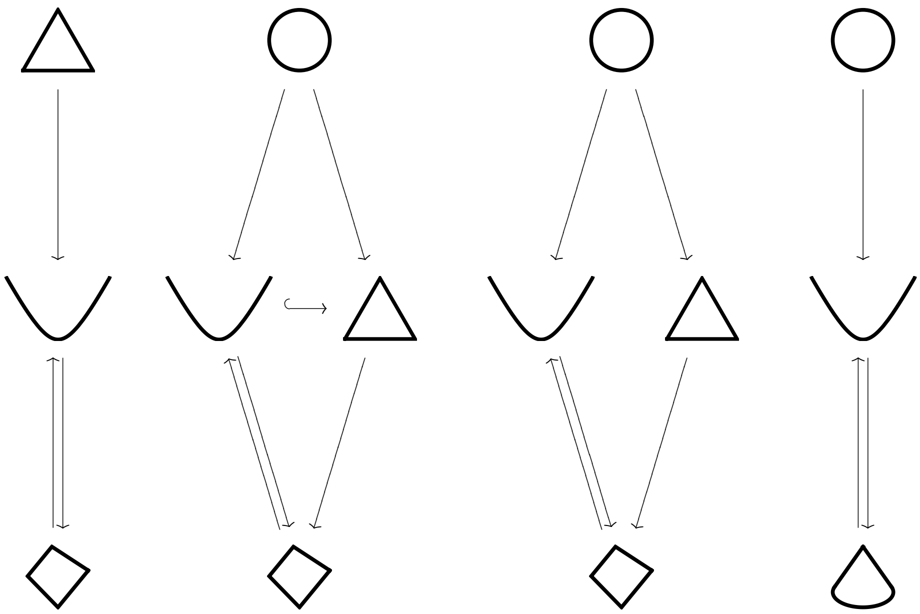
Friedrich-Alexander-Universität
Erlangen-Nürnberg

Exponential Families with Incompatible Statistics and Their Entropy Distance



The diagrams on the cover page (from left to right) sketch the structure of entropy distance from an exponential family in different settings, starting with probability distributions, over quantum systems with compatible statistics to quantum systems with incompatible statistics. More details are explained in Section 1.5.

Exponential Families with Incompatible Statistics and Their Entropy Distance

Exponentialfamilien mit inkompatibler Statistik und ihr Entropieabstand

Der Naturwissenschaftlichen Fakultät
der Friedrich-Alexander-Universität Erlangen-Nürnberg
zur
Erlangung des Doktorgrades Dr. rer. nat.

vorgelegt von
Stephan Wilhelm Weis

aus Lichtenfels

Als Dissertation genehmigt von der Naturwissen-
schaftlichen Fakultät der Friedrich-Alexander-Universität
Erlangen-Nürnberg

Tag der mündlichen Prüfung:	21. Dezember 2009
Vorsitzender der Promotionskommission:	Prof. Dr. Eberhard Bänsch
Erstberichterstatter:	Prof. Dr. Andreas Knauf
Zweitberichterstatter:	Prof. Dr. Hermann Schulz-Baldes
Zusätzlicher Gutachter:	Ing. František Matúš, CSc.

Abstract

This work generalizes the interaction measure of multi-information known in probability theory to finite-level quantum systems. This is done in the more general context of the entropy distance from an exponential family.

One of the most well-known measures of stochastic dependence is multi-information, applied in various fields including Neuroscience and Statistical Mechanics. Multi-information is the entropy distance from an exponential family. In statistics, an exponential family is a very familiar parametric model, it admits a simple geometric description of entropy distance and maximum-likelihood estimation by mean values. But a complete description of these concepts requires to study extensions of a family, an investigation that was started by N. N. Čencov and O. Barndorff-Nielsen and continued principally by I. Csiszár and F. Matúš.

Very little is known about extensions of an exponential family for a finite-level quantum system. In this thesis we consider mean value parameters of the statistic of such a family and a suitable extension thereof. Generalizing probability theory, we prove that the parameters describe the entropy distance from the family and they parametrize its *rI*-closure consisting of the points approximated in relative entropy. The dimension function of a local maximizer of entropy distance is bounded by the dimension of the family. We show that the *rI*-closure of a Gibbs family consists of the maximum entropy ensembles.

A new and generic phenomenon of a non-abelian exponential family is the appearance of non-exposed faces of the convex mean value set. We prove that mean values of the closure of an *e*-geodesic included in the family meet only the relative interior of exposed faces. Unlike in finite probability spaces there are examples with a discontinuous entropy distance, the continuity being equivalent to equality of *rI*-closure and topological closure of the family. Examples suggest that the topology of an exponential family is related to the topology of associated projector lattices and to open projections and symmetrizations of state spaces.

We conclude that multi-information for a quantum system is a continuous function equal to the entropy distance from a factorizable family. For analysis of a factorizable family we supply a partial classification of convex exponential families. As a perspective to a dynamical situation we examine a measure of temporal interaction for abelian systems, which is related to multi-information.

Zusammenfassung

Diese Arbeit verallgemeinert das aus der Wahrscheinlichkeitstheorie bekannte Interaktionsmaß der Multi-Information auf endlichdimensionale Quantensysteme. Dies geschieht im allgemeineren Rahmen des Entropieabstandes zu einer Exponentialfamilie.

Eines der bekanntesten Maße für stochastische Abhängigkeit ist die Multi-Information mit Anwendungen in Gebieten wie den Neurowissenschaften oder der Statistischen Mechanik. Multi-Information ist der Entropieabstand zu einer Exponentialfamilie. In der Statistik ist eine Exponentialfamilie ein vertrautes parametrisches Modell mit einfacher geometrischer Beschreibung von Entropieabstand und Maximum-Likelihood Schätzung durch Mittelwerte. Aber eine vollständige Beschreibung dieser Begriffe verlangt die Betrachtung von Erweiterungen einer Familie. Deren Erforschung wurde begonnen von N. N. Čencov und O. Barndorff-Nielsen und fortgeführt hauptsächlich von I. Csiszár and F. Matúš.

Nur sehr wenig ist bekannt über Erweiterungen einer Exponentialfamilie bei endlichdimensionalen Quantensystemen. In dieser Arbeit betrachten wir Mittelwertparameter der Statistik einer solchen Familie und geeignete Erweiterungen davon. In Verallgemeinerung der Wahrscheinlichkeitstheorie beweisen wir, dass die Parameter den Entropieabstand zur Familie beschreiben und sie ihren rI-Abschluss, bestehend aus den Punkten approximiert in relativer Entropie, parametrisieren. Die Dimensionsfunktion eines lokalen Maximierers des Entropieabstandes ist durch die Dimension der Exponentialfamilie beschränkt. Wir zeigen, dass der rI-Abschluss einer Gibbsfamilie aus den Maximum-Entropie Ensembles besteht.

Ein neues und generisches Phänomen bei einer nicht-abelschen Exponentialfamilie ist das Auftreten von nicht-exponierten Seiten der konvexen Menge der Mittelwerte. Wir zeigen, dass Mittelwerte des Abschlusses einer e-Geodäten in der Familie nur im relativ Inneren von exponierten Seiten liegen. Anders als bei endlichen Wahrscheinlichkeitsräumen gibt es Beispiele mit unstetigem Entropieabstand, die Stetigkeit ist äquivalent zur Gleichheit von rI-Abschluss und topologischem Abschluss der Familie. Beispiele deuten Verbindungen an von der Topologie einer Exponentialfamilie zur Topologie zugehöriger Projektorenverbände und zu offenen Projektionen und Symmetrisierungen von Zustandsräumen.

Wir folgern, dass Multi-Information für Quantensysteme eine stetige Funktion ist und dem Entropieabstand zu einer Produktfamilie gleicht. Zur Untersuchung einer Produktfamilie stellen wir eine teilweise algebraische Klassifikation von konvexen Exponentialfamilien bereit. Als Ausblick auf eine dynamische Situation betrachten wir ein Maß für die zeitliche Interaktion in abelschen Systemen, das verwandt ist mit Multi-Information.

to Margareta Monica

Acknowledgments

Being able to write this thesis in the Mathematical Physics group at Friedrich-Alexander-Universität Erlangen-Nürnberg was a great opportunity. In the first place I want to thank my supervisor Prof. Andreas Knauf for the constant and friendly support in conversations about ideas and literature both in mathematics and physics and for his encouragement to a systematic problem solving.

Luckily the opportunity turned into an exciting and substantial project with a grant at the DFG group "Geometry and Complexity in Information Theory". The regular workshops were very instructive and I enjoyed the lively interchange. Thanks to Dr. Nihat Ay in Leipzig, Prof. Andreas Knauf in Erlangen, Prof. Burkhard Kümmerer in Darmstadt, Prof. Ruedi Seiler in Berlin and their teams, who did an excellent organization.

Studies on multi-information by Dr. Nihat Ay and Prof. Andreas Knauf [AK] in 2006 convinced me to write a quantum generalization of entropy distance from an exponential family. The results and questions are printed in this thesis. Thanks to both authors for many discussions about stochastic interaction and information theory. Special thanks to Dr. Nihat Ay for excursions to his "Information Theory of Cognitive Systems Group" at Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig. I always felt very welcome.

For stimulating discussions or correspondence I wish to thank Prof. Achim Clausing, Dr. Gabriele Eichfelder, Dr. Sylvain Golenia, Dr. Boris Gutkin, Dr. Markus Müller, Prof. Rolf Schneider and Dr. Arleta Szkoła. I am very pleased to have found much help in correcting errors in the thesis, thanks to Prof. Jon Harrison, Gabriele Mühlig, Dr. Markus Müller, Dr. Ina Voigt and Dr. Brian Winn. I kindly thank Ing. František Matúš, CSc. and Prof. Hermann Schulz-Baldes for their expert report and Prof. Wulf-Dieter Geyer, Prof. Gerhard Keller and Prof. Andreas Knauf for their generous oral examinations.

Finally I want to thank my parents Magdalena and Josef and my brothers and sisters Anna, Maria, Wolfgang and Alexander a lot for helping me to relax during times of necessary recreation.

Erlangen, 11. January 2010
Stephan Weis

Contents

Page

Figures

Page

1 Introduction

Measures of stochastic dependence are used in the field of Neural Networks to describe correlations between neural spikes, see Linsker [Lin] or Tononi et al. [To]. Infomax principles are regarded as fundamental structuring rules where the typical result is the emergence of determinism through maximization of a stochastic interaction measure. Multi-information is a very prominent measure of stochastic interaction. We will introduce this measure in Section 1.1 after a few remarks on independence. The emergence of determinism through maximization of multi-information is confirmed, see Ay and Knauf [AK]. As a second area to apply multi-information we mention Statistical Mechanics, see e.g. Ruelle [Ru]. It is believed that stochastic dependencies are large at a point of a phase coexistence of a thermodynamic system. Matsuda et al. [Ma] and Erb and Ay [Er] have proved for Ising systems that multi-information has large values and singular derivatives at a point of a phase coexistence.

Multi-information is the relative entropy distance D from independent systems. These independent systems form an exponential family $\mathcal{E}$. For a probability distribution p representing a given system, we can write multi-information as the entropy distance (1.4)

$$D_{\mathcal{E}}(p) := \inf_{q \in \mathcal{E}} D(p||q),$$

see Amari [Am01] or Geiger, Meek and Sturmfels [Ge]. An example including a maximizer of multi-information is depicted in Figure 1.1. The geometry of exponential families of probability distributions and the geometry of entropy distance from an exponential family are recalled in Section 1.2 and in Section 1.3. The pioneers in this field are Csiszár and Matúš [Cs03].

From the Statistical Mechanics point of view, it is natural to study composite quantum systems, see e.g. Bratteli and Robinson [Bra]. This leads us to the question how multi-information and entropy distance generalize to quantum systems. In this thesis we treat the finite-dimensional case. Quantum systems are introduced in Section 1.4. In Section 1.5 we discuss special cases of entropy distance from a quantum exponential family. The chapter closes with an overview of the whole thesis and with a questionnaire.

1.1 Measures of stochastic dependence

We want to quantify stochastic dependence between stochastic units so we start by recalling what independence means. For our context finite probability spaces will suffice. For general notions see Bauer [BaW], Chapt. II.6., II.9. The event space when throwing a fair die is $\{1,2,3,4,5,6\}$ with the outcome governed by the uniform probability distribution. Two events $A, B \subset \{1,2,3,4,5,6\}$ occur with probabilities $P(A) = \frac{|A|}{6}$ and $P(B) = \frac{|B|}{6}$. If $P(B) > 0$ then the conditional probability of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

This is interpreted as the probability of A if it is known that B has occurred [BaW]. For example we can ask what is the probability of obtaining a 3 when the outcome is an odd number. We take $A = \{3\}$, $B = \{1,3,5\}$ and obtain $P(A|B) = \frac{1}{3}$ which is larger than the probability $\frac{1}{6}$ to obtain 3 without the information that the outcome is odd. It can happen that additional information from B may have no influence on A. Then $P(A) = P(A|B)$ or likewise

$$P(A \cap B) = P(A)P(B).$$

As an example we can ask what is the probability of obtaining a result greater than 2 if the outcome is known to be odd. Taking $A = \{3,4,5,6\}$ and $B = \{1,3,5\}$ one obtains equality $P(A \cap B) = \frac{1}{3} = P(A)P(B)$. In this case we say the events A and B are stochastically independent. More generally, three events $A, B, C \subset \{1,2,3,4,5,6\}$ are stochastically independent if

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

and similarly the concept of independence is applied to a larger number of events.

Let us make precise what is understood by independence of multiple experiments, that may be run in parallel. We consider three stochastic units with finite configuration spaces Ω_1, Ω_2 and Ω_3. The whole experiment has configuration space $\Omega := \Omega_1 \times \Omega_2 \times \Omega_3$ and a realization is governed by a probability distribution p on Ω with probability $P(A) = \sum_{a \in A} p(a)$ for an event $A \subset \Omega$. The probability $P_1(A_1)$ of observing in unit 1 an event $A_1 \subset \Omega_1$ is the probability of the local event $\widetilde{A_1} := A_1 \times \Omega_2 \times \Omega_3$,

$$P_1(A_1) = P(\widetilde{A_1}),$$

the unit 1 being governed by the marginal probability distribution $p_1(\omega_1) = P_1(\{\omega_1\})$ for $\omega_1 \in \Omega_1$. With the analogue definitions for unit 2 and 3, it is known [BaM] that the local events

$$\widetilde{A_1}, \widetilde{A_2} \quad \text{and} \quad \widetilde{A_3}$$

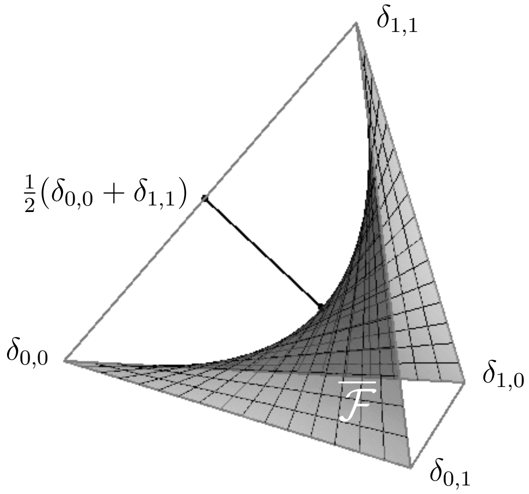


Figure 1.1: The set of factorizable probability distributions $\overline{\mathcal{F}}$ on two units $\Omega_1 = \Omega_2 = \{0, 1\}$ is depicted within the probability simplex having the Dirac measures $\delta_{\sigma_1,\sigma_2}$ concentrated on (σ_1, σ_2) for $\sigma_i \in \Omega_i$ and $i = 1, 2$ as extreme points. The distribution $\frac{1}{2}(\delta_{0,0}+\delta_{1,1})$ is a global maximizer of multi-information.

are stochastically independent for arbitrary $A_i \subset \Omega_i$ $(i = 1, 2, 3)$, only for a probability distribution p on Ω satisfying

$$p(\omega_1, \omega_2, \omega_3) = p_1(\omega_1)p_2(\omega_2)p_3(\omega_3) \tag{1.1}$$

for $\omega_i \in \Omega_i$ and $i = 1, \ldots, 3$. We have arrived at the concept of stochastic independence for stochastic units: local events in distinct units are stochastically independent. This is a special case of independence for families of events, see [BaW] Def. 6.2. A probability distribution p satisfying (1.1) is called factorizable. The set of factorizable probability distributions is denoted $\overline{\mathcal{F}}$. An example of $\overline{\mathcal{F}}$ is depicted in Figure 1.1 with two units $\Omega_1 = \Omega_2 = \{0, 1\}$ that could represent two independently thrown loaded coins.

The first method used to quantify a stochastic dependence between units is to measure correlations between pairs of units. Let us consider two units with configuration spaces Ω_1, Ω_2 and $\Omega := \Omega_1 \times \Omega_2$. Observations can be made through a random variable $X : \Omega \to \mathbb{R}$. Local random variables observe only one sub-system. We use for $\sigma_1 \in \Omega_1$ and $\sigma_2 \in \Omega_2$ the local random variables $(\omega_1 \in \Omega_1,\ \omega_2 \in \Omega_2)$

$$X_{\sigma_1}(\omega_1, \omega_2) := \begin{cases} 1 & \text{if } \omega_1 = \sigma_1, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{and} \qquad Y_{\sigma_2}(\omega_1, \omega_2) := \begin{cases} 1 & \text{if } \omega_2 = \sigma_2, \\ 0 & \text{otherwise.} \end{cases}$$

The expectation of $X : \Omega \to \mathbb{R}$ with respect to a probability distribution p on Ω is defined by

$$\mathbb{E}_p[X] := \sum_{\omega \in \Omega} p(\omega) X(\omega).$$

The expectations of local random variables are the unit marginals, one has $\mathbb{E}_p(X_{\sigma_1}) = p_1(\sigma_1)$ and $\mathbb{E}_p(Y_{\sigma_2}) = p_2(\sigma_2)$ for $\sigma_1 \in \Omega_1$, $\sigma_2 \in \Omega_2$. The covariance of two random variables $X, Y : \Omega \to \mathbb{R}$ with respect to p is

$$\mathrm{Cov}_p[X, Y] := \mathbb{E}_p \left[(X - \mathbb{E}_p[X])(Y - \mathbb{E}_p[Y]) \right].$$

The covariance of local random variables for $\sigma_1 \in \Omega_1$ and $\sigma_2 \in \Omega_2$ is

$$\mathrm{Cov}_p[X_{\sigma_1}, Y_{\sigma_2}] = p(\sigma_1, \sigma_2) - p_1(\sigma_1) p_2(\sigma_2)$$

so the vanishing covariance of local random variables is equivalent to stochastic independence of the two units: a sum of moduli of these covariances can be used to quantify stochastic dependence.

We must keep in mind that covariances between local random variables do not capture all statistical dependencies between units if there are more than two units. As an example we consider the case of three units $\Omega_i := \{0,1\}$ for $i = 1,2,3$ with p the probability distribution on $\Omega := \{0,1\} \times \{0,1\} \times \{0,1\}$ being identically distributed on the event that an even number of 1's is observed. The local random variables $X(\omega_1, \omega_2, \omega_3) := \omega_1$, $Y(\omega_1, \omega_2, \omega_3) := \omega_2$ and $Z(\omega_1, \omega_2, \omega_3) := \omega_3$ for $\omega_i \in \Omega_i$ and $i = 1,2,3$ have the expectations $\mathbb{E}_p[X] = \mathbb{E}_p[Y] = \mathbb{E}_p[Z] = \frac{1}{2}$ and $\mathbb{E}_p[XY] = \mathbb{E}_p[YZ] = \mathbb{E}_p[ZX] = \frac{1}{4}$ so the covariances vanish,

$$\mathrm{Cov}_p[X, Y] = \mathrm{Cov}_p[Y, Z] = \mathrm{Cov}_p[Z, X] = 0.$$

On the other hand there is the functional dependence $Z = X + Y \mod 2$ with respect to p which gives rise to a stochastic dependence between the units. One has $p_1(1)p_2(1)p_3(1) = \frac{1}{8} \neq 0 = p(1,1,1)$. As a preview of the general situation, there is a decompositions of stochastic dependencies into a hierarchy of interactions. This is treated in information geometry, see Amari [Am01].

One way to quantify stochastic dependence between three units, suitable for generalization to an arbitrary number of units, is to use a distance of a probability distribution from the set $\overline{\mathcal{F}}$ of factorizable probability distributions. Let p, q be probability distributions on Ω. The Kullback-Leibler distance [Ku], also called relative entropy, of p from q

$$D(p||q) := \sum_{\omega \in \Omega} p(\omega) \ln(\frac{p(\omega)}{q(\omega)}) \tag{1.2}$$

is a natural measure of distance between probability distributions (see Remark 1.1) being non-negative and with value zero only if $p = q$. The relative entropy distance of p from $\overline{\mathcal{F}}$ has an entropic representation. The Shannon entropy [Sh] of p is[1]

$$H(X) := \sum_{\omega \in \Omega} \eta(p(\omega)) \tag{1.3}$$

with $\eta : [0,1] \to [0, \frac{1}{e}]$, $x \mapsto -x\ln(x)$. One has[2] [Am01]

$$I(X) := \sum_{i=1}^{3} H(X_i) - H(X) = \inf_{q \in \overline{\mathcal{F}}} D(p||q) \tag{1.4}$$

where $H(X_1) = \sum_{\omega_1 \in \Omega_1} \eta(p_1(\omega_1))$ and similarly for unit 2 and 3. This is a (non-negative) measure of stochastic dependence between the three units, vanishing only if p is factorizable. The functional $I(X)$ is called **multi-information**.

Remark 1.1 Shannon entropy is a measure of the uncertainty of a probability distribution. In information theory it is the average description length of repeated independent experiments, see Shannon [Sh]. A generalization, using entropy rates, is possible to ergodic information source, see Cover and Thomas [Co]. Yet, the average coding length of an ergodic information source is treated in quantum theory, see Bjelaković and Szkoła [BS]. The relative entropy, too, has a meaning in information theory. It is the exponent in the probability of error in hypothesis testing between two distributions [Co]. Quantum mechanical generalizations of hypothesis testing are treated by Bjelaković et al. [Bj05] or Petz [Pe08].

1.2 Exponential families in statistics

We introduce exponential families of probability distributions and comment on their geometry—including maximum-likelihood estimation and entropy distance. Geometric specialties of exponential families are the Pythagorean theorem of relative entropy and the mean value parametrization. We recall the situation for a general measurable space

[1]By custom, if only one probability distribution p on Ω is used then a random variable X stands for p while the variables X_i stands for the marginal distributions p_i for $i = 1, 2, 3$.

[2]The article is based on dually flat structures in differential geometry. The representation (1.4) is treated in the article only for probability distributions p that agree with a probability distribution of full support in a certain mean. The representation (1.4) is a special case of our results in Section 8.3.

and in more restrictive settings we also recall that an exponential family is characterized as a parametric statistical model having an efficient estimator, the maximum-likelihood estimator.

A **parametric statistical model** $(X, \mathfrak{A}, \mathfrak{P}, \theta)$ consists of a measurable space $(X, \mathfrak{A})$ with σ-algebra $\mathfrak{A}$ on a set X, called the sample space, and of a real differentiable manifold $\mathfrak{P}$ of probability measures on $\mathfrak{A}$. We assume that $\mathfrak{P}$ has a global chart $\theta : \mathfrak{P} \to \Theta$ to an open subset $\Theta \subset \mathbb{R}^d$. We take the parametrization $\Theta \to \mathfrak{P}$ with respect to a dominating measure μ on $\mathfrak{A}$. Given $\theta \in \Theta$ we denote $p_\theta : X \to \mathbb{R}$ the probability density function for the measure $P_\theta := p_\theta \mu$, [3]

$$\mathfrak{P} = \{p_\theta \mu : \theta \in \Theta\}.$$

Often we denote a parametric statistical model by $\mathfrak{P}$ and specify further components as required. The **expectation** (if it exists) of a measurable function $f : X \to \mathbb{R}$ with respect to a probability measure P on $\mathfrak{A}$ is given by

$$\mathbb{E}_P[f] := \int_X f(x) dP(x). \tag{1.5}$$

If we assume for a parametric statistical model $\mathfrak{P}$ smoothness of the density function $\theta \mapsto p_\theta(x)$ for each $x \in X$ and provided the following expectations exist then the **Fisher information matrix** is defined by

$$g_{ij}(\theta) := \mathbb{E}_{P_\theta}[\frac{\partial}{\partial \theta^i} \ln(p_\theta) \frac{\partial}{\partial \theta^j} \ln(p_\theta)]. \tag{1.6}$$

In many applications, the Fisher information matrix g_{ij} induces a Riemannian metric on $\mathfrak{P}$. Then the inverse is denoted g^{ij}. Fisher information is the starting point for information geometry, where more general affine geometries on $\mathfrak{P}$ are studied in relation to statistical properties, see Amari and Nagaoka [AN] or Murray and Rice [MR].

Remark 1.2 Čencov [Ce] characterized Fisher information as the unique Riemannian metric that is monotone under certain natural transformations. Petz [Pe08] has proved that a whole family of monotone metrics exist for the quantum case. A popular one of these is the BKM-metric, see Petz [Pe94], where the acronym BKM stands for Bogoliubov, Kubo and Mori. We are using this metric in the context of a quantum exponential family. Grasselli and Streater [Gra] have shown that the BKM-metric is the unique monotone metric that satisfies an important duality in information geometry, which holds for the Fisher metric.

[3] It is custom in statistics that θ denotes both a chart and a point in the image of the chart.

Remark 1.3 (Cramér-Rao inequality) One of the most basic links between Fisher information and a statistical quantity is the Cramér-Rao inequality about efficiency of an unbiased estimator. We summarize the proof from [MR]. An **estimator** for the model $(X, \mathfrak{A}, \mathfrak{P}, \theta)$ is a measurable function $\hat{\theta} : X \to \Theta$. An estimator $\widehat{\theta}$ is **unbiased** if for $\theta \in \Theta$ $\mathbb{E}_{P_\theta}(\hat{\theta}) = \theta$ holds. The **covariance** (if it exists) of measurable functions $f, g : X \to \mathbb{R}$ with respect to a probability measure P on $\mathfrak{A}$ is given by the number

$$\mathrm{Cov}_P[f, g] := \mathbb{E}_P\left[(f - \mathbb{E}_P[f])(g - \mathbb{E}_P[g])\right]. \tag{1.7}$$

The **variance** (if it exists) of a measurable function $f : X \to \mathbb{R}$ with respect to a probability measure P on $\mathfrak{A}$ is given by the number

$$\mathrm{Var}_P[f] := \mathrm{Cov}_P[f, f]. \tag{1.8}$$

Under the assumptions of (1.6) and that Fisher information is invertible let us differentiate the equation $\mathbb{E}_{P_\theta}(\hat{\theta}) = \theta$. If partial derivatives commute with the integral then one obtains for $i, j = 1, \ldots, d$

$$\mathbb{E}_{P_\theta}\left[\frac{\partial \ln(p_\theta)}{\partial \theta^j} \hat{\theta}^{\,i}\right] = \delta^i_j := \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{1.9}$$

Using Einstein summation, the positive semi-definite matrix

$$\mathbb{E}_{P_\theta}\left[\left(\hat{\theta}^{\,i} - \mathbb{E}_{P_\theta}[\hat{\theta}^{\,i}] - \frac{\partial \ln(p_\theta)}{\partial \theta^k} g^{ik}\right)\left(\hat{\theta}^{\,j} - \mathbb{E}_{P_\theta}[\hat{\theta}^{\,j}] - \frac{\partial \ln(p_\theta)}{\partial \theta^l} g^{jl}\right)\right]$$

simplifies (with a possibly infinite covariance) to

$$\mathrm{Cov}_{P_\theta}\left[\hat{\theta}^{\,i}, \hat{\theta}^{\,j}\right] - 2\,\mathbb{E}_{P_\theta}\left[(\hat{\theta}^{\,i} - \mathbb{E}[\hat{\theta}^{\,i}])\frac{\partial \ln(p_\theta)}{\partial \theta^l} g^{jl}\right] + g^{ij}.$$

Using (1.9) one obtains $\mathbb{E}_{P_\theta}\left[(\hat{\theta}^{\,i} - \mathbb{E}[\hat{\theta}^{\,i}])\frac{\partial \ln(p_\theta)}{\theta^l} g^{il}\right] = \delta^i_l\, g^{jl} = g^{ij}$. This implies the famous **Cramér-Rao inequality**, that is, the matrix

$$\mathrm{Cov}_{P_\theta}\left[\hat{\theta}^{\,i}, \hat{\theta}^{\,j}\right] - g^{ij}$$

is positive semi-definite. If the equality $\mathrm{Cov}_{P_\theta}\left[\hat{\theta}^{\,i}, \hat{\theta}^{\,j}\right] = g^{ij}$ occurs then the unbiased estimator $\hat{\theta}$ is called **efficient**. If $d = 1$ then the Cramér-Rao inequality is $\mathrm{Var}_{P_\theta}[\hat{\theta}] \geq g^{11}$.

One of the simplest parametric models are exponential families.

Definition 1.4 (Exponential families) An **exponential family** is a parametric statistical model on a measurable space $(X, \mathfrak{A})$ with probability density functions $p_\theta : X \to \mathbb{R}$

$$p_\theta(x) := \exp(C(x) + \sum_{i=1}^{d} \theta^i F_i(x) - \Lambda(\theta))$$

with respect to a dominating measure μ. The measurable functions $C, F_1, \ldots, F_d : X \to \mathbb{R}$ include the **statistic** $\{F_1, \ldots, F_d\}$ and the parameters $\theta^1, \ldots, \theta^d$ are the **canonical parameters**. For normalization the log-Laplace transform for $\theta \in \Theta$

$$\Lambda(\theta) := \ln \int_X e^{C(x)+\sum_{i=1}^{d} \theta^i F_i(x)} \, \mathrm{d}\, \mu(x)$$

is used and it is assumed that $\Lambda(\theta) < \infty$ for $\theta \in \Theta$. An exponential family is **full** if $\Theta = \{\theta \in \mathbb{R}^d : \Lambda(\theta) < \infty\}$.

Example 1.5 (Gauss distributions) For $d = 2$, the normal distribution with mean $\mu \in \mathbb{R}$ and standard deviation $0 < \sigma < \infty$ is defined with respect to Lebesgue measure $\mathrm{d}\, x$ on $X := \mathbb{R}$ by the probability density

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

The set of normal distributions is a full exponential family with $C(x) := 0$, statistic $F_1(x) := x$, $F_2(x) := x^2$ for $x \in \mathbb{R}$ and with canonical parameters $\theta^1 := \frac{\mu}{\sigma^2}$ and $\theta^2 := -\frac{1}{2\sigma^2}$.

More generally, for $k \geq 1$ and $d = k + \frac{k(k+1)}{2}$ the multivariate normal distribution with mean $\mu \in \mathbb{R}^k$ and for a positive definite matrix $\Sigma \in \mathbb{R}^{k \times k}$ is defined with respect to Lebesgue measure $\mathrm{d}\, x$ on $X := \mathbb{R}^k$ by the probability density

$$p(x) := (2\pi)^{-\frac{k}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\langle (x-\mu), \Sigma^{-1}(x-\mu)\rangle\right)$$

using the standard scalar product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^k$. The set of multivariate normal distributions is a full exponential family with $C(x) := 0$ and with statistic $F_i(x) := x_i$ and $F_{i,j}(x) = x_i x_j$ for $1 \leq i \leq j \leq k$. The canonical parameters are $\theta^i := \sum_{j=1}^{k} (\Sigma^{-1})_{i,j} \mu_j$ and $\theta^{i,i} := -\frac{1}{2}(\Sigma^{-1})_{i,i}$ for $1 \leq i \leq k$ as well as $\theta^{i,j} := -(\Sigma^{-1})_{i,j}$ for $1 \leq i < j \leq k$.

Example 1.6 (Poisson distributions) For $d = 1$ the Poisson distribution with parameter $\xi > 0$ is defined with respect to counting measure μ on $X := \{0, 1, 2, \ldots\}$ by the probability density

$$p(x) := e^{-\xi} \frac{\xi^x}{x!}.$$

The set of Poisson distributions is a full exponential family with $C(x) := -\ln(x!)$, identity statistic $F(x) = x$ and canonical parameter $\theta := \ln(\xi)$.

Relative entropy was defined for probability measures P, Q on an arbitrary measurable space $(X, \mathfrak{A})$ by Kullback and Leibler [Ku]. It is non-negative and is zero if and only if $P = Q$. References indicating that relative entropy is a natural distance measure in information theory, are given in Remark 1.1.

As opposed to relative entropy, we can not generalize the Shannon entropy (1.3) in the same way, e.g. see the discussion in Reed and Simon [Ree4] (2.2;6). Also in the quantum case, relative entropy has a much simpler structure compared to the von Neumann entropy. This is demonstrated under the aspect of general relations for composite systems by Ibinson, Linden and Winter [Ib].

Definition 1.7 (Relative entropy) For two probability measures P, Q on a measurable space $(X, \mathfrak{A})$ the **relative entropy** of P from Q is defined by

$$D(P||Q) := \begin{cases} \int_X \ln\left(\frac{\mathrm{d}\,P}{\mathrm{d}\,Q}\right) \mathrm{d}\,P, & \text{if } P \ll Q, \\ +\infty, & \text{otherwise,} \end{cases}$$

with $\frac{\mathrm{d}\,P}{\mathrm{d}\,Q}$ the Radon-Nikodym derivative. The **entropy distance** of P from a statistical model $\mathfrak{P}$ on $(X, \mathfrak{A})$ is defined by

$$D_{\mathfrak{P}}(P) := \inf_{Q \in \mathfrak{P}} D(P||Q).$$

Given statistics (measurable functions) $F_1, \ldots, F_d : X \to \mathbb{R}$ and a vector $a \in \mathbb{R}^d$, a **linear family** is defined by

$$\mathcal{L}_{a,F} := \{P \text{ a probability measure on } \mathfrak{A} \text{ with } \mathbb{E}_P[F] = a\}.$$

The following Pythagorean theorem is the deeper reason why relative entropy as a distance measure is very well compatible with an exponential family.

Remark 1.8 (Pythagorean theorem of relative entropy) On a sample space $(X, \mathfrak{A})$ we consider an exponential family $\mathcal{E}$ with statistics $F_1, \ldots, F_d$. We assume $C = 0$, absorbing the function C into the dominating measure μ if necessary.

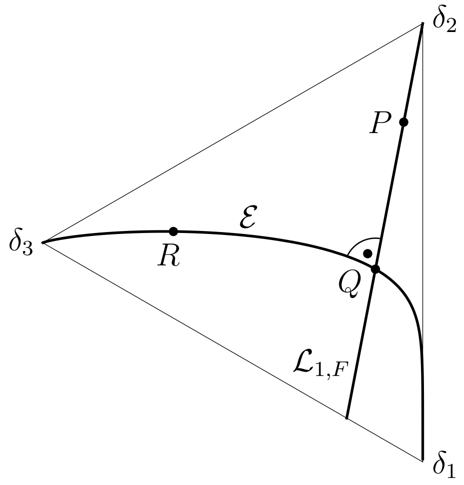


Figure 1.2: For three Dirac measures $\delta_1, \delta_2, \delta_3$ on the sample space $X = \{1, 2, 3\}$ we consider the exponential family $\mathcal{E}$ having density functions $e^{\theta F(x) - \Lambda(\theta)}$ for $\theta \in \mathbb{R}$ with respect to the dominating counting measure $\mu = \delta_1 + \delta_2 + \delta_3$. The statistic is $F(1) = -3, F(2) = 1, F(3) = 2$, the linear family $\mathcal{L}_{1,F}$ is depicted together with $\mathcal{E}$ inside the probability simplex on X. The two families meet orthogonally with respect to Fisher information, and the Pythagorean relation $D(P||Q) + D(Q||R) = D(P||R)$ holds.

(a) Let us consider a linear family $\mathcal{L}_{a,F}$ for some $a \in \mathbb{R}^d$ and choose $P \in \mathcal{L}_{a,F}$ and $R \in \mathcal{E}$ arbitrarily. If $Q \in \mathcal{L}_{a,F} \cap \mathcal{E}$, then the **Pythagorean theorem of relative entropy**

$$D(P||Q) + D(Q||R) = D(P||R) \tag{1.10}$$

holds [Cs03]. Below we comment on consequences and extensions in the literature.

(b) First, let us recall the proof of (1.10). For the probability measure Q resp. R we denote by q resp. r the density function with respect to μ and we denote by θ_q resp. θ_r the canonical parameter. Since $P, Q \in \mathcal{L}_{a,F}$ is a probability measure, the expectations $\mathbb{E}_P[\ln(q)]$, $\mathbb{E}_P[\ln(r)]$, $\mathbb{E}_Q[\ln(q)]$ and $\mathbb{E}_Q[\ln(r)]$ are finite. If $P \not\ll \mu$ then both sides of (1.10) are $+\infty$. Otherwise let $p := \frac{\mathrm{d}\,P}{\mathrm{d}\,\mu}$ denote the probability density function of P. Since $D(P||Q) \geq 0$, see Kullback and Leibler [Ku], the expectation $\mathbb{E}_P[\ln(p)]$ is finite or $-\infty$. The latter case implies that both sides of (1.10) are $+\infty$. Otherwise we obtain

$$\begin{aligned} &D(P||Q) + D(Q||R) - D(P||R) = \mathbb{E}_P[\ln(\tfrac{p}{q})] + \mathbb{E}_Q[\ln(\tfrac{q}{r})] - \mathbb{E}_P[\ln(\tfrac{p}{r})] \\ = \;& \textstyle\int_X (\ln(r) - \ln(q))\,(p - q)\,\mathrm{d}\,\mu = \langle \theta_r - \theta_q, \mathbb{E}_P[F] - \mathbb{E}_Q[F] \rangle = 0 \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard scalar product on $\mathbb{R}^k$.

(c) The geometry of entropy distance from $\mathcal{E}$ is described by linear families. If $a \in \mathbb{R}^d$ then the linear family $\mathcal{L}_{a,F}$ intersects $\mathcal{E}$ in at most one probability measure. Otherwise, if $P, Q \in \mathcal{L}_{a,F} \cap \mathcal{E}$ and $P \neq Q$ then by the Pythagorean theorem of relative entropy follows the contradiction

$$D(P||P) = D(P||Q) + D(Q||P) > 0.$$

In the case that every probability measure in $P \in \mathcal{E}$ has a mean $\mathbb{E}_{P_\theta}[F] \in \mathbb{R}^d$, this implies the **mean value parametrization** of $\mathcal{E}$

$$\{\mathbb{E}_P[F] : P \in \mathcal{E}\} \longrightarrow \mathcal{E}, \qquad a \longmapsto \mathcal{L}_{a,F} \cap \mathcal{E} \tag{1.11}$$

identifying a one-element set with the element. By the Pythagorean theorem, every probability measure P in the **cylinder**

$$\{\widetilde{P} \text{ a probability measure on } \mathfrak{A} : \mathbb{E}_{\widetilde{P}}[F] = \mathbb{E}_{\widetilde{Q}}[F] \text{ for some } \widetilde{Q} \in \mathcal{E}\}$$

there is a unique probability measure $\Pi_{P\to\mathcal{E}} \in \mathcal{E}$, called the ***rI*-projection** of P to $\mathcal{E}$, such that

$$D_{\mathcal{E}}(P) = D(P||\Pi_{P\to\mathcal{E}}). \tag{1.12}$$

The rI-projection is generalized to the case of a missing intersection $\mathcal{L}_{a,F} \cap \mathcal{E} = \emptyset$ by Csiszár and Matúš [Cs03] using a theory of extensions of an exponential family. We will present details to the construction of an extension only for the case of a finite sample space X in the following section.

(d) The Pythagorean theorem of relative entropy owes its name to the fact that a linear family $\mathcal{L}_{a,F}$ and the exponential family $\mathcal{E}$ intersect orthogonally with respect to Fisher information. We assume $P \in \mathcal{L}_{a,F}$ such that $P \ll \mu$ and $Q \in \mathcal{L}_{a,F} \cap \mathcal{E}$ and further $\frac{\partial}{\partial\theta^i}|_{\theta=\theta_q}\Lambda(\theta) < \infty$. Let p, q denote the probability density functions of P, Q respectively and θ_q denotes the canonical parameter of Q. The tangent vector of the linear curve from Q to P and the tangent vector of a canonical coordinate curve $i = 1, \ldots, d$ of $\mathcal{E}$ emanating from Q are orthogonal for Fisher information:

$$\begin{aligned} &\int_X \tfrac{\partial}{\partial t}|_{t=0} \ln(q + t(p-q)) \tfrac{\partial}{\partial\theta^i}|_{\theta=\theta_q} \ln(p_\theta) \,\mathrm{d}\, Q = \int_X \tfrac{1}{q}(p-q)\left(F_i - \tfrac{\partial}{\partial\theta^i}|_{\theta=\theta_q}\Lambda(\theta)\right) q \,\mathrm{d}\,\mu \\ = \;\; &\int_X (p-q) F_i \,\mathrm{d}\,\mu = \mathbb{E}_P[F_i] - \mathbb{E}_Q[F_i] = 0. \end{aligned}$$

An example for the Pythagorean theorem of relative entropy is depicted in Figure 1.2. There exist Pythagorean theorems in information geometry for divergences generalizing the relative entropy, see Amari and Nagaoka [AN].

Still, a theoretical foundation of maximum-likelihood estimation is discussed controversially, see Efron [Ef]. The quantum mechanical generalization is rather involved even in the simplest cases, if the statistics do not commute, see Petz [Pe08]. We mention maximum-likelihood estimation because its calculation for an exponential family is equivalent to minimization of relative entropy from the exponential family. In addition, it makes the mean value parametrization more natural from the statistical point of view.

Definition 1.9 (Maximum-likelihood estimation for exponential families) We consider the full exponential family $\mathcal{E}$ on a sample space $(X, \mathfrak{A})$ with statistic $F : X \to \mathbb{R}^d$. Given the sample mean

$$a := \frac{1}{N}\sum_{i=1}^{N} F(x_i) \in \mathbb{R}^d$$

of an i.i.d. sample $x_1, \ldots, x_N$ from a probability measure $P_\theta \in \mathcal{E}$ with $\theta \in \Theta$ unknown, a maximizer $\theta^* \in \Theta$ of the log-likelihood function

$$\theta \mapsto \langle \theta, a \rangle - \Lambda(\theta)$$

is called a **maximum-likelihood estimate** of the unknown parameter.

Remark 1.10 (Maximum-likelihood for exponential families) As is well-known, see e.g. Csiszár and Matúš [Cs08], any $\theta^* \in \Theta$ with mean $a = \mathbb{E}_{P_{\theta^*}}[F] \in \mathbb{R}^d$ is a maximum-likelihood estimate in the above sense. Indeed, if $C = 0$ (absorbed into the dominating measure μ) then a direct calculation gives for arbitrary $\theta \in \Theta$

$$[\langle \theta^*, a \rangle - \Lambda(\theta^*)] - [\langle \theta, a \rangle - \Lambda(\theta)] = D(P_{\theta^*}||P_\theta) \geq 0.$$

As pointed out in Remark 1.8 (c), the linear family $\mathcal{L}_{a,F}$ intersects $\mathcal{E}$ only in P_{θ^*}, so the estimate is unique. We can see here a connection between maximum likelihood estimation, mean value parametrization (1.11) and minimization of entropy distance: if $P \in \mathcal{L}_{a,F}$ then the rI-projection of P to $\mathcal{E}$ is $\Pi_{P \to \mathcal{E}} = P_{\theta^*}$, so the entropy distance of P from $\mathcal{E}$ is (1.12)

$$\inf_{Q \in \mathcal{E}} D(P||Q) = D(P||P_{\theta_*}).$$

Remark 1.11 (Efficient estimators for exponential families) Under suitable assumption one can prove that maximum-likelihood estimation provides an efficient unbiased estimator for an exponential family, see Section 7.4 in [MR]. Conversely, the existence of an efficient unbiased estimator for a parametric statistical model implies that the model is an exponential family parametrized by mean values, see e.g. Sections 2.5 and 3.5 in [AN].

1.3 Example: mean value chart and closure

With little effort we have found a maximum-likelihood estimate and we have controlled the entropy distance from an exponential family in Remark 1.10 in case of presence of a suitable mean from the exponential family. Now we give an example of the rI-closure of an exponential family and we show how the mean value parametrization (1.11) of the family is extended to define maximum-likelihood estimates for arbitrary sample means and to describe entropy distance for arbitrary probability distributions. The theoretical prerequisites are found in Barndorff-Nielsen [Bar] or in Csiszár and Matúš [Cs03].

The understanding of a finite sample space is sufficient for generalization to finite-level quantum systems. We study the finite sample space $X := \{1, 2, 3, 4\}$, and with respect to a non-zero dominating measure μ on X the full exponential family $\mathcal{E} := \{P_{\mu,\theta}(x) : \theta \in \Theta\}$ with $d = 2$ statistics $F_1, F_2 : X \to \mathbb{R}$ and $\Theta := \mathbb{R}^2$. We denote a density function by $p_{\mu,\theta}(x) := \frac{\mathrm{d}\, P_{\mu,\theta}(x)}{\mathrm{d}\,\mu}$ having $\Lambda_\mu(\theta) := \ln \int_X \exp(\theta^1 F_1(x) + \theta^2 F_2(x))\, \mathrm{d}\,\mu$ and

$$p_{\mu,\theta}(x) = \exp(\theta^1 F_1(x) + \theta^2 F_2(x) - \Lambda_\mu(\theta)), \qquad x \in X, \tag{1.13}$$

see also Definition 1.4. Let us choose the counting measure μ on X. Then the probability density functions $p_{\mu,\theta}$ belong to the probability simplex $\mathbb{P}^4 := \{p \in \mathbb{R}^4 : p_1, p_2, p_3, p_4 \geq 0 \text{ and } p_1 + p_2 + p_3 + p_4 = 1\}$.

Some arguments can be understood without differential calculus. Every probability measure $P \in \mathcal{E}$ has a mean with respect to F, so the mean value parameters (1.11) for $p := \frac{\mathrm{d}\,P}{\mathrm{d}\,\mu}$

$$\mathbb{E}_P[F] = p(1)F(1) + p(2)F(2) + p(3)F(3) + p(4)F(4) \in \mathbb{R}^2$$

parametrize $\mathcal{E}$. More generally, for a probability measure P on X, the mean $\mathbb{E}_P[F]$ is a convex combination of the values of the statistic $F(i) \in \mathbb{R}$ and the $F(i)$'s are the support points of the push-forward measure F_*P on $\mathbb{R}^2$

$$F_*P(A) := P(F^{-1}(A)), \qquad A \subset \mathbb{R}^2.$$

Hence every mean $\mathbb{E}_P[F]$ of a probability measure P on X with respect to the statistic F belongs to the **convex support**

$$\mathrm{cs}(F_*\mu) := \mathrm{conv}(\mathrm{supp}(F_*\mu)), \tag{1.14}$$

which is convex hull of the support $\mathrm{supp}(F_*\mu) = \{F(1), F(2), F(3), F(4)\}$ of $F_*\mu$. The convex support as the convex hull of finitely many points is polytope[4].

For a quantum generalization, we assume the statistics F_1 and F_2 are orthonormal in $\mathbb{R}^X \cong \mathbb{R}^4$. This condition does not restrict the choice of families: a linear change of the statistic neither affects sample means nor means from a measure in $\mathcal{E}$. Using $\mathbb{R}^2$ as a coordinate vector space for $U := \mathrm{Lin}\{F_1, F_2\}$ with respect to the basis F_1, F_2 we have

$$\mathbb{E}_P[F] = \langle F_1, p\rangle F_1 + \langle F_2, p\rangle F_2 = \pi_U(p) \in U$$

where π_U is the orthogonal projection from $\mathbb{R}^X$ to U. In this thesis we prefer the coordinate free description by U. The results are translated back to an actual statistic at the relevant places in Remark 6.32, Remark 7.19 and Section 8.1. We notice

$$\mathrm{cs}(F_*\mu) = \pi_U(\mathbb{P}^4). \tag{1.15}$$

Let us also assume that $(F_1, F_2, \mathrm{Id}\,|_X)$ are linearly independent. Then the canonical parametrization $\Theta \to \mathcal{E}$, $\theta \mapsto P_{\mu,\theta}$ is a diffeomorphism and the inverse of the mean value parametrization (1.11) is the **mean value chart**

$$\pi_{\mathcal{E}} : \qquad \mathcal{E} \longrightarrow \mathrm{int}(\mathrm{cs}(F_*\mu)), \qquad P \longmapsto \mathbb{E}_P[F],$$

[4]A polytope is defined as the convex hull of finitely many points. The convex support is not necessarily a polytope for infinite sample space X. For example, if we consider the normal distributions in Example 1.5, then the convex support is the convex hull of a parabola.

see [Bar]. Here, int denotes the interior of a subset of $\mathbb{R}^2$. We give now some of the arguments to establish the mean value chart since they are used in Section 6.3 for quantum systems. For $i, j = 1, 2$

$$\frac{\partial}{\partial \theta^i} \Lambda_\mu(\theta) = \mathbb{E}_{P_{\mu,\theta}}[F_i]$$

holds and

$$\frac{\partial}{\partial \theta^j} \mathbb{E}_{P_{\mu,\theta}}(F_i) = \frac{\partial^2}{\partial \theta^i \partial \theta^j} \Lambda_\mu(\theta) = \mathbb{E}_{P_{\mu,\theta}}\left[(F_i - \mathbb{E}_{P_{\mu,\theta}}[F_i])(F_j - \mathbb{E}_{P_{\mu,\theta}}[F_j])\right]. \tag{1.16}$$

The Jacobian of $\theta \mapsto \mathbb{E}_{P_{\mu,\theta}}[F]$ is the positive definite Fisher information matrix (1.6) because $\frac{\partial}{\partial \theta^i} \ln(p_{\mu,\theta}) = F_i - \mathbb{E}_{P_{\mu,\theta}}[F_i]$. Here the Fisher information matrix is positive definite, since it is the Gram matrix of the linearly independent vectors $F_1 - \mathbb{E}_{P_{\mu,\theta}}[F_1]$ and $F_2 - \mathbb{E}_{P_{\mu,\theta}}[F_2]$. In the quantum case, the BKM-metric will replace Fisher information. Since the Jacobian is invertible, the mean value chart is a local diffeomorphism. The proof that the image of the mean value chart is $\mathrm{int}(\mathrm{cs}(F_*\mu)))$, but not smaller, requires more effort and is only written in Section 6.3.

The reader is invited to identify constructions around the mean value chart in Figure 1.3 on page 27. There are optimization principles implicit in the mean value chart described in Remark 1.8 and Remark 1.10: if a probability measure P on X has mean $a := \mathbb{E}_P[F]$ equal to the mean $a = \mathbb{E}_Q[F]$ of a member $Q \in \mathcal{E}$, then the *rI*-projection $\Pi_{P \to \mathcal{E}}$ of P to $\mathcal{E}$ is the intersection of the linear family $\mathcal{L}_{a,F}$ with $\mathcal{E}$. The mean value parametrization $\pi_{\mathcal{E}}^{-1}$ is the maximum-likelihood estimator for $\mathcal{E}$. The *rI*-projection may be written

$$\Pi_{P \to \mathcal{E}} = \pi_{\mathcal{E}}^{-1} \circ \pi_U(p)$$

and uniquely minimizes relative entropy $\inf_{Q \in \mathcal{E}} D(P||Q) = D(P||\Pi_{P \to \mathcal{E}})$.

The need for an extension of $\mathcal{E}$ arises for two reasons. Firstly, concerning the maximum-likelihood estimation, if a finite i.i.d. sample $x_1, \ldots, x_N$ is drawn from a probability measure $Q \in \mathcal{E}$ then there is a non-zero probability that the sample mean

$$a := \frac{1}{N} \sum_{i=1}^{N} F(x_i) \in \mathbb{R}^2$$

belongs to the boundary of the convex support. But there the mean value parametrization of $\mathcal{E}$ is not defined; indeed the log-likelihood function $L(\theta) := \langle \theta, a \rangle - \Lambda_\mu(\theta)$ has no maximum for $\theta \in \mathbb{R}^2$ because the function L is strictly convex (1.16) while the gradient

$$\nabla L(\theta) = F_1 \frac{\partial}{\partial \theta^1} L(\theta) + F_2 \frac{\partial}{\partial \theta^2} L(\theta) = a - \mathbb{E}_{P_{\mu,\theta}}[F] \neq 0.$$

In Figure 1.3 this corresponds to a sample mean a on a side of the triangular convex support. Secondly, we want to control the entropy distance $D_{\mathcal{E}}(P)$ from $\mathcal{E}$ for a probability measure P on X, not necessarily coinciding with the probability measure of $\mathcal{E}$. We notice for arbitrary $\theta \in \Theta$ that

$$D(P||P_\theta) = -H(P) - L(\theta)$$

with Shannon entropy H and the log-likelihood function L from above. The missing maximum-likelihood estimate manifests itself in a missing minimum of $\mathcal{E} \mapsto \mathbb{R} \cup \{\infty\}$, $Q \mapsto D(P||Q)$.

A solution to these problems is provided by the ***rI*-closure** of $\mathcal{E}$ that is defined by

$$\mathrm{cl}_{rI}(\mathcal{E}) := \{Q \text{ a probability measure on } X : D_{\mathcal{E}}(Q) = 0\}.$$

For a finite sample space X this closure is known to be equal to the topological closure, $\overline{\mathcal{E}} = \mathrm{cl}_{rI}(\mathcal{E})$, see Barndorff-Nielsen [Bar] and Csiszár and Matúš [Cs03]. In addition, the projection $\pi_{\overline{\mathcal{E}}} := \mathbb{E}_{(\cdot)}[F]|_{\overline{\mathcal{E}}}$

$$\pi_{\overline{\mathcal{E}}} : \mathrm{cl}_{rI}(\mathcal{E}) \to \mathrm{cs}(F_*\mu) \tag{1.17}$$

is a homeomorphism. Using the extended mean value chart (1.17) one can define an rI-projection for an arbitrary probability measure P on X by

$$\Pi_{P\to\mathcal{E}} := \pi_{\overline{\mathcal{E}}}^{-1} \circ \pi_U(p) \tag{1.18}$$

such that the entropy distance is

$$D_{\mathcal{E}}(P) = D(P||\Pi_{P\to\mathcal{E}}).$$

It turns out for a finite sample space X that the closure $\overline{\mathcal{E}} = \mathrm{cl}_{rI}(\mathcal{E})$ is an extension of $\mathcal{E}$ appropriate also for maximum-likelihood estimation, see [Bar] and see Csiszár and Matúš [Cs08] for generalizations and recent development. The case of an arbitrary Borel measure on $\mathbb{R}^k$ is much more complex compared to a finite sample space [Cs03]. For instance one can have $\overline{\mathcal{E}} \supsetneq \mathrm{cl}_{rI}(\mathcal{E})$.

Since the quantum case in Section 7.1 will be analogous, we explain in detail how the extension $\overline{\mathcal{E}}$ is constructed. We will realize in Section 7.2 that results for a finite-level quantum system are much richer in structure than the probabilistic case of finite support. For example, $\overline{\mathcal{E}} \supsetneq \mathrm{cl}_{rI}(\mathcal{E})$ can occur.

The convex support $\mathrm{cs}(F_*\mu)$ is a polytope, see (1.14). The concept of a **face** in convex geometry is defined in another chapter. The faces of the triangular convex support in Figure 1.3 are the empty set $\emptyset$, the three vertices $F(1), F(2), F(4)$, the three sides $[F(1), F(2)]$, $[F(2), F(4)]$, $[F(4), F(1)]$ and the triangular convex support itself. Given a non-empty face G of the convex support, we define μ^G as the counting measure on $F^{-1}(G) \subset X$ and exponential family

$$\mathcal{E}_{\mu^G} := \{P_{\mu^G,\theta} : \theta \in \mathbb{R}^2\}$$

with the notation from (1.13). Then $P_{\mu^G,\theta}$ is the conditional probability measure $P_{\mu^G,\theta}(A) = P_{\mu,\theta}(A|F^{-1}(G))$ and in particular, $\mathcal{E} = \mathcal{E}_\mu = \mathcal{E}_{\mu^G}$ if G is the convex support. If G is a smaller face then the canonical parametrization of the exponential family $\mathcal{E}_{\mu^G}$ is not diffeomorphic, but the expectation with respect to F still is. One has

$$\overline{\mathcal{E}} = \bigcup_G \mathcal{E}_{\mu^G}$$

with the union running over the non-empty faces G of the convex support $\mathrm{cs}(F_*\mu)$. An example is depicted in Figure 1.4 on page 28.

1.4 Quantum systems and measurement

We define a finite-level quantum systems and explain how information about its state is obtained from a measurement. A personal recommendation of research monographs follows, the scope of each one goes far beyond our applications. Minimal sufficient definitions for our treatment are found in Petz [Pe08] or Amari and Nagaoka [AN]. The introduction in Holevo [Ho] develops quantum theory from probability theory in brief, with analogies and differences. Beltrametti and Cassinelli [Bel] convince with physical counterparts motivating many mathematical concepts.

Definition 1.12 For a natural number $n \in \mathbb{N}$ an **n-level quantum system** is modeled on the n-dimensional Hilbert space $H := \mathbb{C}^n$. We denote by $\mathcal{B}(H)$ the algebra of $n \times n$-matrices acting as linear operators on H. A **state** of the quantum system is described by a **density matrix** $\rho \in \mathcal{B}(H)$. By definition, a matrix $\rho \in \mathcal{B}(H)$ is a density matrix, if ρ is positive semi-definite (ρ is self-adjoint and has no negative eigenvalues) and has unit trace $\mathrm{tr}(\rho) = 1$, see e.g. [Ho].

Definition 1.13 [AN] An **operator valued measure** for a non-empty finite set X is a mapping $E : 2^X \to \mathcal{B}(H)$ with 2^X the power set of X such that

(a) $E(\bigcup_{i=1}^k B_i) = \sum_{i=1}^k E(B_i)$ if $\{B_i\}_{i=1}^k$ are mutually disjoint subsets of X and

(b) $\sum_{x\in X} E(x)^* E(x) = \mathbb{1}$, with x substituting the one-element set $\{x\}$,

where the asterisk * denotes the adjoint of a matrix[5] and $\mathbb{1}$ is the identity matrix in $\mathcal{B}(H)$. We call an operator valued measure E a **measurement** for simplicity. If the state of a quantum system is represented by the density matrix $\rho \in \mathcal{B}(H)$ then in the measurement E an event $Y \subset X$ is observed with probability

$$P_\rho(Y) := \sum_{y\in Y} \mathrm{tr}(E(y)\rho E(y)^*). \tag{1.19}$$

If $P_\rho(Y) > 0$ then the state of the system after observation of Y is

$$\rho^Y := \frac{\sum_{y\in Y} E(y)\rho E(y)^*}{\sum_{y\in Y} \mathrm{tr}(E(y)\rho E(y)^*)}. \tag{1.20}$$

A particular case is $P_\rho(X) = 1$ and $\rho^X = \sum_{x\in X} E(x)\rho E(x)^*$. The density matrix ρ^Y is called a **state reduction** of ρ. An important special measurement is a **simple** or **von Neumann measurement**, where the measurement operators are orthogonal projectors $E(x)^2 = E(x) = E(x)^*$ for $x \in X$.

Some special features of a von Neumann measurement $E : 2^X \to \mathcal{B}(H)$ follow. Given an arbitrary event $Z \subset X$ and a density matrix $\rho \in \mathcal{B}(H)$, the measurement probability is

$$P_\rho(Z) = \mathrm{tr}(E(Z)\rho)$$

because $E(x)E(y) = 0$ if $x, y \in X$ and $x \neq y$ (Remark 2.16 (b)). The state reduction of ρ is

$$\rho^X = \sum_{x\in X} E(x)\rho E(x)^* = \sum_{x\in X} P_\rho(x)\rho^x \tag{1.21}$$

where the right-hand side is only defined if $P_\rho(x) > 0$ and x substitutes $\{x\}$ for $x \in X$. If the event $Z \subset X$ has positive probability $P_\rho(Z) > 0$ then

$$P_\rho(\,\cdot\,|Z) = P_{\rho^Z}, \tag{1.22}$$

the conditional measurement probabilities of ρ given Z are the measurement probabilities of the state reduction ρ^Z (the calculation uses $E(Y \cap Z) = E(Y)E(Z)$). The formula (1.21) expresses the notion that a state reduction is a "measurement without selection" [Bel], that is, ρ^X is a post-measurement and pre-observation state, likewise, ρ^X is a density matrix-valued random variable distributed by the measurement probability.

[5] In $\mathbb{C}^n$ the standard scalar product is $\langle x, y\rangle := \overline{x_1}y_1 + \ldots + \overline{x_n}y_n$ for $x, y \in \mathbb{C}^n$.

Definition 1.14 When E is a simple measurement and the possible outcomes of E are labeled by mutually distinct real numbers $\{\alpha_i\}_{i\in X}$, we call the pair $(\{E(i)\}, \{\alpha_i\})$ an **observable**. We may represent an observable $(\{p_i\}, \{\alpha_i\})$ by the self-adjoint matrix

$$a = \sum_{i\in X} \alpha_i p_i.$$

Two observables are said to be **compatible** if the associated self-adjoint matrices, say a and b commute, i.e.

$$ab = ba.$$

Otherwise, the observables are **incompatible**. An idempotent observable, that is, an orthogonal projector, is often called an **elementary event** corresponding to a yes-no experiment.

For a simple measurement E the numbers $\{\alpha_i\}_{i\in X}$ assigned in Definition 1.14 are the eigenvalues and $\{E(i)\}_{i\in X}$ are the orthogonal projectors onto eigenspaces of the observable a. There is a one-to-one correspondence between hermitian matrices and observables.

Definition 1.15 If the state of a system is represented by the density matrix ρ then an observable a defines the random variable (ρ, a) of possible measurement results α_i distributed by P_ρ (1.19). The random variable (ρ, a) has **mean value**

$$\mathbb{E}_\rho[a] := \mathbb{E}[(\rho, a)] = \sum_{i\in X} \alpha_i P_\rho(i) = \mathrm{tr}(\rho a). \tag{1.23}$$

The special case that density matrices and observables of a quantum system are confined to the space of diagonal matrices, then we speak of the **classical case**.

Remark 1.16 (A state space projects to a simplex) If we are given a von Neumann measurement $E : 2^X \to \mathcal{B}(H)$ with rank-one orthogonal projectors $\{E(x)\}_{x\in X}$ then $X = \{1, \dots, n\}$ and the linear mapping describing the state reduction

$$\mathcal{B}(H) \longrightarrow \mathrm{Lin}\{E(x)\}_{x\in X}, \qquad a \longmapsto \sum_{x\in X} E(x)aE(x) = \sum_{x\in X} \mathrm{tr}(E(x)a)E(x)$$

is the orthogonal projection from $\mathcal{B}(H)$ to $\mathrm{Lin}\{E(x)\}_{x\in X}$ with respect to the **Hilbert-Schmidt inner product** $(a, b) \mapsto \mathrm{tr}(a^*b)$ for $a, b \in \mathcal{B}(H)$. The restriction to density matrices assigns state reductions (1.21)

$$\left\{\begin{matrix}\text{density}\\ \text{matrices}\end{matrix}\right\} \quad\longrightarrow\quad \mathbb{P}^n, \qquad \rho \quad\longmapsto\quad \rho^X = \sum_{x\in X} P_\rho(x)E(x). \tag{1.24}$$

When the orthogonal projectors $\{E(x)\}_{x\in X}$ are identified with the canonical basis of $\mathbb{R}^n$, this mapping is onto the **probability simplex**

$$\mathbb{P}^n := \{p \in \mathbb{R}^n : \sum_{i=1}^{n} p_i = 1, p_i \geq 0 \text{ for } i = 1, \dots, n\}, \tag{1.25}$$

Surjectivity follows from convex combinations of the fix points $E(x)^X = E(x)$ for $x \in X$.

1.5 Entropy distance for quantum systems

We will now explain the diagrams on the cover page and show which special similarities of quantum theory to probability theory can arise for the entropy distance from an exponential family. The results in this thesis apply to a finite-dimensional C*-algebra by representation on a matrix algebra. The probabilistic case is included with diagonal matrices.

We define an exponential family $\mathcal{E}$ of density matrices by statistics F_1, F_2 (for simplicity now only two) and an affine parameter C, the matrices $C, F_1, F_2 \in \mathcal{B}(H)$ being self-adjoint on $H := \mathbb{C}^n$. The exponential family $\mathcal{E}$ consists of density matrices for $\theta = (\theta^1, \theta^2) \in \mathbb{R}^2$

$$\rho_\theta := \frac{\exp(C + \theta^1 F_1 + \theta^2 F_2)}{\operatorname{tr}\left(\exp(C + \theta^1 F_1 + \theta^2 F_2)\right)}.$$

The left-hand diagram on the cover page describes entropy distance in probability theory as discussed in Section 1.3. The density matrices in this classical case form the probability simplex $\mathbb{P}^n$, see (1.25), which is drawn as the top triangle in the diagram. The arc in the diagram corresponds to the exponential family $\mathcal{E}$ and the quadrilateral to the convex support $\operatorname{cs}(F_*\mu)$. This is the set of mean values of the simplex $\mathbb{P}^n$ with respect to F_1 and F_2 and this is a polytope, see (1.14). The arrows are the rI-projection (1.18), and the extended mean value chart (1.17) with inverse.

The right-hand diagram on the cover page describes entropy distance for finite-level quantum systems as resolved in Theorem 6. Compared to the left-hand diagram, the triangle is replaced by a circle representing a set of quantum states, for example a Bloch ball. The quadrilateral is replaced by a cheese-shaped area—it is demonstrated in Figure 5.4 on page 104 that the mean value set of a quantum system can combine flat and curved components.

The maximum-likelihood aspect of entropy distance explained in Remark 1.10 disappears for quantum systems. Empirically, only the measurement probabilities (1.19) can be analyzed and not the proper state. However, the structure of entropy distance is perfectly analogous to the probabilistic case. This is explained in the preamble to Section 7.1. Differences to the probabilistic case include possible discontinuities of entropy distance, that appear if and only if the rI-closure of $\mathcal{E}$ is not the topological closure, see the Staffelberg family in Example 7.28.

Simplifications can occur if the statistics F_1 and F_2 are compatible, that is $F_1F_2 = F_2F_1$. Then F_1 and F_2 are simultaneously diagonalizable so there exists a sequence of rank-one orthogonal projectors $\{p_i\}_{i=1}^n$ and real coefficients $\{a_i\}_{i=1}^n$ and $\{b_i\}_{i=1}^n$ such that $\sum_{i=1}^n p_i = \mathbb{1}$ and such that $F_1 = \sum_{i=1}^n a_i p_i$ and $F_2 = \sum_{i=1}^n b_i p_i$. We consider the von Neumann measurement E extending the mapping

$$\{1,\ldots,n\} \to \mathcal{B}(H) \qquad \text{where} \qquad i \mapsto p_i.$$

The state reduction (1.24) of an arbitrary density matrix $\rho \in \mathcal{B}(H)$ with respect to the measurement E is

$$\rho^{\{1,\ldots,n\}} = \sum_{i=1}^n p_i \rho p_i.$$

Mean values of the statistics F_1 and F_2 are the same for ρ and for the state reduction $\rho^{\{1,\ldots,n\}}$,

$$\langle F_1, \rho^{\{1,\ldots,n\}}\rangle = \langle F_1, \rho\rangle \qquad \text{and} \qquad \langle F_2, \rho^{\{1,\ldots,n\}}\rangle = \langle F_2, \rho\rangle.$$

Then the following diagram commutes.

$$\begin{tikzcd}
\left\{\begin{matrix}\text{density}\\ \text{matrices}\end{matrix}\right\} \arrow[rr] \arrow[dr] & & \left\{\begin{matrix}\text{state reductions}\\ \text{of density matrices}\end{matrix}\right\} \cong \mathbb{P}^n \arrow[dl] \\
& \left\{\begin{matrix}\text{mean values}\\ \text{of the statistics } F_1, F_2\end{matrix}\right\} &
\end{tikzcd}$$

This additional structure arising for compatible statistics is displayed on the cover page in the third diagram. The set of mean values of density matrices is a polytope because it is the set of mean values of the simplex $\mathbb{P}^n$ of measurement probabilities of E (1.25).

If in addition to compatible statistics F_1 and F_2 the affine offset C commutes with F_1 and with F_2, then the exponential family $\mathcal{E}$ actually belongs the set of measurement probabilities of E. One has

$$\mathcal{E} \subset \left\{\begin{matrix}\text{state reductions}\\ \text{of quantum systems}\end{matrix}\right\} \cong \mathbb{P}^n.$$

This situation is displayed on the cover page in the second diagram. In part it can be dealt with using only probability theory. However, a density matrix need not be a mixture of post-measurement states, so a probabilistic treatment does not cover the problem in full. For example it will miss the improved bound on the maximal possible disorder of a local maximizer of entropy distance from $\mathcal{E}$, see Remark 7.30 on page 162. This justifies the extensive studies for the main result in Theorem 6 for compatible statistics.

1.6 Overview

This section gives a summary on the purpose of each chapter including only the most important references to the literature and new results. More details are found in the introductory parts of some chapters.

Finite-dimensional C*-algebras. Chapter 2 introduces algebraic fundamentals from the literature. The main focus is on finite-dimensional C*-algebras and their projector lattices. The representation of a finite-dimensional C*-algebra as a matrix algebra follows from the Gelfand-Naimark theorem, see e.g. Davidson [Da]. The structure of the projector lattice as a union of Grassmannian manifolds is recalled and customized.

Convex geometry in Euclidean space. Chapter 3 is two-fold. There is an introduction of the necessary concepts from convex geometry in finite-dimensional Euclidean spaces by reference to the literature mainly from Rochafellar [Ro]. Then we prove new results useful for the study of state spaces. We establish face lattice isomorphisms between a convex set and an orthogonal projection of the set, the two being linked by the cylinder formed by the inverse projection. Another subject is the pair (touching cone, normal cone) which is the dual concept to the pair (face, exposed face), see Schneider [Sch]. A state space has the special feature of exposed face and this is preserved by intersection with an affine space. A state space has the special feature of normal cone and this is preserved by orthogonal projection to a linear space. We note that a projection of a convex set can have non-exposed faces although the convex set does not have any by itself. We prove a characterization of exposed faces for a convex set, where every touching cone is a normal cone.

State spaces. Chapter 4 contains the convex geometric fundamentals of the state space of a matrix algebra. Here we follow the example set by Alfsen and Schultz [Al] who compute the face lattice of the state space in a C*-algebra and a von Neumann algebra. The desired statements about normal cones and relative interiors are easily deduced but we provide own proofs. An original result is the homeomorphism between the projector lattice of the matrix algebra and the face lattice of the state space equipped with Hausdorff distance.

State reflections. Chapter 5 is a new contribution to the field of state spaces. State reflections are defined as orthogonal projections of a state space to some vector space. A state reflection is a generalization of the concept of convex support for a probability measure, see Barndorff-Nielsen [Bar] or Csiszár and Matúš [Cs03]. While a state space has only exposed faces, a state reflection can have non-exposed faces. We describe the face lattices of a state reflection using lattice morphisms to the projector lattice of the algebra. We use access sequences for characterization and computation of projector lattices of a state reflection. We start to clear the topology of the projector lattices.

Exponential families. Chapter 6 is a customized collection of known results about exponential families in a matrix algebra. But the examples of the **swallow** and **Staffelberg family** serve subsequently as models to explain new features of quantum exponential families. We explain geometry of relative entropy including the BKM-metric and the Pythagorean theorem of relative entropy, see e.g. Petz [Pe94, Pe08]. The mean value chart for exponential families is explained following Wichmann [Wic].

Entropy distance. Chapter 7 is a new contribution to the field of quantum exponential families. The rI-closure known from exponential families of probability distributions, see e.g. Csiszár and Matúš [Cs03], is an appropriate extension for an exponential family in a matrix algebra to describe entropy distance. A difference with the classical case is that the rI-closure is not covered by the closures of e-geodesics in the exponential family. This is related to non-exposed faces of the state reflection. Another difference is that an exponential family in a matrix algebra can have the rI-closure strictly included in the topological closure. We prove that the rI-closure is topologically closed if and only if entropy distance is continuous. We can prove a connection between the topology of an exponential family and the topology of associated projector lattices. The rank of a local maximizer of entropy distance is bounded. The bound can improve quadratically in the quantum case over the classical case.

Some examples. In Chapter 8 the new results are applied to examples. The Gibbs ensembles with respect to a set of observables are fundamental to Quantum Statistical Mechanics, see e.g. Ingarden et al. [In]. We prove that the rI-closure of the Gibbs ensembles is the set of maximum entropy ensembles (including singular density matrices). The second example covers abelian matrix algebras where our results simplify to the case of probability distributions of finite support. We show continuity of entropy distance. The initial motivation of this thesis was to study multi-information, see e.g. Ay and Knauf [AK], which is the entropy distance for factorizable families. We obtain continuity of multi-information. For a local maximizer of multi-information we prove an upper rank bound increasing as the square root of the number of units. This is a quadratic improvement over the classical case. We include a partial classification of convex exponential families because it is believed that their structure exhibits further information about factorizable families. For the convex subfamilies of factorizable families the classification is total.

Application: Stationary Markov transitions. In Chapter 9 a dynamical situation is discussed with an interaction measure related to multi-information, see Ay [Ay01]. Similarly to the case of multi-information, the measure can have applications in Mathematical Physics, Mathematical Biology and Neuroscience. Using a cyclic decomposition of Markov chains, we prove the emergence of determinism for Markov transitions of a high temporal interaction.

1.7 Questions

The following questions, while limiting this work, propose new viewpoints that seem within reach.

Convex geometry

We have proved Theorem 1 under the condition that every touching cone of a convex set is a normal cone and we could ask for more:

Question 1 What are the special properties of a convex subset of $\mathbb{R}^m$ where every touching cone is a normal cone?

More generally, we like to know other branches of mathematics where the projection of a convex set plays a role. An example is the Fourier-Motzkin Elimination [Zi].

Topology of projector lattices

We consider a matrix algebra A. Let $V \subset A$ be a vector space of self-adjoint and traceless matrices. The idea is to get information from the projector lattices about singularities of an exponential family of the form $\mathcal{E} := \exp_1(\theta + V)$ for $\theta \in A$ self-adjoint. Question 5, Question 6 and Question 7 on page 119 are ordered in decreasing strength, the strongest one being Question 5 about the inclusion

$$\mathcal{P}_V \subset \overline{\mathcal{P}_{V,\perp}}. \tag{1.26}$$

The affirmative to (1.26) is related to a Grassmannian approximation of the rI-closure of $\mathcal{E}$, see Remark 7.25. A special case of (1.26) is the following.

Question 2 Are isolated points in $\mathcal{P}_V$ included in the exposed projector lattice $\mathcal{P}_{V,\perp}$?

The attempts made at the end of Section 5.4 to answer Question 7 could be structured by the following.

Question 3 Is it true that a stable convex set $C \subset \mathbb{R}^m$ has a homeomorphic symmetrization map at a vector space $W \subset \mathbb{R}^m$ if and only if the projection $\pi_W|_C$ is open?

To investigate (1.26) one can also study the projector lattice $\mathcal{P}_V$ geometrically. Is it a union of pieces of a variety? A simpler question would be: does a "proper quantum case" have an infinite projector lattice? To answer this, it can help to know about generators and their algebras (some results are known for 3×3-matrices [Asl]). Here is the question.

Question 4 If $u, v \in A^0_{\text{sa}}$ and if $\mathcal{P}_{\text{Lin}\{u,v\}}$ is finite, do u and v commute?

rI- and variation closures of exponential families

To study the possible discrepancy $\text{cl}_{rI}(\mathcal{E}) \subsetneq \overline{\mathcal{E}}$ in Example 7.21 with the Staffelberg family in a general context, it is necessary to create and evaluate more examples. We do know that the combinatorial mean value parametrization is discontinuous if and only if $\text{cl}_{rI}(\mathcal{E}) \subsetneq \overline{\mathcal{E}}$ (7.28). In Question 8 on page 158 the positions of these discontinuities are proposed. Question 9 on page 160 and Question 10 on page 161 make shape proposals about the variation closure $\overline{\mathcal{E}}$. Question 11 on page 162 is an extension of Question 8 about the location of discontinuities of entropy distance.

Stochastic dependencies

- Nonseparability [Ho]: a precise knowledge of the set of factorizable states can only be of advantage when studying entanglement.

- Is it possible to find a connection between multi-information and phase coexistence (beyond the Ising model [Ma, Er]) for arbitrary classical and quantum spin systems?

State estimation theory

- Do our geometric results have consequences for quantum state estimation [Pe08]?
- Can a coherent state estimation theory be related to quantum exponential families? This question requires infinite-dimensional Hilbert spaces.

Information geometry

Can the Pythagorean theorem of relative entropy be extended to the whole state space? (This is true for probability measures of finite support [Cs03].)

Convex exponential families

What is the meaning of the commutator relation (see Section 8.4) for the algebraic classification of convex families?

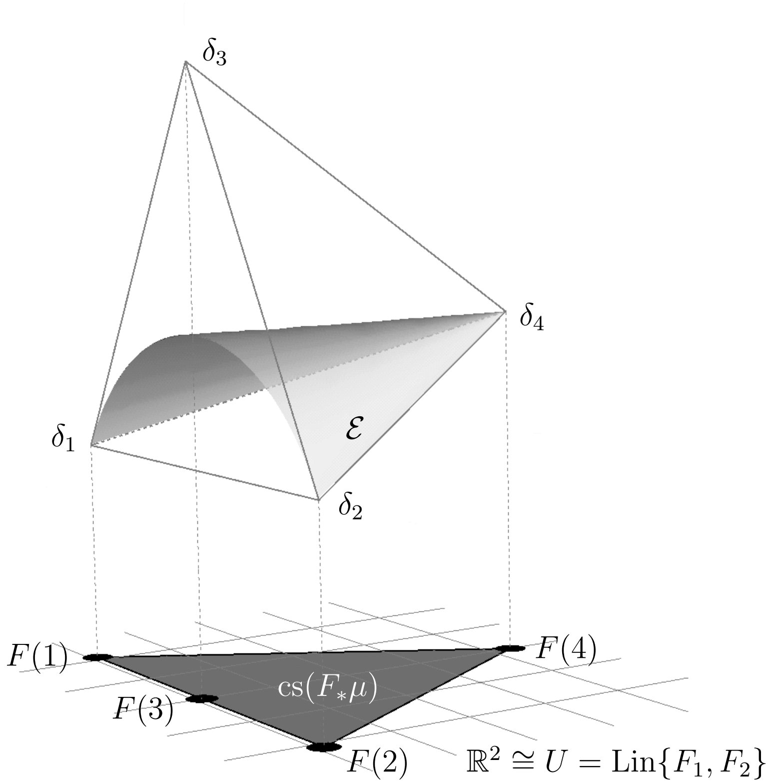


Figure 1.3: **The mean value chart in Section 1.3.** A probability measure P on $X := \{1,2,3,4\}$ is identified with the density function $p := \frac{\mathrm{d}P}{\mathrm{d}\mu}$ in the probability simplex $\mathbb{P}^4 \subset \mathbb{R}^X$ for counting measure μ on X. The drawing is a piece of a hyperplane of $\mathbb{R}^X$ including $F_1 := \frac{1}{\sqrt{2}}(1,-1,0,0)$ and $F_2 := \frac{1}{2\sqrt{3}}(1,1,1,-3)$ and the probability simplex $\mathbb{P}^4$ shifted in a direction perpendicular to $U := \mathrm{Lin}\{F_1, F_2\}$. The mean value of P is the orthogonal projection $\mathbb{E}_P[F] = \pi_U(p)$ to U. Thus, the values of the statistic $F(1), F(2), F(3), F(4)$ are projections of the vertices of $\mathbb{P}^4$ and their convex hull is the convex support $\mathrm{cs}(F_*\mu) = \pi_U(\mathbb{P}^4)$. Here this is a triangle with vertices $F(1), F(2), F(4)$. The exponential family $\mathcal{E}$ with densities $p_{\mu,\theta}(x) := \exp(\theta^1 F_1(x) + \theta^2 F_2(x) - \Lambda(\theta))$ is depicted as the gray curved surface inside $\mathbb{P}^4$. The restriction $\pi_{\mathcal{E}} = \mathbb{E}_{(\cdot)}[F]|_{\mathcal{E}}$ is the mean value chart for $\mathcal{E}$. The inverse $\pi_{\mathcal{E}}^{-1}$, being defined on the interior $\mathrm{int}(\mathrm{cs}(F_*\mu))$, is the maximum-likelihood estimator for $\mathcal{E}$. Provided $\mathbb{E}_P[F] \in \mathrm{int}(\mathrm{cs}(F_*\mu))$, the rI-projection of P to $\mathcal{E}$ is $\Pi_{P\to\mathcal{E}} = \pi_{\mathcal{E}}^{-1}(\mathbb{E}_P(F))$, and one has $\inf_{Q\in\mathcal{E}} D(P||Q) = D(P||\Pi_{P\to\mathcal{E}})$ for the entropy distance of P from $\mathcal{E}$.

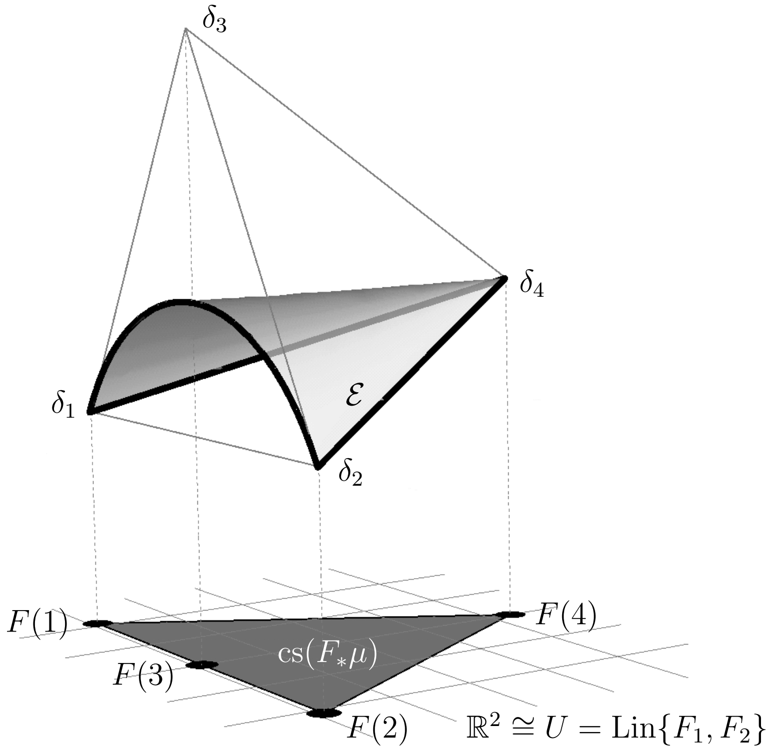


Figure 1.4: **The closure in Section 1.3.** The example from Figure 1.3 is completed to the closure $\overline{\mathcal{E}}$. The convex support $\mathrm{cs}(F_*\mu)$ has 8 faces, the empty set $\emptyset$, the three vertices $F(1), F(2), F(4)$, the three sides $[F(1), F(2)]$, $[F(2), F(4)]$, $[F(4), F(1)]$ and the triangular convex support. To each non-empty face G of $\mathrm{cs}(F_*\mu)$ there corresponds a set of conditional probability distributions $\mathcal{E}_{\mu^G} := \{P(\,\cdot\,|F^{-1}(G)) : P \in \mathcal{E}\}$. This set is an exponential family by itself and the union of these "conditional families" for non-empty faces G is the closure of $\mathcal{E}$. The restriction $\pi_{\overline{\mathcal{E}}} = \mathbb{E}_{(\cdot)}[F]|_{\overline{\mathcal{E}}}$ is a homeomorphism $\overline{\mathcal{E}} \to \mathrm{cs}(F_*\mu)$. The inverse $\pi_{\overline{\mathcal{E}}}^{-1}$ is suitable to extend the rI-projection. One defines for a probability measure P on X the projection $\Pi_{P\to\mathcal{E}} := \pi_{\overline{\mathcal{E}}}^{-1} \circ \pi_U(p)$ where $p = \frac{\mathrm{d}\,P}{\mathrm{d}\,\mu}$ is the Radon-Nikodym derivative with respect to counting measure μ on X. Then, the entropy distance from $\mathcal{E}$ is given by $\inf_{Q\in\mathcal{E}} D(P||Q) = D(P||\Pi_{P\to\mathcal{E}})$.

1.8 Notation

$\mathbb{N} := \{1, 2, 3, \ldots\}$
$\mathbb{N}_0 := \{0, 1, 2, \ldots\}$

Definitions, remarks, examples, lemmas, propositions, corollaries and theorems are labeled by two Roman numbers "chapter.number" and are referred to as "Definition chapter.number", and so on. Equations are labeled by two Roman numbers "chapter.number" and are referred to as "(chapter.number)".

For spatial reasons we write column vectors as rows vectors unless otherwise specified.

2 Finite-dimensional C*-algebras

A C*-algebra is an abstraction of a (closed) algebra of linear operators acting on a Hilbert space. In the finite-dimensional case it is an appropriate setting to study a finite-level quantum system. A C*-algebra has the advantages of a greater flexibility through the axiomatic definition and of the independence of representation compared to an algebra of matrices.

In this chapter we give proofs only for easy statements that may not be obvious for a novice. The literature is cited with the corresponding statements.

2.1 The Gelfand-Naimark theorem

We recall the geometry of real and complex Hilbert spaces, the representation theorem for a C*-algebra and (continuous) functional calculus. We begin with the definition of a C*-algebra.

Definition 2.1 **(a)** A (complex) **algebra** A is a vector space over the complex field $\mathbb{C}$ where a multiplication is defined as a binary operation that satisfies the distributive law with respect to addition, see 1.1 in [Mu] for more details. The algebra A is called **abelian** or **commutative** if for all elements $a, b \in A$

$$ab - ba = 0.$$

In physics language we will sometimes refer to an abelian algebra as the **classical case** and to a non-abelian algebra as the **quantum case**. If the algebra contains a multiplicative identity $\mathbb{1}$ then the algebra is called **unital**. A **subalgebra** B of an algebra A is a subset of A closed under the operations in A, except possibly inclusion of the identity $\mathbb{1}$. A subalgebra of A that contains the identity $\mathbb{1}$ of A is called a **unital subalgebra** of A.

(b) Let $\mathbb{K}$ be the field $\mathbb{R}$ of real numbers or the field $\mathbb{C}$ of complex numbers. A **norm** on a vector space H over $\mathbb{K}$ is a mapping $\|\cdot\| : H \to \mathbb{R}_0^+$ such that for $a, b \in H$ and $\alpha \in \mathbb{K}$ hold

$$\begin{array}{rl} \text{(i)} & \|a\| = 0 \quad \Longleftrightarrow \quad a = 0, \\ \text{(ii)} & \|\alpha a\| = |\alpha| \|a\|, \\ \text{(iii)} & \|a + b\| \leq \|a\| + \|b\|. \end{array}$$

An algebra with a norm is a **normed algebra**.

We follow the definition of a C*-algebra in [Da].

Definition 2.2 A **Banach algebra** A is a complex normed algebra which is complete (as a topological space) and satisfies

$$\|ab\| \leq \|a\| \|b\| \quad \text{for all} \quad a, b \in A.$$

A **Banach *-algebra** A is a complex Banach algebra with a conjugate linear involution * (called the **adjoint**) which is an anti-isomorphism. That is, for all a, b in A and $\lambda \in \mathbb{C}$,

$$\begin{aligned} (a+b)^* &= a^* + b^* \\ (\lambda a)^* &= \overline{\lambda} a^* \\ a^{**} &= a \\ (ab)^* &= b^* a^*. \end{aligned} \tag{2.1}$$

A **C*-algebra** A is a Banach *-algebra with the additional norm condition

$$\|a^* a\| = \|a\|^2 \quad \text{for all} \quad a \in A. \tag{2.2}$$

A norm satisfying (2.2) is a **C*-norm**. An element a of a C*-algebra A is **self-adjoint** if $a^* = a$, it is **normal** if $a^* a = a a^*$, it is unitary if $a^* a = a a^* = \mathbb{1}$.

We introduce notation and recall frequently used properties of Hilbert spaces.

Definition 2.3 Let H be a $\mathbb{K}$-vector space. A **sesquilinear form** on H is a mapping $\langle \cdot, \cdot \rangle : H \times H \to \mathbb{K}$ that is conjugate linear in the first argument and linear in the second. A sesquilinear form $\langle \cdot, \cdot \rangle$ on H is **positive** (semidefinite) if $\langle x, x \rangle \geq 0$ for all $x \in H$. An

inner product $\langle\cdot,\cdot\rangle$ on H is a positive sesquilinear form on H such that $\langle x,x\rangle = 0$ if and only if $x = 0$. Then the tuple $(H, \langle\cdot,\cdot\rangle)$ is called an **inner product space**. If $\mathbb{K} = \mathbb{R}$ then an inner product space is called a **Euclidean space**. If $\mathbb{K} = \mathbb{C}$ it is a **Hermitian space**. A Hermitian space that is complete (as a topological space) is a **Hilbert space**, an Euclidean space that is complete is a **real Hilbert space**. For an inner product $\langle\cdot,\cdot\rangle$ on H we define $\|x\|_2 := \sqrt{\langle x,x\rangle}$, $x \in H$.

Remark 2.4 For an inner product on a $\mathbb{K}$-vector space H, the **Schwarz inequality**

$$|\langle x,y\rangle| \leq \|x\|_2\|y\|_2 \tag{2.3}$$

and the **triangle inequality**

$$\|x+y\|_2 \leq \|x\|_2 + \|y\|_2, \tag{2.4}$$

hold for $x, y \in H$. These inequalities are proved in §64 in [Hal].

Definition 2.5 In an inner product space $(H, \langle\cdot,\cdot\rangle)$ we call the mapping $x \mapsto \|x\|_2$ for $x \in H$ the **two-norm**. A vector $u \in H$ is **normalized** and u is a **unit vector** if $\|u\|_2 = 1$. Two vectors $u, v \in H$ are **orthogonal**, written $u \perp v$, if $\langle u,v\rangle = 0$. Two subsets $U, V \subset H$ are **orthogonal** if each element of U is orthogonal to each element of V, written $U \perp V$. The **orthogonal complement** of U in H is the subspace

$$U^\perp := \{v \in H : v \perp u \quad \text{for all} \quad u \in U\}.$$

A set S of vectors in H is an **orthonormal set** if the vectors in S are normalized and mutually orthogonal. If S is an orthonormal set and no other orthonormal set contains S as a proper subset then S is called an **orthonormal basis** for H or an **ONB** for H. The **operator norm** of a linear operator a on H is defined by

$$\|a\| := \sup_{\substack{x\in H \\ \|x\|_2\leq 1}} \|a(x)\|_2. \tag{2.5}$$

A linear operator a on H is **bounded** if the operator norm $\|a\|$ if finite. The set of bounded operators on H is denoted by $\mathcal{B}(H)$.

Remark 2.6 **(a)** In an inner product space $(H, \langle\cdot,\cdot\rangle)$ the two-norm $x \mapsto \|x\|_2$ for $x \in H$ is a norm in the sense of Definition 2.1 (b). This follows from the triangle inequality (2.4) and the definition of an inner product. In all our applications, the space H has finite dimension. Then H is complete for the topology of the two-norm, a Hermitian space is a Hilbert space and an Euclidean space is a real Hilbert space.

(b) For a bounded operator $a \in \mathcal{B}(H)$ there exists a unique bounded operator $a^* \in \mathcal{B}(H)$ such that $\langle a^*(x), y \rangle = \langle x, a(y) \rangle$ holds for all $x, y \in H$ (Theorem 2.3.1 in [Mu]). We call a^* the **adjoint** of a.

(c) The algebra $\mathcal{B}(H)$ of bounded operators on a Hilbert space H is a C*-algebra with C*-norm the operator norm (2.5) and with the adjoint defined in (b), see Lemma 2.1.3 in [Mu]. One can argue as follows. If $a \in \mathcal{B}(H)$ then the inequality $\|a^*a\| \leq \|a^*\|\|a\|$ is elementary to prove. Using the Schwarz inequality (2.3) we have $\|a\|^2 \leq \|a^*a\|$. This proves $\|a^*\| = \|a\|$ and one obtains $\|a\|^2 = \|a^*a\|$. More general, any closed and self-adjoint subalgebra of $\mathcal{B}(H)$ is a C*-algebra. A closed and self-adjoint subalgebra of $\mathcal{B}(H)$ is called a **concrete C*-algebra**.

(d) The **Gelfand-Naimark theorem** (Theorem I.9.12 in [Da]) says a concrete C*-algebra is the general example of a C*-algebra: Every C*-algebra is isometrically *-isomorphic to a concrete C*-algebra of operators. A simple example for $k \in \mathbb{N}$ is the **full matrix algebra** M_k. This is the space of complex $k \times k$ matrices. For completeness we put $M_0 = \{0\}$. By Theorem III.1.1 in [Da] a finite-dimensional C*-algebra A is *-isomorphic to the direct sum of full matrix algebras

$$A \cong M_{n_1} \oplus \cdots \oplus M_{n_N}, \tag{2.6}$$

where $n = (n_1, \ldots, n_N) \in \mathbb{N}_0^N$ is a multi-index and $N \in \mathbb{N}$.

Definition 2.7 A finite-dimensional concrete C*-algebra A of the form (2.6) will be referred to as a **matrix algebra**. We abbreviate $|n| := \sum_{i=1}^N n_i$. Then $A \hookrightarrow M_{|n|}$ is a natural embedding and a C*-norm on A is given by $\|a\| := \sup_{\substack{x \in H \\ \|x\|_2 = 1}} \|a(x)\|_2$. where a matrix a is considered a linear operator on the Hilbert space

$$H := H_1 \oplus \cdots \oplus H_N \tag{2.7}$$

for $H_i := \mathbb{C}^{n_i}$, $i = 1, \ldots, N$ such that $H = \mathbb{C}^{|n|}$. The two-norm $\| \cdot \|_2$ above is the norm on the Hilbert space $H = \mathbb{C}^{|n|}$ induced by the inner product for $x, y \in H$

$$\langle x, y \rangle := \sum_{i=1}^{|n|} \overline{x_i} y_i. \tag{2.8}$$

The abelian shaping of the Gelfand-Naimark Theorem is the Gelfand transform and it is the representation of an abelian C*-algebra as an algebra of complex valued functions. This is useful to define continuous functional calculus.

Definition 2.8 Let A be a unital C*-algebra. The **spectrum** of $a \in A$ is the set

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

the **spectral radius** of a is the non-negative number $\mathrm{spr}(a) := \max\{|\lambda| : \lambda \in \mathrm{spec}(a)\}$. The **spectral norm** of $a \in A$ is defined by $\sqrt{\mathrm{spr}\,(a^*a)}$. A C*-subalgebra of A is a subset of A that is closed under the algebraic operations of A. The C*-algebra **generated** by an element $a \in A$ is the smallest C*-subalgebra of A that contains a and this algebra is denoted $C^*(a)$. The operator a is **positive** if $\mathrm{spec}(a)$ is non-negative. If a is positive then we write $a \geq 0$.

Remark 2.9 **(a) Spectral invariance.** Let B be a unital C*-algebra and $A \subset B$ be a unital C*-subalgebra. *A priori*, the spectrum of an element $a \in A$ calculated in A can be larger than the spectrum of a calculated in B because there are more possible inverses for a in the larger algebra B. It turns out that the spectrum of an element a in a C*-algebra is invariant under a unital algebra embedding (Corollary I.5.7 in [Da]).

(b) If a is an element in a unital C*-algebra and a is normal, that is $a^*a = aa^*$, then there exists a C*-isomorphism (Corollary I.3.2 in [Da])

$$\Gamma : C^*(a) \to C(\mathrm{spec}(a))$$

called the **Gelfand transform**. Here we denote by $C(\mathrm{spec}(a))$ the abelian algebra of continuous functions $\mathrm{spec}(a) \to \mathbb{C}$ endowed with the supremum norm $\|b\| := \sup_{x \in \mathrm{spec}(a)} |b(x)|$ for $b \in C(\mathrm{spec}(a))$.

(c) If $a \in A$ for a unital C*-algebra A then the spectral norm and C*-norm are equal,

$$\|a\| = \sqrt{\mathrm{spr}(a^*a)} \tag{2.9}$$

(Corollary I.3.4). This follows immediately from the Gelfand transform in (b) and from equality of the spectrum of a in $C^*(a)$ and in A, see (a).

(d) If a is a normal element in a unital C*-algebra and f is a continuous function on $\mathrm{spec}(a)$ then there is an operator $f(a)$ defined $f(a) := \Gamma^{-1} \circ f \circ \Gamma$ by the Gelfand transform Γ. The association that transforms a complex valued function on $\mathrm{spec}(a)$ into an operator valued function on $C^*(a)$ is called **continuous functional calculus**. If g is continuous on $f(\mathrm{spec}(a))$ then $g(f(a)) = g \circ f(a)$. This is proved in in Corollary I.3.3 in [Da]. The continuous functional calculus extends a polynomial definition of operator functions because the Gelfand transform is an algebra isomorphism. In the case of a finite dimensional C*-algebra, the spectrum of a is finite. Then it follows that an arbitrary function on $\mathrm{spec}(a)$ is polynomial, for it can be written by interpolation with Newton polynomials.

(e) If A is a concrete C*-algebra of linear operators on a Hilbert space H then from the Gelfand-Naimark Theorem and from the Gelfand transform we obtain equality of operator norm (2.5) and spectral norm. For $a \in A$ holds

$$\|a\| = \sup_{\substack{x \in H \\ \|x\|_2 \leq 1}} \|a(x)\|_2 = \sqrt{\mathrm{spr}(a^*a)}.$$

2.2 Analysis in matrix algebras

A matrix algebra (Definition 2.7) is our working model of a finite-dimensional C*-algebra. It is specified by a natural number $N \in \mathbb{N}$ and a multi-index $n = (n_1, \ldots, n_N) \in \mathbb{N}_0^N$. A matrix algebra A is a direct sum of full matrix algebras

$$A := M_{n_1} \oplus \cdots \oplus M_{n_N}$$

acting on the Hilbert space $H := H_1 \oplus \cdots \oplus H_N$ for $H_i := \mathbb{C}^{n_i}$, $i = 1, \ldots, N$. We describe the most frequently used methods.

The positive cone of a matrix algebra induces a partial ordering on the self-adjoint matrices. We recall the Hilbert-Schmidt inner product and trace norm. Orthogonal projectors are used to write the spectral theorem and functional calculus. Compressions are introduced in analogy with Alfsen and Schultz [Al]. They can generalize conditional probability distributions to non-abelian algebras.

Definition 2.10 The **positive cone** of A consists of positive matrices

$$A^+ := \{a \in A : a \geq 0\}. \tag{2.10}$$

Remark 2.11 Let $a \in A$ be a matrix. **(a)** Then a is positive if and only if $a = b^2$ for some self-adjoint matrix $b \in A$ (Lemma I.4.3 in [Da]).

(b) The matrix a^*a is positive (Theorem I.4.5 in [Da]).

(c) The relation $a \geq 0$ implies $b^*ab \geq 0$ for every $b \in A$ as a consequence of (a) and (b).

(d) One has $a \geq 0$ if and only if $\langle a(x), x\rangle \geq 0$ for all $x \in H$ (Theorem 2.3.5 in [Mu]).

(e) If $a \geq 0$ then by functional calculus in Remark 2.9 (d) there exists a (positive) **square-root** $\sqrt{a} \in \mathrm{C}^*(a) \subset A$, that satisfies $\sqrt{a} \geq 0$ and $(\sqrt{a})^2 = a$.

(f) Every self-adjoint matrix $a \in A_{\mathrm{sa}}$ has a decomposition $a = a^+ - a^-$ for two positive matrices $a^+, a^- \in A^+$. This follows from functional calculus applied to the real functions $x \mapsto \max(0, x)$ and $x \mapsto \max(0, -x)$.

On the matrix algebra A we can use the C*-norm. The following two norms are used, too. The Hilbert-Schmidt norm is a two-norm, it is induced by an inner product.

Definition 2.12 If $\{x_i\}$ is an ONB of the Hilbert space $H = \mathbb{C}^{|n|}$ then the **trace** of a matrix $a \in A$ is

$$\mathrm{tr}(a) := \sum_i \langle a(x_i), x_i\rangle \tag{2.11}$$

The **Hilbert-Schmidt inner product** or **HS inner product** for $a, b \in A$ is given by

$$(a, b) \mapsto \langle a, b\rangle := \mathrm{tr}(a^*b)$$

and the **Hilbert-Schmidt norm** is $\|a\|_2 := \sqrt{\langle a, a\rangle}$. The **trace norm** of a matrix $a \in A$ is given by

$$\|a\|_1 = \mathrm{tr}\,|a|$$

where $|a| := \sqrt{a^*a}$ is defined by functional calculus.

Remark 2.13 **(a)** The HS inner product is an inner product on A [Du].

(b) If $a, b \in A$ are positive matrices then $a \perp b$ if and only if $ab = 0$. [Proof on page 190]

(c) The trace norm is a norm [Ni]. By Lemma 2.57 in [Al] one has the inequality for $a, b \in A$

$$|\,\mathrm{tr}(ab)| \leq \|a\|_1 \|b\| \tag{2.12}$$

with trace norm and spectral norm on the right-hand side.

On the matrix algebra A some special notation is used.

Definition 2.14 The real vector space of self-adjoint matrices is denoted

$$A_{\mathrm{sa}} := \{a \in A : a^* = a\} \tag{2.13}$$

The HS inner product on A restricts to the Euclidean inner product $(a, b) \mapsto \langle a, b\rangle = \mathrm{tr}(ab)$ on A_{sa}. Unless otherwise specified, we understand the orthogonal complement of a subset $U \subset A_{\mathrm{sa}}$ as the complement in A_{sa}, that is

$$U^{\perp} := \{b \in A_{\mathrm{sa}} : b \perp a \quad \text{for all} \quad a \in U\}.$$

We define the real vector space of **traceless self-adjoint** matrices ($i = 0$) and the real affine space of **trace one self-adjoint** matrices ($i = 1$) as

$$A^i_{\mathrm{sa}} := \{a \in A_{\mathrm{sa}} : \mathrm{tr}(a) = i\}. \tag{2.14}$$

The space of self-adjoint matrices has dimension

$$\dim_{\mathbb{R}}(A_{\mathrm{sa}}) = \dim_{\mathbb{C}}(A) = \sum_{i=1}^{N} n_i^2. \tag{2.15}$$

Definition 2.15 An **orthogonal projector** in A is a matrix $p \in A$ with $p = p^* = p^2$.

$$\mathcal{P}(A) := \{p \in A : p = p^* = p^2\} \tag{2.16}$$

is the **projector lattice** of A. The projectors 0 and $\mathbb{1}$ are **improper** projectors. All other projectors are **proper**. A set $F \subset \mathcal{P}(A)$ of projectors is a **complete set of projectors** if $\sum_{p \in F} p = \mathbb{1}$.

Remark 2.16 **(a)** Every orthogonal projector $p \in \mathcal{P}(A)$ is positive. This follows from $p^2 = p$ by Remark 2.11 (a).

(b) Distinct members p, q of a complete set of projectors in A satisfy $pq = 0$.

[Proof on page 190]

(c) If $a \in A$ is a normal matrix then the **spectral theorem** holds: there exists a complete set $\{p_\mu(a)\}_{\mu \in \mathrm{spec}(a)}$ of orthogonal projectors such that

$$a = \sum_{\mu \in \mathrm{spec}(a)} \mu p_\mu(a). \tag{2.17}$$

This sum is called **spectral decomposition** of a. For $\mu \in \mathrm{spec}(a)$ the orthogonal projector $p_\mu(a)$ is the **spectral projector** of a for μ. Notice that the projector $p_\mu(a)$ belongs to A, indeed it is a polynomial in a, see Satz 5.25 in [Kn01].

(d) If $a \in A$ is a normal matrix with spectral decomposition $a = \sum_{\mu \in \mathrm{spec}(a)} \mu p_\mu(a)$ and if f is a complex function on the spectrum $\mathrm{spec}(a)$ of a, then the continuous functional calculus in Remark 2.9 (d) is (reasoning with [Da] on page 190)

$$f(a) = \sum_{\mu \in \mathrm{spec}(a)} f(\mu) p_\mu(a). \tag{2.18}$$

Definition 2.17 Let $a \in A$ be a normal matrix and with spectral decomposition (2.17) given by $a = \sum_{\mu \in \mathrm{spec}(a)} \mu p_\mu(a)$. The **kernel projector** of a is defined as the orthogonal projector

$$k(a) := \begin{cases} p_0(a) & \text{if } 0 \in \mathrm{spec}(a), \\ 0 & \text{otherwise.} \end{cases} \tag{2.19}$$

The **support projector** of a is the orthogonal projector

$$s(a) := \mathbb{1} - k(a). \tag{2.20}$$

A normal matrix a **dominates** a normal matrix b if $s(b) \leq s(a)$. Let $a \in A$ be a self-adjoint matrix. We will use very frequently the **minimal eigenvalue** and **minimal projector** of a,

$$\mu_-(a) := \min\{\mathrm{spec}(a)\} \qquad \text{and} \qquad p_-(a) := p_{\mu_-(a)}(a), \tag{2.21}$$

as well as the **maximal eigenvalue** and **maximal eigenprojector** of a,

$$\mu_+(a) := \max\{\mathrm{spec}(a)\} \qquad \text{and} \qquad p_+(a) := p_{\mu_+(a)}(a). \tag{2.22}$$

Remark 2.18 If $a \in A$ is a normal matrix, then kernels and images of a and of the adjoint a^* coincide (Satz 8.13 in [Kn01])

$$\mathrm{Im}(a^*) = \mathrm{Im}(a) \qquad \text{and} \qquad \ker(a^*) = \ker(a)$$

and the Hilbert space H is the direct orthogonal sum (Satz 8.14 in [Kn01])

$$H = \mathrm{Im}(a) \oplus \ker(a).$$

By the finite dimensionality of H one has

$$\mathrm{Im}(a) = \ker(a)^{\perp} \qquad \text{and} \qquad \ker(a) = \mathrm{Im}(a)^{\perp}. \tag{2.23}$$

The image of a and of the support projector $s(a)$ are equal,

$$\mathrm{Im}(s(a)) = \mathrm{Im}(a). \tag{2.24}$$

This equation is not standard in the infinite-dimensional case. Usually, the projector with image $\mathrm{Im}(a)$ is the range projector, whereas the support projector of a has the image $\ker(a)^{\perp}$ [Al]. Further, from the spectral theorem we get $s(a)_i = s(a_i)$ for the direct sum summands $i = 1, \ldots, N$ of the matrix algebra $A = M_{n_1} \oplus \cdots \oplus M_{n_N}$. Thus by (2.24) holds

$$\mathrm{Im}(s(a)_i) = \mathrm{Im}(a_i). \tag{2.25}$$

Another consequence of (2.24) and using (2.23) for normal matrices $a, b \in A$ is

$$ab = 0 \qquad \Longleftrightarrow \qquad s(a)s(b) = 0.$$

If $a, b \in A^+$ are positive matrices then by Remark 2.13 (b) one has also

$$a \perp b \qquad \Longleftrightarrow \qquad s(a)s(b) = 0. \tag{2.26}$$

The following concept will be used to describe the generalization of conditional probability distributions in non-abelian algebras. For an orthogonal projector $p \in \mathcal{P}(A)$ we consider the finite dimensional C*-algebra

$$pAp = \{pap : a \in A\}.$$

This algebra does not contain the identity $\mathbb{1}$ of A unless $p = \mathbb{1}$. The identity of pAp is p. The algebra pAp is called compression in [Al]. We give a slightly different definition.

Definition 2.19 For an orthogonal projector $p \in \mathcal{P}(A)$ we define the **compression** of A by p as the matrix algebra

$$A^p := M_{\mathrm{rk}(p_1)} \oplus \cdots \oplus M_{\mathrm{rk}(p_N)} \tag{2.27}$$

We denote by $\mathbb{1}^p$ the multiplicative identity of A^p and by 0^p the zero element of A^p. The compression A^p is naturally embedded in the full matrix algebra $M_{\mathrm{rk}(p)}$. The matrices in A^p are considered linear operators on the Hilbert space

$$H^p := H_1^p \oplus \cdots \oplus H_N^p$$

for $H_i^p := \mathbb{C}^{\mathrm{rk}(p_i)}$, $i = 1, \ldots, N$. On $H^p = \mathbb{C}^{\mathrm{rk}(p)}$ we use the inner product for $x, y \in \mathbb{C}^{\mathrm{rk}(p)}$

$$\langle x, y \rangle := \sum_{i=1}^{\mathrm{rk}(p)} x_i^* y_i.$$

For an ONB $\{x_i\}_{i=1}^{\mathrm{rk}(p)}$ of H^p we use on the compression A^p the trace

$$\mathrm{tr}(a) := \sum_{i=1}^{\mathrm{rk}(p)} \langle a(x_i), x_i \rangle.$$

Remark 2.20 Modulo a trace-preserving *-automorphism of A^p there exists a unique trace-preserving *-isomorphism $pAp \to A^p$. To construct one at least, we proceed as follows. Since $\dim(H_i^p) = \mathrm{rk}(p_i)$ we can choose an isometric linear bijection $\iota : H^p \to \mathrm{Im}(p)$ to the image of p such that for $i = 1, \ldots, N$

$$\iota(0 \oplus \cdots \oplus 0 \oplus H_i^p \oplus 0 \oplus \cdots \oplus 0) = \mathrm{Im}(0 \oplus \cdots \oplus 0 \oplus p_i \oplus 0 \oplus \cdots \oplus 0)$$

holds. Then $a \mapsto \iota^{-1} \circ a \circ \iota$ is a trace-preserving *-isomorphism from pAp to A^p. If two trace-preserving *-isomorphism $\alpha, \beta : pAp \to A^p$ are given then the concatenation $\beta \circ \alpha^{-1}$ is a trace-preserving *-automorphism of A^p.

Definition 2.21 With Remark 2.20 we choose for each orthogonal projector $p \in \mathcal{P}(A)$ a trace-preserving *-isomorphism $pAp \to A^p$ and we denote the inverse by

$$\kappa^p : A^p \to pAp. \tag{2.28}$$

The choice of κ^p is unique modulo a trace-preserving *-automorphism of A^p and κ^p will not be specified further than these properties.

Remark 2.22 Let $p = p_1 \oplus \cdots \oplus p_N \in \mathcal{P}(A)$ be an orthogonal projector.

(a) By definition (2.27) of the compression and by the dimension formula (2.15) for the self-adjoint part of A we have

$$\dim_{\mathbb{R}}((A^p)_{\mathrm{sa}}) = \dim_{\mathbb{C}}(A^p) = \sum_{i=1}^{N} \mathrm{rk}(p_i)^2. \tag{2.29}$$

(b) By the spectral theorem we have for a normal matrix $a \in (A^p)_{\rm sa}$

$$\kappa^p(s(a)) = s(\kappa^p(a)) \tag{2.30}$$

with the support projector (2.20). Since κ^p is a *-monomorphism, we obtain the relations

$$\begin{array}{rcl} \kappa^p\left((A^p)_{\rm sa}\right) & = & \{a \in A_{\rm sa} : s(a) \leq p\}, \\ \kappa^p\left((A^p)^+\right) & = & \{a \in A^+ : s(a) \leq p\}, \\ \kappa^p\left(\mathcal{P}(A^p)\right) & = & \{q \in \mathcal{P}(A) : q \leq p\} \end{array} \tag{2.31}$$

for the space of self-adjoint matrices (2.13), the positive cone (2.10) and the projector lattice (2.16). Since κ^p is trace-preserving we also find for $i = 0, 1$

$$\kappa^p\left((A^p)^i_{\rm sa}\right) = \{a \in A^i_{\rm sa} : s(a) \leq p\} \tag{2.32}$$

for the traceless and trace one self-adjoint matrix spaces (2.14).

(c) For a function $f : U \subset \mathbb{C} \to \mathbb{C}$ and a normal matrix $a \in A^p$ with $\operatorname{spec}(a) \subset U$, the functional calculus (2.18) is transported. Provided that $0 \in U$ one has

$$\kappa^p\left(f(a)\right) = p\, f\left(\kappa^p(a)\right). \tag{2.33}$$

2.3 The Grassmannian

The logic of classical physics is typically the Boolean algebra of subsets of a phase space, the logic of quantum physics is an orthomodular lattice consisting of closed linear subspaces of a Hilbert space, called the Grassmannian. In general this is not a Boolean algebra for the missing distributive law, see Figure 2.1. In both cases the logic is a space of elementary events in the sense introduced in Definition 1.14. Its axiomatization in lattice theory is due to Birkhoff, von Neumann and Husimi mainly. The lattice theory of events is treated in Kalmbach [KaG], the operator algebra counterpart including geometry of state spaces is treated by Alfsen and Schultz [Al]. Our recommended reference to general lattice theory is Birkhoff [Bi].

We recall the necessary lattice theory to describe the Grassmannian of a von Neumann algebra. In the finite dimensional case we recall the structure of the Grassmannian as a union of compact differentiable manifolds by standard arguments from algebraic geometry [Har, Wey] and differential topology [Hi]. These are complemented by metric properties described in Avron, Seiler and Simon [Av].

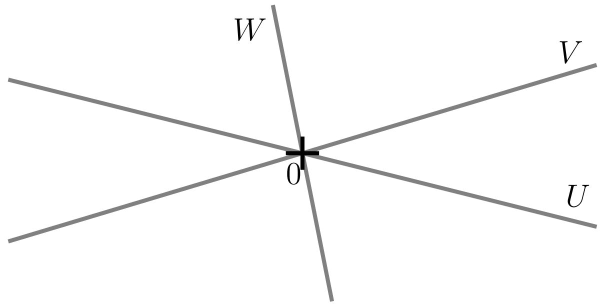


Figure 2.1: The Grassmannian $\mathcal{G}(\mathbb{C}^2) = \{\text{linear subspaces of } \mathbb{C}^2\}$ has no distributive law. Three mutually distinct one-dimensional subspaces $U, V, W \subset \mathbb{C}^2$ have $U + (V \cap W) = U$ and $(U + V) \cap (U + W) = \mathbb{C}^2$.

Definition 2.23 The **Grassmannian** $\mathcal{G}(\mathcal{H})$ of a Hilbert space $\mathcal{H}$ is the space of all closed linear subspaces. Denote $\mathcal{B}(\mathcal{H})$ the set of bounded operators on a Hilbert space $\mathcal{H}$ and let $\mathcal{D}' = \{a \in \mathcal{B}(\mathcal{H}) : ab = ba \text{ for all } b \in \mathcal{D}\}$ for a subalgebra $\mathcal{D}$ of $\mathcal{B}(\mathcal{H})$. A C*-subalgebra $\mathcal{A}$ of $\mathcal{B}(\mathcal{H})$ is a **von Neumann algebra** if $\mathcal{A} = \mathcal{A}''$. The **projector lattice** is denoted

$$\mathcal{P}(\mathcal{A}) := \{p \in \mathcal{A} : p^2 = p = p^*\},$$

the Grassmannian of $\mathcal{A}$ is defined as the set of images of projectors

$$\mathcal{G}(\mathcal{A}) := \{\mathrm{Im}(p) \subset \mathcal{H} : p \in \mathcal{P}(\mathcal{A})\}. \tag{2.34}$$

We denote the **positive cone** of $\mathcal{A}$ by $\mathcal{A}^+ := \{a \in \mathcal{A} : a \geq 0\}$. The space of self-adjoint operators is denoted by $\mathcal{A}_{\text{sa}} := \{a \in \mathcal{A} : a^* = a\}$.

A **partially ordered set** or a **poset** is a tuple $(M, \leq)$ consisting of a set M and a relation $\leq$ on $M \times M$ such that for $x, y, z \in M$ holds

$$\begin{array}{lllll} \text{(a)} & x \leq x, & & & \\ \text{(b)} & x \leq y & \text{and} \quad y \leq x & \text{implies} & x = y, \\ \text{(c)} & x \leq y & \text{and} \quad y \leq z & \text{implies} & x \leq z. \end{array}$$

The relation $\leq$ is called a **partial ordering** on M.

Remark 2.24 The positive cone induces a partial ordering on $\mathcal{A}_{\text{sa}}$. The relation

$$a \leq b :\iff b - a \geq 0 \tag{2.35}$$

defined for self-adjoint operators $a, b \in \mathcal{A}_{\mathrm{sa}}$ is a partial ordering on $\mathcal{A}_{\mathrm{sa}}$. One way to prove this is using Remark 2.11 (d) [Mu].

Remark 2.25 (a) By definition, the Grassmannian $\mathcal{G}(\mathcal{A})$ of a von Neumann algebra $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$ for a Hilbert space $\mathcal{H}$ consists of the images of orthogonal projectors. Any projector $p^2 = p$ of $\mathcal{A}$ induces the direct sum decomposition of the Hilbert space

$$\mathcal{H} = \mathrm{Im}(p) + \ker(p),$$

given for a vector $x \in \mathcal{H}$ by $x = p(x) + (\mathbb{1} - p)(x)$, cf. Dunford and Schwartz [Du], VI.3. If p is an orthogonal projector, $p^2 = p = p^*$, then one has $\mathrm{Im}(p) \perp \ker(p)$ and this implies

$$\ker(p) = \mathrm{Im}(p)^{\perp} \qquad \text{and} \qquad \mathrm{Im}(p) = \ker(p)^{\perp}. \tag{2.36}$$

So

$$(\ker(p)^{\perp})^{\perp} = \ker(p) \qquad \text{and} \qquad (\mathrm{Im}(p)^{\perp})^{\perp} = \mathrm{Im}(p). \tag{2.37}$$

We notice from (2.37) that kernel and image of a projector are closed subspaces. Since a projector p is characterized by the condition $p(x) = x$ for $x \in \mathrm{Im}(p)$ and by $p(x) = 0$ for $x \in \ker(p)$, we see that the mapping

$$\mathcal{P}(\mathcal{A}) \to \mathcal{G}(\mathcal{A}), \quad p \mapsto \mathrm{Im}(p) \tag{2.38}$$

from the projector lattice $\mathcal{P}(\mathcal{A})$ to the Grassmannian $\mathcal{G}(\mathcal{A})$ is a bijection.

(b) While true in the general C*-algebraic context [KaG] we give short proofs in the finite dimensional case for the following relations of the bijection (2.38). Equivalent are for two orthogonal projectors $p, q \in \mathcal{P}(A)$

$$\begin{array}{rrcl} \text{(i)} & \mathrm{Im}(p) & \subset & \mathrm{Im}(q), \\ \text{(ii)} & \mathrm{Im}(p) & \perp & \mathrm{Im}(q)^{\perp}, \\ \text{(iii)} & p & \leq & q, \\ \text{(iv)} & p & \perp & \mathbb{1} - q, \\ \text{(v)} & pq & = & p. \end{array} \tag{2.39}$$

The bijection $p \mapsto \mathrm{Im}(p)$ is an isomorphism of partially ordered sets. The projector lattice is ordered by the positive cone and the Grassmannian is ordered by set inclusion.

[Proof on page 190]

We discuss the order theoretic structure of the Grassmannian $\mathcal{G}(\mathcal{A})$.

Definition 2.26 A mapping $f : X \to Y$ between two posets $(X, \leq)$ and $(Y, \leq)$ is **isotone**, if $x_1 \leq x_2$ implies $f(x_1) \leq f(x_2)$ for any $x_1, x_2 \in X$. The mapping f is **antitone** if $x_1 \leq x_2$ implies $f(x_2) \leq f(x_1)$. In a poset $\mathcal{L}$, a **lower bound** of a subset $X \subset \mathcal{L}$ is an element $y \in \mathcal{L}$ such that $y \leq x$ for all $x \in X$. An **infimum** of X is a lower bound z of X such that $y \leq z$ for every lower bound y of X. Dually, an **upper bound** of a subset $X \subset \mathcal{L}$ is an element $y \in \mathcal{L}$ such that $x \leq y$ for all $x \in X$. A **supremum** of X is an upper bound z of X such that $z \leq y$ for every upper bound y of X. A **lattice** $(\mathcal{L}, \leq, \wedge, \vee)$ is a poset $(\mathcal{L}, \leq)$ where any two elements $x, y \in \mathcal{L}$ have an infimum and a supremum (then being unique). The infimum of x and y is denoted by $x \wedge y$, the supremum of x and y by $x \vee y$. The partial ordering of $\mathcal{L}$ restricts to a subset $X \subset \mathcal{L}$. We call X a **sublattice** of $\mathcal{L}$ if for all $x, y \in X$ the infimum $x \wedge y$ and the supremum $x \vee y$ (calculated in $\mathcal{L}$) belongs to X.

A lattice $(\mathcal{L}, \leq, \wedge, \vee)$ is **complete** if every subset X of $\mathcal{L}$ has an infimum and a supremum. The infimum of X is denoted $\inf(X)$, the supremum of X is denoted $\sup(X)$ in case of existence. The **smallest element** of $\mathcal{L}$ is $0 := \inf(\mathcal{L})$, the **greatest element** of $\mathcal{L}$ is $1 := \sup(\mathcal{L})$. We denote a complete lattice by $(\mathcal{L}, \leq, \wedge, \vee, 0, 1)$. An element $x \in \mathcal{L}$ is an **atom** if $y \leq x$ and $y \neq x$ implies $y = 0$, and x is a **coatom** if $y \geq x$ and $y \neq x$ implies $y = 1$, for $y \in \mathcal{L}$.

If a lattice $\mathcal{L}$ has a smallest and a greatest element, then by a **complement** of $x \in \mathcal{L}$ we mean an element $y \in \mathcal{L}$ such that $x \wedge y = 0$ and $x \vee y = 1$. If every element in $\mathcal{L}$ has a complement then $\mathcal{L}$ is a **complemented lattice**. An **ortholattice** is a structure $(\mathcal{L}, \leq, \wedge, \vee, ', 0, 1)$ which is a lattice with universal bounds $0 = \inf(\mathcal{L})$, $1 = \sup(\mathcal{L})$ and with an involution $x \mapsto x'$ on $\mathcal{L}$ such that $(x')' = x$ and

$$\begin{array}{rl} x \wedge x' = 0, & \qquad x \vee x' = 1, \\ (x \wedge y)' = x' \vee y', & \qquad (x \vee y)' = x' \wedge y' \end{array} \tag{2.40}$$

holds for all $x, y \in \mathcal{L}$.

A lattice $(\mathcal{L}, \leq, \wedge, \vee)$ is **distributive**, if the **distributive law**

$$\begin{aligned} x \wedge (y \vee z) &= (x \wedge y) \vee (x \wedge z), \\ x \vee (y \wedge z) &= (x \vee y) \wedge (x \vee z) \end{aligned}$$

holds for all $x, y, z \in \mathcal{L}$. A **Boolean lattice** is a complemented distributive lattice. A lattice $(\mathcal{L}, \leq, \wedge, \vee)$ is **modular** if the **modular law**

$$x \leq z \quad \text{implies} \quad x \vee (y \wedge z) = (x \vee y) \wedge z \tag{2.41}$$

holds for elements $x, y, z \in \mathcal{L}$. An **orthomodular lattice** is an ortholattice where the **orthomodular law**

$$x \leq y \quad \text{implies} \quad x \vee (x' \wedge y) = y$$

holds for all $x, y \in \mathcal{L}$.

The following three lemmas describe the order structure for an infinite dimensional von Neumann algebra. The Grassmannian is partially ordered by set inclusion. The infimum of closed linear spaces is their intersection. The orthogonal complement of the inner product of the Hilbert space is used as involution for an orthomodular lattice structure.

Lemma 2.27 (Proposition 1, page 65 in [KaG]) The Grassmannian $(\mathcal{G}(\mathcal{H}), \subset, \cap, \vee, ^{\perp}, \{0\}, \mathcal{H})$ is a complete orthomodular lattice.

Lemma 2.28 (Theorem 6, page 69 in [KaG]) If $\mathcal{A}$ is a von Neumann algebra on a Hilbert space $\mathcal{H}$ then the Grassmannian $(\mathcal{G}(\mathcal{A}), \subset, \cap, \vee, ^{\perp}, \{0\}, \mathcal{H})$ is a complete sub-orthomodular lattice of $\mathcal{G}(\mathcal{H})$.

The property that $\mathcal{G}(\mathcal{A})$ is a sub-orthomodular lattice of $\mathcal{G}(\mathcal{H})$ means that $\mathcal{G}(\mathcal{A})$ is an orthomodular lattice by itself and that the lattice operations of $\mathcal{G}(\mathcal{A})$ coincide with these of $\mathcal{G}(\mathcal{H})$ obtained by inclusion $\mathcal{G}(\mathcal{A}) \subset \mathcal{G}(\mathcal{H})$. The next statement characterizes the order theoretic qualities of the bijection discussed in Remark 2.25.

Lemma 2.29 (Theorem 2.104, page 112 in [Al]) If $\mathcal{A}$ is a von Neumann algebra on a Hilbert space $\mathcal{H}$ then the projector lattice $\mathcal{P}(\mathcal{A})$ is a complete orthomodular lattice isomorphic to the Grassmannian $\mathcal{G}(\mathcal{A})$ under the mapping $p \mapsto \mathrm{Im}(p)$.

We discuss a few tools from lattice theory.

Definition 2.30 A property of subsets of a set M is a **closure property** when (i) M has the property, and (ii) any intersection of subsets having the given property itself has this property.

Lemma 2.31 (Birkhoff: Corollary in [Bi] on page 7) Those subsets of any set which have a given closure property form a complete lattice, in which the lattice infimum of any family of subsets S_α is their intersection, and their lattice supremum is the intersection of all subsets T_β which contain every S_α.

Lemma 2.32 (Birkhoff: Lemma 1 in [Bi] on page 24) An isotone bijection between two lattices with isotone inverse is a lattice isomorphism.

Definition 2.33 A poset $\mathcal{L}$ is a **chain** if for any two elements $x, y \in \mathcal{L}$ holds $x \leq y$ or $y \leq x$. The **length** of a chain $\mathcal{L}$ is $|\mathcal{L}| - 1$. The **length** of a poset $\mathcal{L}$ is the least upper bound of the lengths of the chains in $\mathcal{L}$.

Remark 2.34 A lattice of finite length is complete. See page 111 in [Bi].

We discuss some connections among the various lattices introduced.

Remark 2.35 **(a)** If the infimum of a subset X in a poset $\mathcal{L}$ exists, then it is unique. Similarly, the supremum of X is unique [Bi].

(b) A subset X of a lattice $\mathcal{L}$ (with the partial ordering induced by $\mathcal{L}$) can be a lattice without being a sublattice of $\mathcal{L}$, that is, infima and suprema can be different in X and in $\mathcal{L}$. An example is explained in Figure 3.4 on page 59.

(c) The ordering of a lattice is recovered from infimum and supremum by

$$x \leq y \iff x \wedge y = x \qquad \text{and} \qquad x \leq y \iff x \vee y = y, \tag{2.42}$$

see Lemma 1 on page 8 in [Bi]. In particular a Boolean lattice $(\mathcal{L}, \leq, \wedge, \vee)$ can be considered as an algebra $(\mathcal{L}, \wedge, \vee)$ with two binary operations and under this aspect it is often called a **Boolean algebra**.

(d) In a Boolean algebra $(\mathcal{L}, \wedge, \vee, 0, 1)$ there exists a unique complement x' to each $x \in \mathcal{L}$, see Theorem 10, Chapter I in [Bi], indeed uniqueness is true for every distributive lattice. As an example, consider Figure 2.1: given a one-dimensional subspace, every distinct one-dimensional subspace is a complement, hence the lattice cannot be distributive. Moreover, in a Boolean algebra holds $(x')' = x$ and

$$(x \wedge y)' = x' \vee y', \qquad (x \vee y)' = x' \wedge y'. \tag{2.43}$$

In particular, every Boolean algebra is an orthomodular lattice.

Conversely, if every element x in a lattice $\mathcal{L}$ has a unique complement x' and if (2.43) holds, then $\mathcal{L}$ is a distributive lattice, see Theorem 17, Chapter II in [Bi].

(e) If $(\mathcal{L}, \wedge, \vee, ')$ is a lattice with an arbitrary involution $x \mapsto x'$ then the equations (2.43) are equivalent to [Bi]

$$x \leq y \quad \Longleftrightarrow \quad x' \geq y'. \tag{2.44}$$

[Proof of (e) on page 191]

(f) Trivially, a distributive lattice is modular, but not conversely, see page 13 in [Bi]. Also, a modular ortholattice is orthomodular, but not conversely: the Grassmannian $\mathcal{G}(\mathcal{H})$ is modular if and only if the Hilbert space $\mathcal{H}$ is finite-dimensional, see Proposition 5 on page 67 in [KaG]. The modularity for finite dimensional $\mathcal{H}$ follows from the fact that the normal subgroups of any group are a modular lattice, see Theorem 11 on page 13 in [Bi], and from the fact that every linear subspace is closed in the finite dimensional case.

We comment on the matrix algebra case. For $N \in \mathbb{N}$ we assume $A = A_1 \oplus \cdots \oplus A_N$ is a direct sum where $n = (n_1, \ldots, n_N) \in \mathbb{N}_0^N$ is a multi-index and for $i = 1, \ldots, N$ we use full matrix algebras $A_i := M_{n_i}$. The algebra A is a von Neumann algebra on the Hilbert space $H := H_1 \oplus \cdots \oplus H_N$ for $H_i := \mathbb{C}^{n_i}$, $i = 1, \ldots, N$.

Remark 2.36 **(a)** If $N \geq 2$ and $n_i \geq 1$ for two indices $i \in \{1, \ldots, N\}$ then the algebra A is smaller than the algebra of linear operators on H and the Grassmannian is smaller than the space of all linear subspaces of H.

(b) The Grassmannian of A is the direct sum $\bigoplus_{i=1}^N \mathcal{G}(A_i)$. For $U_i, V_i \subset H_i$, $i = 1, \ldots, N$ holds $\bigoplus_{i=1}^N U_i \subset \bigoplus_{i=1}^N V_i$ if and only if $U_i \subset V_i$ for $i = 1, \ldots, N$. Therefore the lattice operations in $\mathcal{G}(\mathcal{A})$ are obtained from the individual direct summands

$$(\bigoplus_{i=1}^N U_i) \wedge (\bigoplus_{i=1}^N V_i) = \bigoplus_{i=1}^N (U_i \wedge V_i), \qquad (\bigoplus_{i=1}^N U_i) \vee (\bigoplus_{i=1}^N V_i) = \bigoplus_{i=1}^N (U_i \vee V_i)$$

(Theorem 7 on page 8 in [Bi]). As a consequence the distributive law (2.41) in $\mathcal{G}(A)$ is equivalent with the distributive law in each direct summand $\mathcal{G}(H_i)$, $i = 1, \ldots, N$. The Grassmannian lattice $\mathcal{G}(\mathbb{C}^2)$ has no distributive law, see Figure 2.1. Thus, the Grassmannian $\mathcal{G}(A)$ is a Boolean lattice if and only if each summand of A is at most one-dimensional, that is, A is abelian. The inner product of two vectors $x_1 \oplus \cdots \oplus x_N, y_1 \oplus \cdots \oplus y_N \in H$ is $\sum_{i=1}^N \langle x_i, y_i \rangle$ so the complement of $\bigoplus_{i=1}^N U_i$ in $\mathcal{G}(A)$ is $(\bigoplus_{i=1}^N U_i)^\perp = \bigoplus_{i=1}^N (U_i^\perp)$.

(c) In the classical commutative case $A = \mathbb{C}^N$ the Grassmannian is a Boolean algebra by the previous discussion. There is more additional structure. The **support** of a vector $x \in A = \mathbb{C}^N$ is

$$\operatorname{supp}(x) := \{i \in \{1, \ldots, N\} : x_i \neq 0\}. \tag{2.45}$$

The support is a useful generalization from elements of the probability simplex (1.25). There is a lattice isomorphism between the projector lattice

$$\mathcal{P}(A) = \{a_1 \oplus \cdots \oplus a_N \in \mathbb{C}^N : a_i \in \{0, 1\}, i = 1, \ldots, N\}$$

and the power set $2^{\{1,\ldots,N\}}$ of $\{1, \ldots, N\}$ given by

$$\operatorname{supp}|_{\mathcal{P}(A)} : \quad \mathcal{P}(A) \to 2^{\{1,\ldots,N\}}, \qquad p \mapsto \operatorname{supp}(p).$$

The support is connected to the support projector $p \in \mathcal{P}(A)$ of x (2.20) which has a zero coefficient p_i if and only if x_i is zero for $i = 1, \ldots, N$. One has

$$\operatorname{supp} \circ s(x) = \operatorname{supp}(x).$$

This connection is the reason to prefer the name "support projector" to the name "range projector". These distinct concepts in the theory of C*-algebras [Al] coincide in the finite-dimensional case.

(d) Returning to the general matrix algebra case, the support projector $s(a)$ of a normal matrix $a \in A$ is the infimum [Al]

$$s(a) = \bigwedge \{p \in \mathcal{P}(A) : pa = a\}. \tag{2.46}$$

[Proof on page 191]

(e) The infimum and supremum of two projectors $p, q \in \mathcal{P}(A)$ that commute, $pq = qp$, is

$$p \wedge q = pq \qquad \text{and} \qquad p \vee q = p + q - pq.$$

This is proved using only (2.39) (v). See also Theorem 2.104 in [Al] for the von Neumann algebra case.

We describe the topology of the projector lattice by deduction from the Grassmannian. The partition of the Grassmannian into analytic manifolds is one of the most popular examples in algebraic geometry [Har, Wey] and differential topology [Hi]. We formulate the results at our desire using metric conditions [Av] with C*-norm.

Definition 2.37 **(a)** Two matrices $a, b \in A$ are **conjugate** if there is a unitary $v \in A$ such that

$$b = vav^*. \tag{2.47}$$

The property to be conjugate defines an equivalence relation on A and the relation restricts to an equivalence relation on the projector lattice $\mathcal{P}(A)$. The equivalence classes are the **conjugation classes**.

(b) For two multi-indices $k = (k_1, \ldots, k_N), n = (n_1, \ldots, n_N) \in \mathbb{N}_0^N$ we say that k is **smaller** than n, if $k_i \leq n_i$ for $i = 1, \ldots, N$ and we denote this by $k \leq n$.

(c) For two multi-indices $k, n \in \mathbb{N}_0^N$ with $k \leq n$ and for $A := M_{n_1} \oplus \cdots \oplus M_{n_N}$ we define the **conjugation manifold**

$$\mathcal{P}_k(A) := \mathcal{P}_{k_1}(M_{n_1}) \oplus \cdots \oplus \mathcal{P}_{k_N}(M_{n_N}) \tag{2.48}$$

where $\mathcal{P}_{k_i}(M_{n_i}) := \{p \in \mathcal{P}(M_{n_i}) : \mathrm{rk}(p) = k_i\}$ for $i = 1, \ldots, N$.

Lemma 2.38 If $n \in \mathbb{N}_0^N$ and $A = M_{n_1} \oplus \cdots \oplus M_{n_N}$ then the conjugation classes of the projector lattice $\mathcal{P}(A)$ are the conjugation manifolds $\mathcal{P}_k(A)$ for the multi-indices $k \in \mathbb{N}_0^N$ bounded by $0 \leq k \leq n$. Each $\mathcal{P}_k(A)$ is a compact real analytic manifold of dimension $2\sum_{i=1}^N k_i(n_i - k_i)$. The diameter of the projector lattice is (for $A \neq \{0\}$)

$$\sup_{p,q \in \mathcal{P}(A)} \|p - q\| = 1.$$

Distinct conjugation manifolds are maximally separated, for multi-indices $k \neq l$ with $0 \leq k \leq n$ and $0 \leq l \leq n$ holds $\inf_{p \in \mathcal{P}_k(A), q \in \mathcal{P}_l(A))} \|p - q\| = 1$. [Proof on page 191]

Remark 2.39 For a matrix algebra A, if two orthogonal projectors $p, q \in \mathcal{P}(A)$ satisfy $\|p - q\| < 1$, then they belong to the same conjugation manifold $\mathcal{P}_k(A)$ by Lemma 2.38. Moreover, the unitary v such that $q = vpv^*$ can be chosen as a hermitean unitary. Using the **signum function** $\mathrm{sgn} : \mathbb{R} \to \{-1, 0, 1\}$ which is -1 for negative numbers, 0 at 0 and 1 for positive numbers, a possible choice for v is

$$v := \mathrm{sgn}(\mathbb{1} - p - q) \tag{2.49}$$

defined by functional calculus [Av]. Observe the following asymptotic property of v. If $p \in \mathcal{P}$ and (p_n) is a sequence of orthogonal projectors with limit $p = \lim_{n\to\infty} p_n$ then

$$\lim_{i\to\infty} \mathrm{sgn}(\mathbb{1} - p - p_n) = \mathbb{1} - 2p. \tag{2.50}$$

[Proof of (2.50) on page 192]

3 Convex geometry in Euclidean space

This chapter introduces the terms of convex geometry necessary for our work. In the first two sections most relations are cited by a reference to the monographs [Ro, Sch]. A standard book on the underlying affine geometry is [Kl]. In section three we prove a number of natural relations for normal cones that were not found in the literature. The penultimate section on cylinders is original where we study projections of convex sets. The last section studies two special convex geometric properties of a state space and is substantially original.

3.1 Convex sets

Consider the Euclidean vector space $(\mathbb{R}^m, \langle \cdot, \cdot \rangle)$ and an arbitrary subset $C \subset \mathbb{R}^m$. We recall some working concepts of affine and convex geometry.

Definition 3.1 Let $\mathbb{A} \subset \mathbb{R}^m$ be an affine subspace. The **translation vector space** of $\mathbb{A}$ is defined

$$\operatorname{lin}(\mathbb{A}) := \{x - y : x, y \in \mathbb{A}\} \tag{3.1}$$

provided that $\mathbb{A} \neq \emptyset$. If $\mathbb{A} \neq \emptyset$ then the **orthogonal projection** $\pi_{\mathbb{A}} : \mathbb{R}^m \to \mathbb{A}, x \mapsto \pi_{\mathbb{A}}(x)$ is specified by the relation

$$(x - \pi_{\mathbb{A}}(x)) \perp \operatorname{lin}(\mathbb{A}). \tag{3.2}$$

Recall that $\pi_{\mathbb{A}}$ is an affine map. The finite ($k \in \mathbb{N}$) weighted sum $\lambda_1 x_1 + \cdots + \lambda_k x_k$ of points $x_1, \ldots, x_k \in \mathbb{R}^m$ for real scalars $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ that add up to one, $\lambda_1 + \cdots + \lambda_k = 1$, is an **affine combination** of the points $x_1, \ldots, x_k \in \mathbb{R}^m$. The **affine hull** $\operatorname{aff}(C)$ is the set of all affine combinations of points in C. The **relative interior** of C,

$$\operatorname{ri}(C) \tag{3.3}$$

is the interior of C with respect to the relative topology of the affine hull of C. The **relative boundary** of C is

$$\mathrm{rb}(C) := C \setminus \mathrm{ri}(C). \tag{3.4}$$

The finite ($k \in \mathbb{N}$) weighted sum $\lambda_1 x_1 + \cdots + \lambda_k x_k$ of points $x_1, \ldots, x_k \in \mathbb{R}^m$ for non-negative scalars $\lambda_1, \ldots, \lambda_k \in \mathbb{R}_0^+$ that add up to one, $\lambda_1 + \cdots + \lambda_k = 1$, is a **convex combination** of the points $x_1, \ldots, x_k \in \mathbb{R}^m$. The **convex hull** $\mathrm{conv}(C)$ is the set of all convex combinations of points in C. For $x, y \in \mathbb{R}^m$ we use the short hands

$$\begin{aligned} [x,y] &:= \mathrm{conv}\{x,y\}, \\ ]x,y[&:= \mathrm{ri}(\mathrm{conv}\{x,y\}), \end{aligned} \tag{3.5}$$

where $[x, y]$ is called the **closed segment** and $]x, y[$ the **open segment** with endpoints x, y. The set C is **convex** if

$$[x,y] \subset C \tag{3.6}$$

for all $x, y \in C$. If C is convex and non-empty then the **translation vector space** of C is defined as the translation vector space of the affine hull

$$\mathrm{lin}(C) := \mathrm{lin}(\mathrm{aff}(C)) \tag{3.7}$$

and the **dimension** of C is

$$\dim(C) := \dim(\mathrm{lin}(C)), \tag{3.8}$$

whereas $\dim(\emptyset) := -1$. The **codimension** of C is

$$\mathrm{codim}(C) := m - \dim(\mathrm{lin}(C)). \tag{3.9}$$

For $k \in \mathbb{N}$ a set of $k + 1$ points $b_0, b_1, \ldots, b_k$ is said to be **affinely independent** if $\mathrm{aff}\{b_0, b_1, \ldots, b_k\}$ is k-dimensional. A k-dimensional **simplex** in $\mathbb{R}^m$ is the convex hull of any $k + 1$ affinely independent points in $\mathbb{R}^m$. The set C is a **cone** if it is closed under positive scalar multiplication, that is $\lambda x \in C$ when $x \in C$ and $\lambda > 0$. A **convex cone** is a cone which is a convex set.

Example 3.2 For $k \in \mathbb{N}$ consider the $(k - 1)$-dimensional probability simplex (1.25)

$$\mathbb{P}^k := \{p \in \mathbb{R}^k : p_i \geq 0 \text{ for } i = 1, \ldots, k \text{ and } p_1 + \cdots + p_k = 1\}. \tag{3.10}$$

This is the convex hull of vectors $\delta_i := (0, \ldots, 0, \underset{i}{1}, 0, \ldots, 0)$ corresponding to the **Dirac measures** on $\{1, \ldots, k\}$ for $i = 1, \ldots, k$. The affine hull of $\mathbb{P}^k$ is the $(k - 1)$-dimensional hyperplane $\{x \in \mathbb{R}^k : \sum_{i=1}^k x_i = 1\}$. Thus, the Dirac measures are affinely independent and the probability simplex $\mathbb{P}^k$ is a simplex in the above definition. The relative interior of $\mathbb{P}^k$ consists of the probability distributions $p \in \mathbb{P}^k$ with full support $|\,\mathrm{supp}(p)| = k$.

This is a special case of (4.27) with the matrix algebra $\mathbb{C}^k$. The relative boundary of $\mathbb{P}^k$ is the set of probability distributions that do not dominate the uniform distribution, this is an empty set for $k = 1$.

For graphical applications we seek points $b_i \in \mathbb{R}^{k-1}$, $i = 1, \ldots, k-1$, which together with the origin $b_k := 0 \in \mathbb{R}^{k-1}$ form an isometric copy of the Dirac measures:

$$\|b_i - b_j\|_2 = \sqrt{2} \text{ for } 1 \leq i < j \leq k, \tag{3.11}$$

where $\| \cdot \|_2$ denotes the two-norm on $\mathbb{R}^{k-1}$. Equation (3.11) is solved by the points

$$(b_i)_j = \begin{cases} \frac{1}{\sqrt{j(j+1)}} & \text{for } 1 \leq j < i \\ \sqrt{\frac{i+1}{i}} & \text{for } \quad j = i \\ 0 & \text{for } \quad j > i \end{cases}, \qquad i, j = 1, \ldots, k-1. \tag{3.12}$$

For the tetrahedron ($k = 4$) the vectors in (3.12) are

$$b_1 = \begin{pmatrix} \sqrt{2} \\ 0 \\ 0 \end{pmatrix}, \qquad b_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \sqrt{\frac{3}{2}} \\ 0 \end{pmatrix} \qquad \text{and} \qquad b_3 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{3}} \end{pmatrix}.$$

Remark 3.3 (a) Every affine map $L : \mathbb{R}^m \to \mathbb{R}^k$ commutes with assignment of the affine hull (see page 8 in [Ro]), that is

$$\operatorname{aff}(L(C)) = L(\operatorname{aff}(C)). \tag{3.13}$$

From this it follows that a linear map L commutes with assignment of the translation vector space,

$$\operatorname{lin}(L(C)) = L(\operatorname{lin}(C)). \tag{3.14}$$

If C is convex then an affine map L commutes with the assignment of the relative interior (see Theorem 6.6 in [Ro]),

$$\operatorname{ri}(L(C)) = L(\operatorname{ri}(C)). \tag{3.15}$$

(b) For any two convex subset C_1, C_2 in $\mathbb{R}^m$ we have

$$\operatorname{ri}(C_1 + C_2) = \operatorname{ri}(C_1) + \operatorname{ri}(C_2), \tag{3.16}$$

see Corollary 6.6.2 in [Ro]. Here addition is understood element-wise. If the convex sets C_1 and C_2 share a relative interior point then

$$\operatorname{ri}(C_1 \cap C_2) = \operatorname{ri}(C_1) \cap \operatorname{ri}(C_2) \qquad \text{and} \qquad \overline{C_1 \cap C_2} = \overline{C_1} \cap \overline{C_2} \tag{3.17}$$

by Theorem 6.5 in [Ro]. On page 192 we prove that if the convex sets C_1 and C_2 share a relative interior point then

$$\mathrm{aff}(C_1 \cap C_2) = \mathrm{aff}(C_1) \cap \mathrm{aff}(C_2). \tag{3.18}$$

(c) The convex hull $\mathrm{conv}(C)$ of C is the smallest convex set in $\mathbb{R}^m$ containing C. The set C is convex if and only if $C = \mathrm{conv}(C)$.

Definition 3.4 An **extended real-valued function** on C is a function

$$f : C \to \mathbb{R} \cup \{-\infty, \infty\}.$$

Let f be an extended real-valued function on C. The **epigraph** of f is the set

$$\{(x, \mu) : x \in C, \mu \in \mathbb{R}, \mu \geq f(x)\} \subset \mathbb{R}^{m+1}.$$

The **level set** of f for $\alpha \in \mathbb{R} \cup \{-\infty, \infty\}$ is

$$\{x \in C : f(x) \leq \alpha\}.$$

The function f is **lower semi-continuous** at $x \in C$ if

$$f(x) \leq \liminf_{i \to \infty} f(x_i) \tag{3.19}$$

for every sequence $(x_i)_{i \in \mathbb{N}}$ in C that converges to x. The function f is lower semi-continuous if f is lower semi-continuous at each point $x \in C$. The function f is **upper semi-continuous** if $-f$ is lower semi-continuous. A point $x \in C$ is a **global maximizer** of f if $f(y) \leq f(x)$ for all $y \in C$. A point $x \in C$ is a **local maximizer** of f if there is a neighborhood U of x in C such that $f(y) \leq f(x)$ for all $y \in U$.

Lemma 3.5 ([Ro]) If C is closed then an extended real-valued function f on C is lower semi-continuous if and only if the level set of f is closed in $\mathbb{R}^m$ for each $\alpha \in \mathbb{R}$.

[Proof on page 193]

Definition 3.6 An extended real-valued function f on C is **convex** if the epigraph of f is a convex subset of $\mathbb{R}^{m+1}$. The function f is **concave** if the function $-f$ is convex. Let C be convex. A real-valued function f on C is **strictly convex** if for all distinct points $x, y \in C$ and for all $\lambda \in (0, 1)$ we have

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

The function f is **strictly concave** if $-f$ is strictly convex.

Remark 3.7 If C is convex then $f : C \to (-\infty, +\infty]$ is convex if and only if for all $x, y \in C$ and all $\lambda \in [0,1]$ we have

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

see Theorem 4.1 in [Ro].

3.2 Face lattices

Let C be a convex subset of the Euclidean vector space $(\mathbb{R}^m, \langle \cdot, \cdot \rangle)$. Face and exposed face structures of a convex set are an important convex geometric property.

Definition 3.8 A convex subset $F \subset C$ is a **face** of C if for all $x, y \in C$

$$]x, y[\cap F \neq \emptyset \quad \implies \quad [x, y] \subset F. \tag{3.20}$$

The convex set C itself and $\emptyset$ are faces of C, the **improper** faces. All other faces of C are **proper**. A faces of dimension zero consists of a single point, which is called an **extreme points** of C. The set of faces of C

$$\mathcal{F}(C) \tag{3.21}$$

is called the **face lattice** of C.

Remark 3.9 ([Ro, Bi]) **(a)** For each face F of C we have $F = \operatorname{aff}(F) \cap C$.

(b) If F is a face of C and G is a face of F then G is a face of C.

(c) The intersection of an arbitrary family of faces of C is a face of C.

(d) The face lattice is partially ordered by set inclusion. For subsets of C, the property of being a face is a closure property by (c). Hence by Lemma 2.31 the face lattice

$$(\mathcal{F}(C), \subset, \cap, \vee, \emptyset, C) \tag{3.22}$$

is a complete lattice with smallest element $\emptyset$ and greatest element C. Coatoms of the face lattice are called **facets**. In detail, the closure property of the face lattice implies for a family of faces (F_α) of C the infimum

$$\bigwedge_\alpha F_\alpha = \bigcap_\alpha F_\alpha$$

and the supremum

$$\bigvee_{\alpha} F_\alpha = \bigcap \{F \in \mathcal{F}(C) : F \supset F_\alpha \quad \text{for all} \quad \alpha\}.$$

(e) If F is a face of C and if D is a convex subset of C such that $\mathrm{ri}(D)$ meets F then $D \subset F$.

(f) Every convex set C has the decomposition into the disjoint union

$$C = \dot{\bigcup_{F \in \mathcal{F}(C)}} \mathrm{ri}(F) \tag{3.23}$$

of relative interiors of faces. This decomposition is called **stratification**. In particular, the dimension of a proper face F of C is strictly smaller than the dimension of C. [Proof on page 193]

Definition 3.10 The **face of a point** $x \in C$ is defined as the unique face $F \in \mathcal{F}(C)$ with $x \in \mathrm{ri}(F)$, and is denoted by

$$F(C, x). \tag{3.24}$$

Remark 3.11 ([Ro]) **(a)** For any $x \in C$ and any face G of C the inclusion $F(C, x) \subset G$ is equivalent to $x \in G$.

(b) Let two convex sets $C_1, C_2 \subset \mathbb{R}^m$ be given. For every face F of C_1 the set $F \cap C_2$ is a face of $C_1 \cap C_2$.

Moreover, if C_2 is an affine space then the face structure of $C_1 \cap C_2$ is obtained by intersecting the faces of C_1 with C_2: for all $x \in C_1 \cap C_2$ we have

$$F(C_1 \cap C_2, x) = F(C_1, x) \cap C_2. \tag{3.25}$$

See Figure 3.1 for an example where the equation is wrong under weaker conditions.

(c) The face structure of a face G of C is obtained by restriction. One has $F(G, x) = F(C, x)$ for all $x \in G$.

(d) The **Minkowski theorem** says that a compact convex set $C \subset \mathbb{R}^m$ is the convex hull of its extreme points,

$$C = \mathrm{conv}(\{x \in C : \{x\} \text{ is a face of } C\}). \tag{3.26}$$

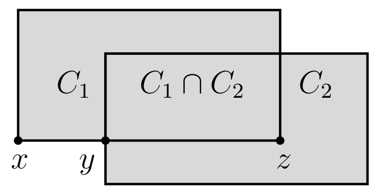


Figure 3.1: The rectangles C_1 and C_2 have $F(C_1, y) \cap C_2 = [x, z] \cap C_2 = [y, z]$ and $F(C_1 \cap C_2, y) = \{y\}$. The equation $F(C_1, y) \cap C_2 = F(C_1 \cap C_2, y)$ in (3.25) is violated.

This is a special case of the famous Krein-Milman theorem [Wer]. [Proof on page 193]

Definition 3.12 The **dimension function** of C assigns to a point $x \in C$ the dimension of the face $F(C, x)$:

$$x \mapsto \dim(F(C, x)). \tag{3.27}$$

The level set of the dimension function for $0 \leq d \leq m$ is the d-**skeleton** of C

$$\operatorname{skel}(C, d) := \{x \in C : \dim(F(C, x)) \leq d\}. \tag{3.28}$$

Remark 3.13 **(a)** If C is closed then the dimension function of C is lower semi-continuous if and only if all d-skeletons of C are closed. This follows from Lemma 3.5 because the skeletons are the level sets of the dimension function.

(b) As an example where the 0-skeleton is not closed consider the skew double cone $C \subset \mathbb{R}^3$ defined as the convex hull of the circle $c := \{(x, y, 0) : (x-1)^2 + y^2 = 1\}$ and the two points $(0, 0, \pm 1)$. Points on the circle c are extreme points of C except for the origin 0, which lies in the relative interior of the one-dimensional face $[(0, 0, -1), (0, 0, 1)]$. See Figure 5.8 for a convex set affinely isomorphic to C.

Some faces of C are obtained by intersection of C with a hyperplane. This makes them special and easier to study.

Definition 3.14 The **barrier cone** of C is

$$B(C) := \{u \in \mathbb{R}^m : \sup_{x \in C} \langle u, x \rangle < \infty\}. \tag{3.29}$$

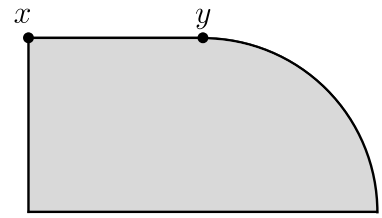


Figure 3.2: The extreme point x is an exposed point. The extreme point y is not an exposed point because every supporting hyperplane through y contains the face $\{x\} \vee \{y\} = [x,y]$. The face $\{y\}$ has the same normal cone as $[x,y]$ although $[x,y]$ is strictly larger than $\{y\}$.

The **support function** of C is defined as $h(C) : \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\}$,

$$h(C,u) := \sup_{x \in C} \langle u, x \rangle. \tag{3.30}$$

For a non-zero vector $u \in B(C) \setminus \{0\}$ the **supporting half-space** of C for u is

$$H^-(C,u) := \{x \in \mathbb{R}^m : \langle u, x \rangle \leq h(C,u)\}.$$

The **supporting hyperplane** of C for u is

$$H(C,u) := \{x \in \mathbb{R}^m : \langle u, x \rangle = h(C,u)\}. \tag{3.31}$$

The **exposed face** of C for u is

$$F_\perp(C,u) := H(C,u) \cap C. \tag{3.32}$$

For completeness we define for a vector $u \in \mathbb{R}^m \setminus B(C)$ the exposed face $F_\perp(C,u) := \emptyset$. The improper faces $\emptyset$ and C are exposed faces of C by definition. For a non-zero vector $u \in \mathbb{R}^m$ we say that $F_\perp(C,u)$ is the **face exposed by** u. An exposed extreme point is called an **exposed point**. The **exposed face lattice** of C is

$$\mathcal{F}_\perp(C) := \{F \in \mathcal{F}(C) : F \text{ is exposed}\}. \tag{3.33}$$

Remark 3.15 **(a)** If C is bounded then $B(C) = \mathbb{R}^m$ and the cylinder $C + V$ has barrier cone $B(C+V) = V^\perp$ for $V \subset \mathbb{R}^m$ a linear space. The barrier cone is not closed in general: if $C = \{(x,y) \in \mathbb{R}^2 : y \leq -x^2\}$ then $B(C) = \{(x,y) \in \mathbb{R}^2 : y > 0\} \cup \{0\}$.

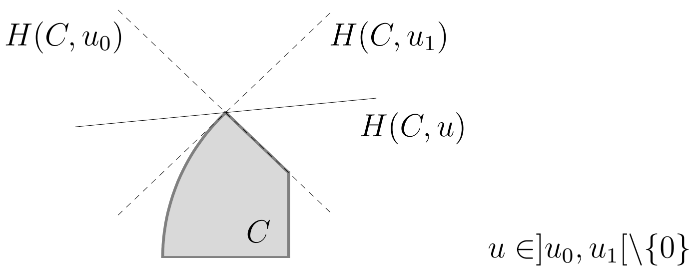


Figure 3.3: The supporting hyperplane $H(C, u)$ of the depicted convex set C meets the intersection of the two faces exposed by the hyperplanes $H(C, u_0)$ and $H(C, u_1)$.

(b) The support function $h(C) : \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\}$ is convex, see Theorem 13.2 in [Ro]. If C is non-empty and bounded then $h(C)$ is finite and continuous throughout $\mathbb{R}^m$ because a convex function is continuous on the relative interior of the domain of finiteness, see Theorem 10.1 in [Ro].

(c) If C is non-empty and closed then for every non-zero vector $u \in B(C)$ the exposed face $F_\perp(C, u)$ is non-empty [consider the orthogonal projection of C to $\mathrm{Lin}(u)$].

(d) The face exposed by a non-zero vector $u \in \mathbb{R}^m$ is the set of maximizers in C of the linear functional $\langle u, \cdot \rangle$,

$$F_\perp(C, u) = \{x \in C : \langle u, x \rangle = h(C, u)\}. \tag{3.34}$$

(e) Every exposed face of C is a face of C.

(f) The relative boundary of C is the union of proper exposed faces of C, see Theorem 11.6 in [Ro].

(g) The difference between an exposed face an a non-exposed face is explained in Figure 3.2.

Lemma 3.16 Any intersection of exposed faces of C is an exposed face of C. In detail, for non-empty $U \subset \mathbb{R}^m \setminus \{0\}$ we have $\mathrm{ri}(\mathrm{conv}(U)) \setminus \{0\} \neq \emptyset$ and for any $v \in \mathrm{ri}(\mathrm{conv}(U)) \setminus \{0\}$ we have $\bigcap_{u \in U} F_\perp(C, u) = F_\perp(C, v)$ unless the intersection is empty. [Proof on page 194]

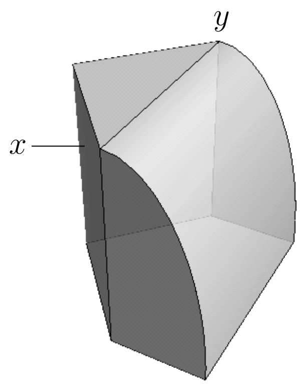


Figure 3.4: The points x, y are exposed points in the depicted closed and convex set. The supremum $x \vee y$ in the lattice of all subsets, the face lattice resp. the exposed face lattice is the two point set $\{x, y\}$, the segment $[x, y]$ resp. the top triangle. The lattices are not sublattices of each other.

Apart from scaling by positive real numbers, the set $\mathrm{ri}(\mathrm{conv}(U)) \setminus \{0\}$ to choose a single exposing vector is maximal in Lemma 3.16. As an example we can consider the relative boundary vector $u_0 \in [u_0, u_1]$ in Figure 3.3. The face exposed by u_0 is larger than the intersection $F_\perp(C, u_0) \cap F_\perp(C, u_1)$.

Remark 3.17 **(a)** The exposed face lattice is partially ordered by inclusion of sets. Since an intersection of exposed faces is an exposed face (Lemma 3.16) the property to be an exposed face of C is a closure property. Hence by Lemma 2.31

$$(\mathcal{F}_\perp(C), \subset, \cap, \vee, \emptyset, C) \tag{3.35}$$

is a complete lattice with smallest element $\emptyset$ and greatest element C. From the closure property follows for a family of exposed faces (F_α) of C the infimum $\bigwedge_\alpha F_\alpha = \bigcap_\alpha F_\alpha$ and the supremum $\bigvee_\alpha F_\alpha = \bigcap\{F \in \mathcal{F}_\perp(C) : F \supset F_\alpha \quad \text{for all} \quad \alpha\}$.

(b) The exposed face lattice is not a sublattice of the face lattice (3.22) in general. The supremum taken in the exposed face lattice may be larger than the supremum taken in the face lattice, see Figure 3.4.

3.3 The normal cone lattice

Let C be a convex subset of the Euclidean vector space $(\mathbb{R}^m, \langle\cdot,\cdot\rangle)$. We study normal cones of C, the dual concept of exposed faces. This section was created at the beginning of the dissertation project with the intention to prove that a state reflection has only exposed faces—which is disproved by the simple counter example in Figure 5.4 on page 104. Existence of non-exposed faces in this context are not documented in the literature as far as we know. Thus, the section contains many original statements with their proofs. Luckily the work was not completely in vain, we prove Theorem 1 in the last section of this chapter.

Definition 3.18 Let x be a point of C. The **normal cone** of C at x is the set of vectors $u \in \mathbb{R}^m$ that do not make an acute angle with any line segment in C with endpoint x,

$$\mathrm{N}(C,x) := \{u \in \mathbb{R}^m : \langle u, y-x\rangle \leq 0 \text{ for all } y \in C\}. \tag{3.36}$$

A vector $u \in \mathrm{N}(C,x)$ is called a **normal vector** of C at x.

Remark 3.19 **(a)** The normal cone of C at x is a closed convex cone and $\mathrm{N}(\overline{C},x) = \mathrm{N}(C,x)$.

(b) The normal cone of C at x is included in the barrier cone (3.29),

$$\mathrm{N}(C,x) \subset B(C) = \{u \in \mathbb{R}^m : h(C,u) < \infty\}. \tag{3.37}$$

(c) The normal cone of C at x consists of those vectors $u \in \mathbb{R}^m$ where x maximizes the linear functional $\langle u,\cdot\rangle$ on C,

$$\mathrm{N}(C,x) = \{u \in \mathbb{R}^m : \langle u, x\rangle = h(C,u)\}. \tag{3.38}$$

(d) The following duality is fundamental. For non-zero vectors $u \in \mathbb{R}^m$ and $x \in C$ we have

$$x \in F_\perp(C,u) \quad\Longleftrightarrow\quad \langle u,x\rangle = h(C,u) \quad\Longleftrightarrow\quad u \in \mathrm{N}(C,x). \tag{3.39}$$

This follows from the maximization principles for exposed faces (3.34) and for normal cones (3.38).

(e) Let $C_1, C_2 \subset \mathbb{R}^m$ be convex sets and $x \in C_1, y \in C_2$. It is elementary to verify that

$$\mathrm{N}(C_1 + C_2, x+y) = \mathrm{N}(C_1,x) \cap \mathrm{N}(C_2,y). \tag{3.40}$$

(f) Let two convex sets $C_1, C_2 \subset \mathbb{R}^m$ share a relative interior point and let $x \in C_1 \cap C_2$. We prove on page 194 ([Ro, Sch]) that

$$\mathrm{N}(C_1 \cap C_2, x) = \mathrm{N}(C_1, x) + \mathrm{N}(C_2, x). \tag{3.41}$$

The premise with the common relative interior point can not be left out. Consider the case of two closed disks in $\mathbb{R}^2$ that share a boundary point but no interior point.

(g) The following relations are easy to verify, let $x, y \in C$.

$$\begin{array}{lrcl}
\text{(i)} & & \mathrm{N}(C,x) \perp \mathrm{lin}(F(C,x)), & \\
\text{(ii)} & \text{if } y \in F(C,x) & \text{then} & \mathrm{N}(C,y) \supset \mathrm{N}(C,x), \\
\text{(iii)} & \text{if } y \in \mathrm{ri}(F(C,x)) & \text{then} & \mathrm{N}(C,y) = \mathrm{N}(C,x), \\
\text{(iv)} & \text{if } u, -u \in \mathrm{N}(C,x) & \text{then} & u \in \mathrm{lin}(C)^\perp.
\end{array}$$

Lemma 3.20 Let $x \in C$. Then $\mathrm{lin}(C)^\perp \subset \mathrm{N}(C,x) \subset B(C)$. Equivalent are

(a) the normal cone $\mathrm{N}(C,x)$ is a vector space,
(b) $x \in \mathrm{ri}(C)$,
(c) $\mathrm{N}(C,x) = \mathrm{lin}(C)^\perp$.

The equality $\mathrm{N}(C,x) = B(C)$ implies that the closure of C is the translate of a convex cone. [Proof on page 194]

Definition 3.21 The **normal cone of a non-empty face** F of C is defined as

$$\mathrm{N}(C,F) := \mathrm{N}(C,x) \tag{3.42}$$

for any $x \in \mathrm{ri}(F)$. This definition is consistent by Remark 3.19 (g) (iii). The **normal cone of the empty set** is defined as the ambient space

$$\mathrm{N}(C,\emptyset) := \mathbb{R}^m.$$

The **normal cone lattice** of a convex set $C \subset \mathbb{R}^m$ is

$$\mathcal{N}(C) := \{\mathrm{N}(C,F) : F \in \mathcal{F}(C)\}. \tag{3.43}$$

The normal cone lattice is partially ordered by set inclusion.

Remark 3.22 **(a)** $\mathrm{N}(\emptyset, \emptyset) = \mathbb{R}^m$ so $\mathcal{N}(\emptyset) = \{\mathbb{R}^m\}$.

(b) The assignment of the normal cone to a face, $\mathrm{N}(C) : \mathcal{F}(C) \to \mathcal{N}(C)$ is antitone. As a consequence of Remark 3.19 (g) (ii) we have for all faces $F, G \in \mathcal{F}(C)$

$$F \subset G \quad \Longrightarrow \quad \mathrm{N}(C, G) \subset \mathrm{N}(C, F). \tag{3.44}$$

(c) The converse to (3.44) is wrong. The faces $\{x\}$ and $\{y\}$ in Figure 3.2 on page 57 are unrelated but $\mathrm{N}(C, \{y\}) \subset \mathrm{N}(C, \{x\})$. In Proposition 3.29 we show that the converse to (3.44) is true for exposed faces.

(d) On page 195 we prove the following facial variant of the duality (3.39). If $F \in \mathcal{F}(C)$ and $u \in \mathbb{R}^m \setminus \{0\}$ then

$$F \subset F_\perp(C, u) \quad \Longleftrightarrow \quad u \in \mathrm{N}(C, F). \tag{3.45}$$

In particular, one has $u \in \mathrm{N}(C, F_\perp(C, u))$.

Proposition 3.23 For faces F, G of C we have $\mathrm{N}(C, F \vee G) = \mathrm{N}(C, F) \cap \mathrm{N}(C, G)$. The cone $\mathrm{N}(C, F \vee G)$ is a face of $\mathrm{N}(C, F)$ and of $\mathrm{N}(C, G)$. [Proof on page 195]

Corollary 3.24 The normal cone lattice $\mathcal{N}(C)$ is a complete lattice with smallest element $\mathrm{lin}(C)^\perp$ and greatest element $\mathbb{R}^m$. The infimum of two cones $K, L \in \mathcal{N}(C)$ is $K \wedge L = K \cap L$. [Proof on page 196]

Definition 3.25 The **smallest exposed face** of C that contains a face $F \in \mathcal{F}(C)$ is denoted by

$$\overset{\perp}{F} := \bigcap \{G \in \mathcal{F}_\perp(C) : F \subset G\}. \tag{3.46}$$

This definition is consistent by completeness of the exposed face lattice $\mathcal{F}_\perp(C)$, see (3.35). The dependence on C will be indicated if confusion can arise.

Lemma 3.26 If $F \in \mathcal{F}(C)$ is a non-empty face with non-zero normal cone then the smallest exposed face of C containing F is

$$\overset{\perp}{F} = \bigcap_{u \in \mathrm{N}(C,F) \setminus \{0\}} F_\perp(C, u). \tag{3.47}$$

In particular, if F is exposed then $F = \bigcap_{u \in \mathrm{N}(C,F)\setminus\{0\}} F_\perp(C,u)$. If $F \in \mathcal{F}(C)$ is a face then

$$\mathrm{N}(C, \overset{\perp}{F}) = \mathrm{N}(C, F). \tag{3.48}$$

[Proof on page 196]

Lemma 3.27 If $F \neq \emptyset$ is an exposed face of C and $u \in \mathrm{ri}(\mathrm{N}(C,F)) \setminus \{0\}$ then $F = F_\perp(C,u)$. [Proof on page 196]

We addressed some effort to the study of exposed faces producing the following results.

Proposition 3.28 Let F and G denote proper faces of C. Then

$$\begin{array}{llll}
\text{(a)} & \forall F,G: \; F \subset G & \Longrightarrow & \mathrm{N}(G) \subset \mathrm{N}(F), \\
\text{(b)} & \forall G: \; G = \overset{\perp}{G} & \Longleftrightarrow & \left(\forall F: \; \mathrm{N}(G) \subset \mathrm{N}(F) \implies F \subset G\right), \\
\text{(c)} & \forall F: \; F = \overset{\perp}{F} & \Longleftrightarrow & \left(\forall G: \; F \subsetneq G \implies \mathrm{N}(G) \subsetneq \mathrm{N}(F)\right), \\
\text{(d)} & \forall F,G: \; G = \overset{\perp}{G} & \Longrightarrow & \left(\mathrm{N}(G) \subsetneq \mathrm{N}(F) \implies F \subsetneq G\right)
\end{array}$$

with abbreviations $\mathrm{N}(F) := \mathrm{N}(C,F)$ and $\mathrm{N}(G) := \mathrm{N}(C,G)$. [Proof on page 197]

The condition (d) in Proposition 3.28 has no converse. In Figure 3.8 on page 71, left drawing, the normal cones of all proper faces are one-dimensional rays so the condition is void but the four corners are non-exposed faces.

Proposition 3.29 Assume that C has not exactly one point. Then the assignment of normal cones to exposed faces $\mathrm{N}(C) : \mathcal{F}_\perp(C) \to \mathcal{N}(C)$, $F \mapsto \mathrm{N}(C,F)$ is an antitone lattice isomorphism. [Proof on page 197]

3.4 Cylinders on a convex set

This section contains a lifting construction to study projections of convex sets. All proofs are original, no equivalent results in the literature are known to us.

Throughout this section let C be a convex subset of the Euclidean vector space $(\mathbb{R}^m, \langle\cdot,\cdot\rangle)$ and let V be a linear subspace of $\mathbb{R}^m$. We study the orthogonal projection of C onto V. Here we present face lattice isomorphisms and we calculate normal cones. The orthogonal projection to V (3.2)

$$\pi_V : \mathbb{R}^m \to V,$$

thought of as acting on sets, establishes the identity

$$\pi_V(M) = (M + V^\perp) \cap V \tag{3.49}$$

for arbitrary subsets $M \subset \mathbb{R}^m$. In addition to the projection $\pi_V(C)$ we will study the cylinder $C + V^\perp$, which connects the convex set C and the projection $\pi_V(C)$.

Remark 3.30 **(a)** When applied to the Grassmannian (2.34) of a finite-dimensional C*-algebra, formula (3.49) describes the projection of a subspace to another subspace in pure lattice terms.

(b) In our application, the convex set C will be the state space of a finite-dimensional C*-algebra and V will be the tangent space of an exponential family.

We wish to make more transparent a basic tool for the study of cylinders.

Lemma 3.31 Let $X, Y, Z \subset \mathbb{R}^m$ such that $Z \pm X \subset Z$. Then $(X+Y) \cap Z = X + (Y \cap Z)$.
[Proof on page 198]

Definition 3.32 We want to emphasize two special cases of Lemma 3.31. For a vector space $U \subset \mathbb{R}^m$ and arbitrary subsets $X, Y, M \subset \mathbb{R}^m$ with $X \subset U$ we have

$$(X + Y) \cap (M + U) = X + (Y \cap (M + U)).$$

This equation is the **modular law for cylinders**. In the special case $|M| = 1$ we obtain the **modular law for affine spaces**. For an affine space $\mathbb{A} \subset \mathbb{R}^m$ with translation vector space $\mathrm{lin}(\mathbb{A})$ and under the assumption $X \subset \mathrm{lin}(\mathbb{A})$ we have

$$(X + Y) \cap \mathbb{A} = X + (Y \cap \mathbb{A}).$$

Definition 3.33 With respect to C and V we define the **lift** as the mapping

$$L_V^C : \quad 2^{\mathbb{R}^m} \to 2^C, \quad M \mapsto (M + V^\perp) \cap C \tag{3.50}$$

where $2^{\mathbb{R}_m}$ denotes the power set of $\mathbb{R}^m$ and 2^C the power set of C.

Lemma 3.34 The projection π_V is isotone with respect to set inclusion. One has

$$L_V^C = L_V^C \circ \pi_V = L_V^C \circ L_V^C$$

on the power set of $\mathbb{R}^m$. If $\mathcal{M}$ is a family of subsets of $\pi_V(C)$ then π_V is left inverse to $L_V^C|_{\mathcal{M}}$. In particular

$$L_V^C|_{\mathcal{M}} : \mathcal{M} \to \{L_V^C(M) : M \in \mathcal{M}\}$$

is a bijection with inverse π_V and $L_V^C|_{\mathcal{M}}$ is isotone with respect to set inclusion. The proofs are elementary.

Remark 3.35 Let $C \neq \emptyset$. **(a)** The support functions of a cylinder and its projection compare as follows. For $v \in \mathbb{R}^m$ we have

$$h(C + V^\perp, v) = h(\pi_V(C) + V^\perp, v) = \begin{cases} h(C, v) = h(\pi_V(C), v) & \text{if } v \in V, \\ \infty & \text{otherwise.} \end{cases}$$

(b) If C is bounded then the barrier cone of the cylinder $C + V^\perp$ is $B(C + V^\perp) = V$. The barrier cone of the projection $\pi_V(C)$ is $\mathbb{R}^m$.

(c) Still, if C is unbounded then $B(C + V^\perp) \subset V$. Indeed, $B(C + V^\perp) = B(C) \cap V$ and $B(\pi_V(C)) = (B(C) \cap V) + V^\perp$. However it may happen that $B(\pi_V(C)) \not\supset B(C)$, see Figure 3.5.

Lemma 3.36 (Lifted faces) If F is a face of $\pi_V(C)$ then the lift $L_V^C(F)$ is a face of C. For non-zero $v \in V$ we have $L_V^C(F_\perp(\pi_V(C), v)) = F_\perp(C, v)$. [Proof on page 198]

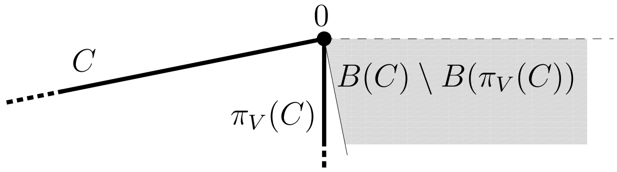


Figure 3.5: The closed ray C projects to a distinct ray $\pi_V(C)$ both emanating from the origin 0. The barrier cones are closed half spaces and they differ according to the angle between the rays.

Definition 3.37 With respect to C and V, the face $L_V^C(F) \in \mathcal{F}(C)$ is called the **lifted face** of $F \in \mathcal{F}(\pi_V(C))$. The **lifted face lattice** is

$$\mathcal{F}_V^C := \{L_V^C(F) : F \in \mathcal{F}(\pi_V(C))\}.$$

The **lifted exposed face lattice** is

$$\mathcal{F}_{V,\perp}^C := \{L_V^C(F) : F \in \mathcal{F}_\perp(\pi_V(C))\} \tag{3.51}$$

where $\mathcal{F}(\pi_V(C))$ is the face lattice of $\pi_V(C)$ and $\mathcal{F}_\perp(\pi_V(C))$ is the exposed face lattice of $\pi_V(C)$. We consider $\mathcal{F}_V^C$ and $\mathcal{F}_{V,\perp}^C$ partially ordered by set inclusion.

Proposition 3.38 (Lifted face lattices) The restricted lifts

$$\begin{array}{rrcl} L_V^C|_{\mathcal{F}(\pi_V(C))}: & \mathcal{F}(\pi_V(C)) & \to & \mathcal{F}_V^C \subset \mathcal{F}(C), \\ L_V^C|_{\mathcal{F}_\perp(\pi_V(C))}: & \mathcal{F}_\perp(\pi_V(C)) & \to & \mathcal{F}_{V,\perp}^C \subset \mathcal{F}_\perp(C) \end{array}$$

are lattice isomorphisms with inverse π_V. One has $L_V^C(F_\perp(\pi_V(C), v)) = F_\perp(C, v)$ and $\pi_V(F_\perp(C, v)) = F_\perp(\pi_V(C), v)$ for non-zero $v \in V$. The infimum in the lifted face lattices is given by intersection. [Proof on page 198]

The lifted face lattice and the lifted exposed face lattice need not be sublattices of $\mathcal{F}(C)$. Consider an equilateral tetrahedron C as a convex subset in $\mathbb{R}^3$ with V the translation vector space of a triangular face F of C. Then two different edges s_1, s_2 of F have

$$L_V^C(\pi_V(s_1) \vee \pi_V(s_2)) = C \supsetneq F = L_V^C(\pi_V(s_1)) \vee L_V^C(\pi_V(s_2)).$$

The supremum on the left-hand side is taken in the face lattice of $\pi_V(C)$ and on the right-hand side in the face lattice of C.

Proposition 3.39 (Lift invariance) A face $F \in \mathcal{F}(C)$ belongs to the lifted face lattice $\mathcal{F}_V^C$ if and only if $L_V^C(F) = F$. [Proof on page 199]

Lemma 3.40 (Relative boundary) The following statements for a convex subset $F \subset C$ are equivalent:

(a) F projects to the relative boundary $\mathrm{rb}(\pi_V(C))$ under π_V,
(b) $F \subset G$ for some proper face $G \in \mathcal{F}_{V,\perp}^C$,
(c) $F \subset G$ for some proper face $G \in \mathcal{F}_V^C$.

[Proof on page 200]

Lemma 3.41 (Normal cones) Let $a \in C + V^\perp$. Then $\mathrm{N}(\pi_V(C), \pi_V(a)) = \mathrm{N}(C + V^\perp, a) + V^\perp$. If a belongs to C then $\mathrm{N}(C + V^\perp, a) = \mathrm{N}(C, a) \cap V$. [Proof on page 200]

3.5 Acute relations

Let C be a convex subset of the Euclidean vector space $(\mathbb{R}^m, \langle \cdot, \cdot \rangle)$. Below we study modified variants of the duality (3.39):

$$x \in F_\perp(C, u) \quad \Longleftrightarrow \quad u \in \mathrm{N}(C, x)$$

for $x \in C$ and $u \in \mathbb{R}^m \setminus \{0\}$. The variants are true for the state spaces of a finite-dimensional C*-algebra. The relations are sharper than the duality above. One is related to normal cones, the other to exposed faces.

Definition 3.42 A vector $u \in \mathbb{R}^m \setminus \{0\}$ is **acute normal** for C if

$$x \in \mathrm{ri}(F_\perp(C, u)) \quad \Longrightarrow \quad u \in \mathrm{ri}(\mathrm{N}(C, x)). \tag{3.52}$$

A point $x \in C$ is **acute exposed** for C if

$$u \in \mathrm{ri}(\mathrm{N}(C, x)) \setminus \{0\} \quad \Longrightarrow \quad x \in \mathrm{ri}(F_\perp(C, u)). \tag{3.53}$$

At first we study acute normal vectors. Related to acute normal vectors only Lemma 3.43

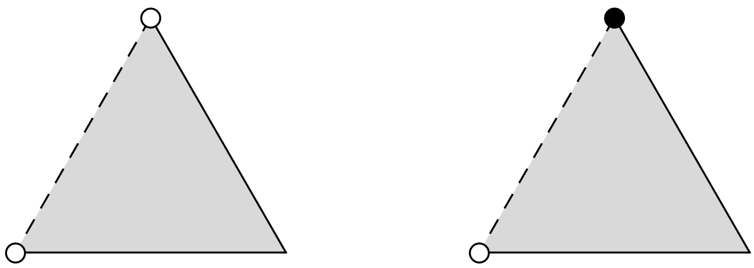


Figure 3.6: Empty circles denote deleted points, dashed lines denote deleted lines. All non-zero vectors are acute normal for the convex set on the left-hand side. One can check Theorem 1, an exposed face is an intersection of coatoms. The convex set on the right-hand side has an exposed face (the top vertex) which is not an intersection of coatoms. Thus there is a non-zero vector which is not acute normal. Which one?

and Theorem 1 will be applied in this thesis. They are used to detect non-exposed faces of a state reflection in Example 5.7 (e) on page 105.

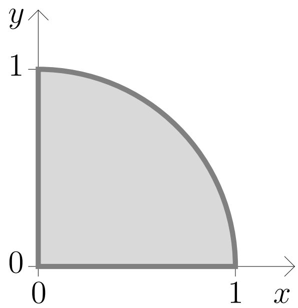
Lemma 3.43 If $C \neq \emptyset$ and if the (non-zero) vector $u \in \text{lin}(C)$ is acute normal for C, then u is acute normal for the restriction of the ambient space $\mathbb{R}^m$ to $\text{aff}(C)$. Assume $V \subset \mathbb{R}^m$ is a linear subspace. If the (non-zero) vector $v \in V$ is acute normal for C then v is acute normal for $\pi_V(C)$. [Proof on page 200]

Theorem 1 If every vector $u \in \mathbb{R}^m \setminus \{0\}$ is acute normal for C then a proper face F of C is exposed if and only if F is an intersection of coatoms of the face lattice $\mathcal{F}(C)$.
[Proof on page 201]

Remark 3.44 (a) In dimension $\dim(C) = 2$ Theorem 1 assumes an easy statement. If the premise of the theorem is true then a face F of C is non-exposed if and only if $F = \{x\}$ where x is the endpoint of some one-dimensional face of C but x is not the endpoint of two distinct one-dimensional faces of C.

(b) The characterizations of exposed faces in (a) and in Theorem 1 does not require that C is closed, see Figure 3.6.

In the following we prove that acute normal vectors have information about the normal cone lattice of a convex set and they are linked to the concept of a touching cone defined



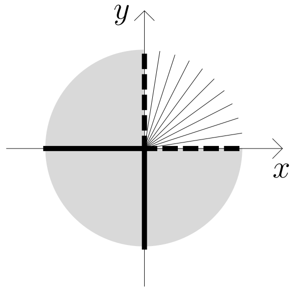


Figure 3.7: The normal cones of the quarter disk (left) are sketched on the unit ball (right). The relative interiors of normal cones do not cover the two dashed rays.

in [Sch]. We want to motivate the definition of a touching cone by the observation that the relative interiors of normal cones $\{\mathrm{ri}(\mathrm{N}(C,F)) : F \in \mathcal{F}_\perp(C) \setminus \{\emptyset\}\}$ do not cover the barrier cone $B(C) \setminus \{0\}$ in general. An example is given in Figure 3.7.

Definition 3.45 (Schneider: page 74 in [Sch]) The **touching cone** of C for a non-zero vector $u \in B(C)$ is defined as

$$\mathrm{T}(C,u) := F(\mathrm{N}(C, F_\perp(C,u)), u) \tag{3.54}$$

provided that $F_\perp(C,u) \neq \emptyset$. This is the face of the normal cone $\mathrm{N}(C, F_\perp(C,u))$ that has u in the relative interior.

Lemma 3.46 Non-zero normal cones of non-empty faces of C are touching cones of C. If K is a touching cone of C then

(a) K is a closed convex cone included in $B(C)$,

(b) $F_\perp(C,u) = F_\perp(C,v)$ for all $u,v \in \mathrm{ri}(K) \setminus \{0\}$,

(c) $K = T(C,u)$ for all $u \in \mathrm{ri}(K) \setminus \{0\}$,

(d) if $0 \in \mathrm{ri}(K)$ then $K = \mathrm{lin}(C)^\perp$.

[Proof on page 202]

Proposition 3.47 Two distinct touching cones of C do not meet in their relative interior. If C is closed then $B(C) \setminus \{0\}$ is covered by relative interiors of touching cones. [Proof on page 202]

Proposition 3.48 A touching cone K of C is the normal cone of a non-empty face of C if and only if there is an acute normal vector in $\mathrm{ri}(K) \setminus \{0\}$. Likewise all vectors in $\mathrm{ri}(K) \setminus \{0\}$ are acute normal for C. [Proof on page 203]

Corollary 3.49 If C is closed then a vector $u \in B(C) \setminus \{0\}$ belongs to the relative interior of the normal cone of a non-empty face of C if and only if u is acute normal for C. [Proof on page 203]

In the following two lemmas we consider acute exposed points.

Lemma 3.50 If $x \in C \cap \mathbb{A}$ for an affine space $\mathbb{A} \subset \mathbb{R}^m$ and if the face $F(C, x)$ is exposed then $F(C \cap \mathbb{A}, x)$ is an exposed face of $C \cap \mathbb{A}$. [Proof on page 203]

Lemma 3.51 A non-empty face F of C is exposed if and only if there is an acute exposed point in $\mathrm{ri}(F)$. Likewise all points in $\mathrm{ri}(F)$ are acute exposed for C. [Proof on page 203]

There is a connection between acute normal vectors and acute exposed points through polarity of convex sets. We do not prove this here and conclude with an example. The stadium in Figure 3.8 has four non-exposed faces. Its polar convex set has four touching cones, which are not normal cones. The polar is computed in Example 3.54.

Definition 3.52 Let $C \subset \mathbb{R}^m$ be a convex set. The **polar** of C is

$$C^\circ := \{x \in \mathbb{R}^m : \langle x, y \rangle \leq 1 \text{ for all } y \in C\}. \tag{3.55}$$

Remark 3.53 • If C is closed and contains the origin 0 then $C^{\circ\circ} = C$, see Theorem 14.5 in [Ro].

- If C is closed then the polar C° is bounded if and only if 0 lies in the interior of C, see Corollary 14.5.1 in [Ro].
- Let C be a compact convex set and $0 \in \mathrm{int}(C)$. The **radial function** of C is

$$\rho(C, x) := \max\{\lambda \geq 0 : \lambda x \in C\}$$

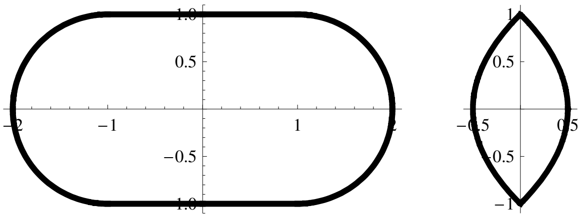


Figure 3.8: The stadium and its polar (for $L = 1$)

with $x \in \mathbb{R}^m \setminus \{0\}$. Using the relation $\rho(C, \lambda x) = \frac{1}{\lambda}\rho(C, x)$ with $x \in \mathbb{R}^m \setminus \{0\}$ and $\lambda > 0$, it follows

$$C = \{x \in \mathbb{R}^m \setminus \{0\} : \|x\| \leq \rho(C, \frac{x}{\|x\|})\} \cup \{0\}. \tag{3.56}$$

This is useful for finding the polar of C because the radial function of the polar is connected to the support function

$$\rho(C^\circ, x) = \frac{1}{h(C, x)} \quad \text{for} \quad \|x\| = 1, \tag{3.57}$$

see Remark 1.7.7 in [Sch].

Example 3.54 We calculate the polar body for the stadium C defined as the union of the rectangle $[-L, L] \times [-1, 1]$ for $L \geq 0$ with two semi-circles of radius one attached on either vertical sides, see Figure 3.8, left drawing. As a coordinate for normal vectors of C we parametrize the unit circle in $\mathbb{R}^2$ by $u(\phi) = \left(\begin{smallmatrix} \cos(\phi) \\ \sin(\phi) \end{smallmatrix}\right)$. By symmetry we restrict to the first quadrant. Let $\phi \in [0, \frac{\pi}{2}]$. The support function of C is

$$h(C, u(\phi)) = \langle \left(\begin{smallmatrix} \cos(\phi) \\ \sin(\phi) \end{smallmatrix}\right), \left(\begin{smallmatrix} L+\cos(\phi) \\ \sin(\phi) \end{smallmatrix}\right) \rangle = 1 + L\cos(\phi). \tag{3.58}$$

Using (3.56), the polar is given in polar coordinates by

$$C^\circ = \{(r, \phi) : r \leq \rho(C^\circ, u(\phi))\}$$

with the radial function (3.57)

$$\rho(C^\circ, u(\phi)) = h(C, u(\phi))^{-1} = (1 + L\cos(\phi))^{-1}. \tag{3.59}$$

Following the involutive character of polarization by direct calculation, let us derive the support function of the polar C°. The slope of the boundary curve of C° in the first quadrant at the intersection point with the ray $\mathbb{R}_0^+ u(\phi)$ is

$$s(\phi) := \frac{\rho'(C, u(\phi))\sin(\phi) + \rho(C, u(\phi))\cos(\phi)}{\rho'(C, u(\phi))\cos(\phi) - \rho(C, u(\phi))\sin(\phi)} = -\frac{L + \cos(\phi)}{\sin(\phi)}.$$

The intersection points of C° with the coordinate axes are in polar respectively Cartesian coordinates

$$(\rho(C^\circ, u(0)), 0) \simeq \begin{pmatrix} \frac{1}{1+L} \\ 0 \end{pmatrix} \qquad \text{and} \qquad (\rho(C^\circ, u(\frac{\pi}{2})), \frac{\pi}{2}) \simeq \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right).$$

There the slope of the boundary curve of C° is $s(0) = -\infty$ and $s(\frac{\pi}{2}) = -L$, see Figure 3.8, right drawing. Accordingly, we calculate the support function of C° separately for slopes a in the segments $[-\infty, -L]$ and $[-L, 0]$.

As a coordinate for a normal vector $v(\psi) = \begin{pmatrix} \cos(\psi) \\ \sin(\psi) \end{pmatrix}$ of C° we use in addition to the angle $\psi \in [0, \frac{\pi}{2}]$ also the slope

$$a = -\tan(\psi)^{-1} \in [-\infty, 0]$$

of the supporting hyperplane $H(C^\circ, v(\psi))$ and denote $\widetilde{v}(a) := v(\psi)$ with inverse coordinate transformation $\psi = \arctan(-\frac{1}{a})$.

In the case $a \in [-L, 0]$, the supporting hyperplane $H(C^\circ, \widetilde{v}(a))$ intersects C° in $\left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$. The support function is thus

$$h(C^\circ, \widetilde{v}(a)) = \langle \begin{pmatrix} \cos(\psi) \\ \sin(\psi) \end{pmatrix}, \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right) \rangle = \sin(\psi) = \frac{1}{\sqrt{a^2+1}}. \tag{3.60}$$

For slopes $a \in [-\infty, -L]$ the supporting hyperplane $H(C^\circ, \widetilde{v}(a))$ is a tangent to the boundary curve of C°. It meets C° at $(\rho(C^\circ, u(\phi)), \phi)$ for

$$\phi = s^{-1}(a) = \arccos\left(\frac{1}{a^2+1}\left(-L - a\sqrt{a^2 - L^2 + 1}\right)\right).$$

This expression is readily verified. By elementary trigonometry we have

$$h(C^\circ, \widetilde{v}(a)) = \rho(C^\circ, u(\phi))\cos(\phi - \psi) = \rho(C^\circ, u(\phi))\cos\left(s^{-1}(a) - \arctan\left(-\frac{1}{a}\right)\right)$$

and easy calculations[1] yield

$$\cos(\phi - \psi) = \sqrt{1 - \frac{L}{a^2+1}}$$

[1]It is beneficial to substitute $a^2 + 1$ and $\sqrt{a^2 - L^2 + 1}$ by single variables.

and finally

$$h(C^\circ, \widetilde{v}(a)) = \frac{\sqrt{a^2+1}}{\sqrt{a^2-L^2+1}-aL}. \tag{3.61}$$

Using (3.56) we obtain the radial function of $C^{\circ\circ}$ from (3.61) and (3.60):

$$\rho(C^{\circ\circ}, \widetilde{v}(a)) = \begin{cases} \frac{\sqrt{a^2-L^2+1}-aL}{\sqrt{a^2+1}} & \text{for} \quad a \in [-\infty, -L] \\ \sqrt{a^2+1} & \text{for} \quad a \in [-L, 0] \end{cases}. \tag{3.62}$$

This radial function indeed describes the stadium C from the beginning. For a slope $a \in [-\infty, -L]$ we have

$$\left(\rho(C^{\circ\circ}, \widetilde{v}(a))\cos(\psi) - L\right)^2 + \left(\rho(C^{\circ\circ}, \widetilde{v}(a))\sin(\psi)\right)^2 = 1.$$

This describes the arc of the stadium. In terms of the angle coordinate $\psi \in [0, \arctan(\frac{1}{L})]$ the radial function of $C^{\circ\circ}$ is

$$\rho(C^{\circ\circ}, v(\psi)) = L\cos(\psi) + \sqrt{1 - L^2\sin^2(\psi)}.$$

Special cases are $\rho(C^{\circ\circ}, \psi) = 2\cos(\psi)$ for $L = 1$ and $\rho(C^{\circ\circ}, \psi) = 1$ for the sphere $L = 0$. For the slope $a \in [-L, 0]$ of a supporting hyperplane of C° the radial function of $C^{\circ\circ}$ satisfies

$$\rho(C^{\circ\circ}, a)\cos(\psi) = -a \qquad \text{and} \qquad \rho(C^{\circ\circ}, a)\sin(\psi) = 1$$

and resembles the straight side of the stadium. We have verified $C^{\circ\circ} = C$. This example confirms the formula in Remark 3.53.

4 State spaces

In the first section we define the state space in a matrix algebra and give examples. Encouraged by Alfsen and Schultz [Al] we first study the cone of positive matrices in Section 4.2 including the isomorphism between orthogonal projectors of the algebra and faces of the cone. In Section 4.3 the isomorphism is translated to the state space. This is only a special case of the analogue results for von Neumann and C*-algebras [Al]. We describe relative interiors of faces and normal cones, the latter by duality to faces. Another news is the homeomorphism between the projector lattice and the face lattice of the state space equipped with the Hausdorff distance. This is established in Section 4.4.

4.1 Examples and illustrations

The state space in a matrix algebra is defined. Examples provide arithmetics for further usage and they show common ground and differences for abelian and non-abelian matrix algebras.

Definition 4.1 A **state** on a complex unital algebra A is a linear functional $f : A \to \mathbb{C}$ that has non-negative (real) values on positive operators and such that $f(\mathbb{1}) = 1$. Unless otherwise stated we assume that A is a matrix algebra throughout the thesis.

The assumption that A is a matrix algebra implies significant simplifications. We recall the definition of a matrix algebra.

Remark 4.2 The following conventions are introduced broadly in Section 2.1. A matrix algebra is defined for $N \in \mathbb{N}$ and a multi-index $n = (n_1, \ldots, n_N) \in \mathbb{N}_0^N$ as

$$A = M_{n_1} \oplus \cdots \oplus M_{n_N}.$$

Here M_k is the full matrix algebra of complex $k \times k$-matrices for $k \in \mathbb{N}$ and by definition $M_0 = \{0\}$ holds. The classical case is $A = \mathbb{C}^{|n|}$ with indices $n_i \leq 1$ for $i = 1, \ldots, N$ and with $|n| = \sum_{i=1}^{N} n_i$. The algebra A is an algebra of linear operators on the Hilbert space

$$H = H_1 \oplus \cdots \oplus H_N$$

for $H_i = \mathbb{C}^{n_i}$, $i = 1, \ldots, N$ such that $H = \mathbb{C}^{|n|}$. The inner product

$$\langle x, y \rangle = \sum_{i=1}^{|n|} \overline{x_i} y_i$$

is used for $x, y \in H$ and the trace

$$\mathrm{tr}(a) = \sum_{i=1}^{|n|} \langle x_i, a(x_i) \rangle$$

is used for $a \in A$, where $\{x_i\}_{i=1}^{|n|}$ is an arbitrary orthonormal basis of H. The Hilbert-Schmidt inner product (also HS inner product) is given by

$$\langle a, b \rangle = \mathrm{tr}(a^* b)$$

for $a, b \in A$. A matrix algebra A is a Hilbert space with the HS inner product. The following geometric studies take place mainly in the self-adjoint part (2.13)

$$A_{\mathrm{sa}} = \{a \in A : a^* = a\}.$$

This is a Euclidean space with HS inner product $\langle a, b \rangle = \mathrm{tr}(a^* b) = \mathrm{tr}(ab)$ for $a, b \in A_{\mathrm{sa}}$.

Remark 4.3 **(a)** There is a bijection between the states f of A and the positive matrices ρ in A of trace one. The correspondence is given by the relation $f(a) = \mathrm{tr}(a\rho)$ for $a \in A$. See for example Definition (2.2,21) in [Th3] and (2.1,5) in [Th4].

(b) A direct sum representation of a matrix algebra A may seem technical. However there are physical situations having this structure. In elementary particle physics some quantities like electric charge have never been observed in a superposition of different values. The corresponding restriction on the algebra is called a **superselection rule**, see [Bel, Ben, Va].

The following definitions are prepared in Section 2.2.

Definition 4.4 One of the main issues is the study of the projector lattice of A (2.16) which we abbreviate by

$$\mathcal{P} := \mathcal{P}(A) = \{p \in A : p^2 = p = p^*\}.$$

A positive matrix $\rho \in A$ with trace one is called a **density matrix** or a **state**. The **state space** of A is the set of density matrices

$$\overline{S}(A) := \{\rho \in A : \rho \geq 0, \mathrm{tr}(\rho) = 1\} = A^+ \cap A^1_{\mathrm{sa}}. \tag{4.1}$$

Here, we use the affine space $A^1_{\mathrm{sa}} = \{a \in A_{\mathrm{sa}} : \mathrm{tr}(a) = 1\}$ of trace one self-adjoint matrices, see (2.14), and the positive cone (2.10) $A^+ = \{a \in A : a \geq 0\}$. The space of **invertible** density matrices is

$$S(A) := \{\rho \in \overline{S}(A) : s(\rho) = \mathbb{1}\} \tag{4.2}$$

where $s(a)$ denotes the support projector (2.20) of a normal matrix $a \in A$. A state is **pure**, if it is an extreme point of the state space.

The state space $\overline{S}(A)$ is a compact convex set. A state $\rho \in \overline{S}(A)$ is pure if and only if ρ is a rank one orthogonal projector. We give two metric properties that are easily deduced from the literature.

Remark 4.5 **(a)** The HS distance between two states $\rho, \sigma \in \overline{S}(A)$ is bounded by

$$\|\rho - \sigma\|_2 \leq \sqrt{2}$$

with equality if and only if ρ and σ are orthogonal pure states. The HS-norm is the Euclidean norm for convex geometric analysis of the state space.

(b) The spectral norm plays only a minor role for our analysis of the state space albeit of great utility for the study of the projector lattice $\mathcal{P}$. We notice for $\rho, \sigma \in \overline{S}(A)$ the spectral norm bound

$$\|\rho - \sigma\| \leq 1$$

with equality if and only if ρ and σ are orthogonal and one of the states is pure.

[Proof on page 204]

Before approaching more advanced problems we recall the situation for the algebra M_2 of 2×2 matrices. We will come back frequently to the following example.

Example 4.6 The **Pauli matrices** [Lev] are the matrices in M_2

$$\sigma_1 := \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right), \quad \sigma_2 := \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right) \quad \text{and} \quad \sigma_3 := \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right). \tag{4.3}$$

One has $\sigma_i^* = \sigma_i$ and $\sigma_i^2 = \mathbb{1}$ for $i = 1, 2, 3$ so the Pauli matrices are self-adjoint and unitary. The anti-commutators are $\sigma_i\sigma_j + \sigma_j\sigma_i = 0$ for distinct $i, j \in \{1, 2, 3\}$. Then $\langle \sigma_i, \sigma_j \rangle = \frac{1}{2}\operatorname{tr}(\sigma_i\sigma_j + \sigma_j\sigma_i) = 0$ for $i \neq j$ shows that

$$\frac{1}{\sqrt{2}}\{\mathbb{1}, \sigma_1, \sigma_2, \sigma_3\}$$

is an ONB for the self-adjoint part $(M_2)_{\mathrm{sa}}$ of the algebra of complex 2×2 matrices. Let $a, b \in \mathbb{R}^3$. The mapping

$$\mathbb{R} \times \mathbb{R}^3 \longrightarrow (M_2)_{\mathrm{sa}}, \quad (\lambda, a) \mapsto \lambda\mathbb{1} + a\widehat{\sigma}$$

is a homothety with ratio $\sqrt{2}$ when the Euclidean scalar product $\langle (\lambda, a), (\mu, b) \rangle = \lambda\mu + a_1b_1 + a_2b_2 + a_3b_3$ is used on on $\mathbb{R}^4 \simeq \mathbb{R} \times \mathbb{R}^3$ and when the HS scalar product is used on $(M_2)_{\mathrm{sa}}$. We use the Euclidean scalar product $\langle a, b \rangle = a_1b_1 + a_2b_2 + a_3b_3$ on $\mathbb{R}^3$ and the Euclidean norm $|a| := \sqrt{\langle a, a \rangle}$. We abbreviate $\widehat{\sigma} := (\sigma_1, \sigma_2, \sigma_3)$ and define

$$a\widehat{\sigma} := a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3.$$

Spectral theory of 2×2-matrices is easy to describe. Let $\lambda \in \mathbb{R}$ and assume $b \neq 0$. The matrix $\lambda\mathbb{1} + b\widehat{\sigma}$ has eigenvalues and eigenvectors

$$\mu_\pm(\lambda\mathbb{1} + b\widehat{\sigma}) = \lambda \pm |b| \quad \text{and} \quad p_\pm(\lambda\mathbb{1} + b\widehat{\sigma}) = \frac{1}{2}(\mathbb{1} \pm \frac{b}{|b|}\widehat{\sigma}). \tag{4.4}$$

It follows that $\lambda\mathbb{1} + a\widehat{\sigma}$ is a density matrix if and only if $\lambda = \frac{1}{2}$ and $|a| \leq \frac{1}{2}$. The state space of M_2 is the **Bloch ball**

$$\overline{S}(M_2) = \left\{ \frac{1}{2}(\mathbb{1} + a\widehat{\sigma}) : |a| \leq 1, a \in \mathbb{R}^3 \right\}, \tag{4.5}$$

see Figure 4.1. The surface of $\overline{S}(M_2)$ consisting of the pure states in M_2 is known [Ni] as the **Bloch sphere**. Conjugation by a projector $p_\pm(b\widehat{\sigma})$ gives

$$p_\pm(b\widehat{\sigma})(\lambda\mathbb{1} + a\widehat{\sigma})p_\pm(b\widehat{\sigma}) = (\lambda \pm \langle a, \frac{b}{|b|} \rangle)p_\pm(b\widehat{\sigma}). \tag{4.6}$$

This can be proved by orthogonal projection: for $v \in A_{\mathrm{sa}}$ and for a rank one projector $p = p_{\pm}(b\widehat{\sigma})$ we have $pvp = \langle p, v\rangle p$.

By functional calculus for a function $f : U \to \mathbb{C}$ with $\{\lambda + |b|, \lambda - |b|\} \subset U \subset \mathbb{C}$

$$f(\lambda\mathbb{1} + b\widehat{\sigma}) = \frac{1}{2}(f(\lambda + |b|) + f(\lambda - |b|))\mathbb{1} + \frac{1}{2}(f(\lambda + |b|) - f(\lambda - |b|))\frac{b}{|b|}\widehat{\sigma}$$

holds. We have an invariance for a vector space $W \subset \mathrm{Lin}\{\sigma_1, \sigma_2, \sigma_3\}$. If $M \subset (\mathbb{R}\mathbb{1} + W)$ and if f is defined on M then

$$f(M) \subset \mathbb{R}\mathbb{1} + W. \tag{4.7}$$

As an example for functional calculus, the exponential function is

$$\exp(\lambda\mathbb{1} + b\widehat{\sigma}) = e^{\lambda}\left(\cosh(|b|)\mathbb{1} + \sinh(|b|)\frac{b}{|b|}\widehat{\sigma}\right) \tag{4.8}$$

and for $\lambda > |b|$ the logarithm is

$$\ln(\lambda\mathbb{1} + b\widehat{\sigma}) = \frac{1}{2}\left(\ln(\lambda^2 - |b|^2)\mathbb{1} + \ln\left(\frac{\lambda + |b|}{\lambda - |b|}\right)\frac{b}{|b|}\widehat{\sigma}\right). \tag{4.9}$$

Remark 4.7 The classical resp. quantum state spaces shown in the top row of Figure 4.1 are discussed in Example 4.9 resp. Example 4.6. We notice that the tetrahedron has only flat boundary components while the Bloch ball has a curved surface.

The images in the bottom row of the figure combine flat and curved shapes. The intersection $\overline{S}(M_3) \cap (\frac{\mathbb{1}}{3} + V)$ for the vector space V specified in the figure is $\overline{S}(M_2 \oplus \mathbb{C}) \cap (\frac{\mathbb{1}}{3} + V)$ because $V \subset M_2 \oplus \mathbb{C}$. The elliptic shape of the intersection is calculated on page 204. The transition to the smaller algebra $M_2 \oplus \mathbb{C}$ is also possible for the projection. This fact is proved in Proposition 5.15. The generation of the projection shape is explained in Example 5.7 where the present case is picked up in (5.36) on page 105. An interesting quantity calculated in the example is the angle $\varphi = \angle(z, V) \approx 0.28\pi$ between V and the vector $z = (-\frac{\mathbb{1}_2}{2}) \oplus 1$ because φ can serve as another argument for the elliptic shape of the intersection: the conic frustum in Figure 5.2 has angular aperture $\frac{\pi}{6} < \varphi$ hence the conic section is in the elliptic regime.

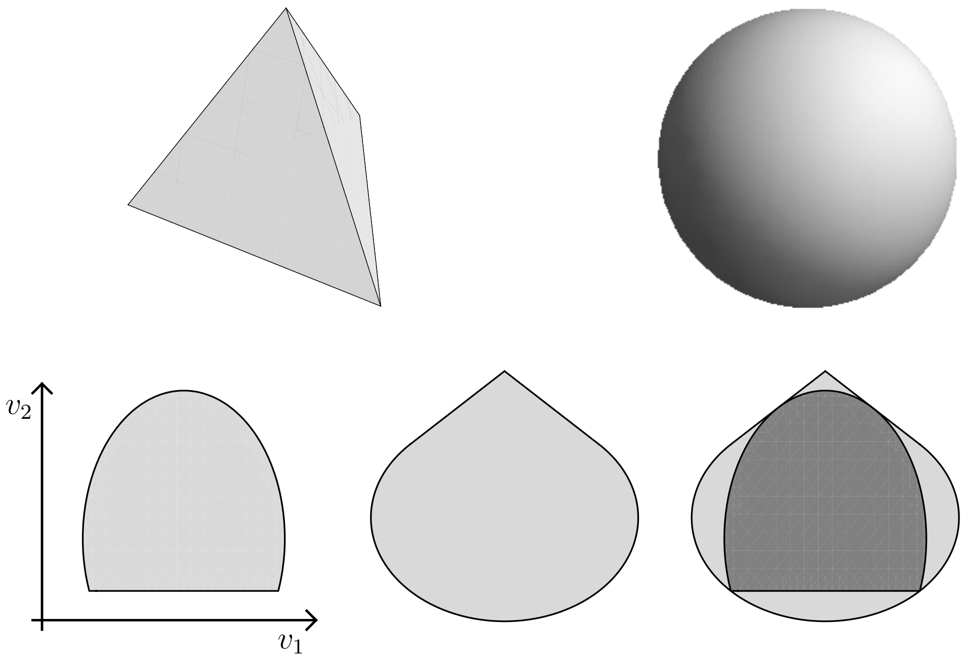


Figure 4.1: The top row shows the tetrahedron $\overline{S}(\mathbb{C}^4)$ and the state space $\overline{S}(M_2)$ known as the Bloch ball (from left to right). Below the state space of the algebra $\overline{S}(M_3)$ is shown in two-dimensional reductions. The left image is the intersection with the affine space $\frac{\mathbb{1}}{3}+V$ and the middle image is the projection to V where V is spanned by the perpendicular vectors $v_1 := \frac{1}{2}(\sigma_1 - \sigma_2) \oplus 0$ and $v_2 := \frac{1}{2}(\sigma_1 + \sigma_2) \oplus 1 - \frac{\mathbb{1}}{3}$ with Pauli matrices σ_1 and σ_2. On the right the two images are overlaid ($\mathbb{1}$ is perpendicular to V). See Remark 4.7 for further details.

Lemma 4.8 Let A be a matrix algebra $A = A_1 \oplus \cdots \oplus A_N$ with matrix algebras A_i $i = 1, \ldots, N$. Then the state space of A is the convex hull

$$\overline{S}(A) = \mathrm{conv}\left(\bigcup_{i=1}^{N} 0 \oplus \cdots 0 \oplus \overline{S}(A_i) \oplus 0 \cdots \oplus 0 \right).$$

[Proof on page 205]

Example 4.9 Lemma 4.8 applied to the classical case $\mathbb{C}^N$ with state space $\overline{S}(\mathbb{C}) = \{1\}$ shows that $\overline{S}(\mathbb{C}^N)$ is the probability simplex (1.25) defined on page 20. This is an $(N-1)$-dimensional equilateral simplex of edge length $\sqrt{2}$.

4.2 The cone of positive matrices

Here we review and prove relevant issues about the positive cone $A^+ = \{a \in A : a \geq 0\}$ (2.10) of a matrix algebra A. The isomorphism between the face lattice of A^+ (without $\emptyset$) and the projector lattice of A is a special case of the situation in a von Neumann algebra [Al]. Another parallel to Alfsen and Schultz's work is the usage of compressions and exposed faces.

As our own contribution we can add a description of the normal cone structure of A^+ (which follows from duality) and we include details about the relative interiors and affine hulls of faces and normal cones. An easy example of a positive cone is depicted in Figure 4.2.

Remark 4.10 **(a)** For a normal matrix $a \in A$ we denote the kernel projector (2.19) of a by $k(a)$. This is the eigenprojector of zero in the spectral decomposition (2.17) of a if zero belongs to the spectrum of a. Otherwise $k(a) = 0$. The support projector of a is $s(a) = \mathbb{1} - k(a)$.

(b) The following notation may seem cryptic but it makes formulas easier to handle. If the matrix algebra is the direct sum $A = M_{n_1} \oplus \ldots \oplus M_{n_N}$ for a multi-index $n \in \mathbb{N}_0^N$ and $N \in \mathbb{N}$ then for an orthogonal projector $p \in \mathcal{P} = \{p \in A : p^2 = p = p^*\}$ we denote by

$$\kappa^p : A^p \to pAp$$

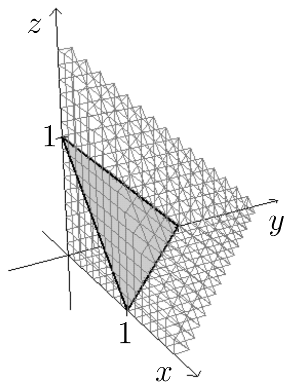


Figure 4.2: The octant $\{(x, y, z) \in \mathbb{R}^3 : x \geq 0, y \geq 0, z \geq 0\}$ is the positive cone of the algebra $\mathbb{C}^3$. The picture shows the state space (triangle) of $\mathbb{C}^3$ inside the positive cone (truncated) inside the space $\mathbb{R}^3$ of self-adjoint matrices.

a *-isomorphic and trace-preserving embedding of the compression

$$A^p = M_{\mathrm{rk}(p_1)} \oplus \cdots \oplus M_{\mathrm{rk}(p_N)}$$

into A with image pAp. The morphism κ^p maps normal matrices of A^p to normal matrices in A dominated by p, for example

$$\kappa^p((A^p)^+) = \{a \in A^+ : s(\rho) \leq p\}. \tag{4.10}$$

Further relations are summarized at (2.28) and (2.31) on page 40.

Proposition 4.11 The positive cone A^+ is a closed convex cone with affine hull and translation vector space equal to the space of self-adjoint matrices:

$$\mathrm{aff}(A^+) = \mathrm{lin}(A^+) = A_{\mathrm{sa}}. \tag{4.11}$$

The barrier cone (3.29) is

$$B(A^+) = -A^+. \tag{4.12}$$

The support function (3.30) of the positive cone evaluated in $b \in B(A^+)$ is

$$h(A^+, b) = 0. \tag{4.13}$$

The relative interior (3.3) of the positive cone is an open set in A_{sa},

$$\mathrm{ri}(A^+) = \{a \in A^+ : s(a) = \mathbb{1}\}. \tag{4.14}$$

The exposed face (3.32) of the positive cone, exposed by a vector $b \in B(A^+) \setminus \{0\}$ is

$$F_\perp(A^+, b) = \kappa^{k(b)}\left((A^{k(b)})^+\right). \tag{4.15}$$

The affine hull and translation vector space of this face are

$$\operatorname{aff}(F_\perp(A^+, b)) = \operatorname{lin}(F_\perp(A^+, b)) = \kappa^{k(b)}\left((A^p)_{\mathrm{sa}}\right). \tag{4.16}$$

The relative interior is

$$\operatorname{ri}\left(F_\perp(A^+, b)\right) = \kappa^{k(b)}\left(\operatorname{ri}\left((A^{k(b)})^+\right)\right). \tag{4.17}$$

[Proof on page 205]

Example 4.12 For $A = \mathbb{C}^3$ in Figure 4.2 we consider the proper exposed faces. These are exposed by the vectors $-p$ for non-zero orthogonal projectors p, see Proposition 4.11. The projector lattice is

$$\mathcal{P} = \{0, (0,0,1), (0,1,0), (1,0,0), (0,1,1), (1,0,1), (1,1,0), \mathbb{1}\},$$

and yields as faces the three quarter-planes

$$\begin{aligned} &\{(x,y,z) \in \mathbb{R}^3 : x \geq 0, y \geq 0, z = 0\},\\ &\{(x,y,z) \in \mathbb{R}^3 : z \geq 0, x \geq 0, y = 0\},\\ &\{(x,y,z) \in \mathbb{R}^3 : y \geq 0, z \geq 0, x = 0\},\end{aligned}$$

the tree half-lines

$$\begin{aligned} &\{(x,y,z) \in \mathbb{R}^3 : x \geq 0, y = z = 0\},\\ &\{(x,y,z) \in \mathbb{R}^3 : y \geq 0, x = z = 0\},\\ &\{(x,y,z) \in \mathbb{R}^3 : z \geq 0, x = y = 0\},\end{aligned}$$

and the origin $\{(0,0,0)\}$ in this order.

Corollary 4.13 Every face of the positive cone A^+ is exposed. The mapping

$$\mathcal{P} \to \mathcal{F}(A^+) \setminus \{\emptyset\}, \quad p \mapsto \kappa^p((A^p)^+)$$

is an isomorphism of complete lattices. [Proof on page 207]

Corollary 4.14 The normal cone (3.36) of the positive cone at $a \in A^+$ is

$$\mathrm{N}(A^+, a) = -\kappa^{k(a)}\left((A^{k(a)})^+\right). \tag{4.18}$$

The relative interior of the normal cone at $a \in A^+$ is

$$\operatorname{ri}\left(\mathrm{N}(A^+, a)\right) = -\kappa^{k(a)}\left(\operatorname{ri}((A^{k(a)})^+)\right). \tag{4.19}$$

[Proof on page 207]

4.3 Faces and normal cones

In the context of a matrix algebra A we follow Alfsen and Schultz's [Al] arguments for C*-algebras and von Neumann algebras. The state space is a base for the positive cone, that is $A^+ = \{\lambda\overline{S}(A) : \lambda \geq 0\}$. We use the intersection property (4.1)

$$\overline{S}(A) = A^+ \cap A^1_{\mathrm{sa}}$$

and translate the face and normal cone lattices from the positive cone A^+ to the state space $\overline{S}(A)$. The isomorphism between the face lattice of $\overline{S}(A)$ and the projector lattice of A are special cases of the infinite dimensional results.

As an own contribution we translate from the positive cone the structure of normal cones and details about the relative interiors and affine hulls of faces and normal cones.

Definition 4.15 If $p \neq 0$ is an orthogonal projector in A, the **trace state** in the compression A^p is

$$\widehat{\mathbb{1}^p} := \frac{\mathbb{1}^p}{\mathrm{tr}(\mathbb{1}^p)} \in \overline{S}(A^p), \tag{4.20}$$

with $\mathbb{1}^p$ the identity in A^p, see Definition 2.19 on page 39. Since $\mathbb{1}^{\mathbb{1}} = \mathbb{1}$ we write $\widehat{\mathbb{1}}$ for the trace state of A. The state

$$\frac{p}{\mathrm{tr}(p)} = \kappa^p(\widehat{\mathbb{1}^p}) \in \overline{S}(A) \tag{4.21}$$

is called the **centroid** with support p. The centroid with full support $\mathbb{1}$ is the trace state of A.

Remark 4.16 If p is an orthogonal projector in A then the embedding of the state space $\overline{S}(A^p)$ of the compression A^p into A under κ^p consists of all states $\rho \in \overline{S}(A)$ of the algebra A that are dominated by p (Remark 2.22 (b))

$$\kappa^p(\overline{S}(A^p)) = \{\rho \in \overline{S}(A) : s(\rho) \leq p\}. \tag{4.22}$$

The embedding of the invertible states

$$\kappa^p(S(A^p)) = \{\rho \in \overline{S}(A) : s(\rho) = p\} \tag{4.23}$$

contains the centroid with support p.

The following is the state space counterpart to Proposition 4.11 for the positive cone, concerning exposed faces. We recall for a self-adjoint matrix $v \in A_{\mathrm{sa}}$ that the maximal eigenvalue of v is denoted $\mu_+(v)$ and the maximal projector by $p_+(v)$, this is the eigenprojector of v for the eigenvalue $\mu_+(v)$.

Definition 4.17 The face lattice of the state space is denoted

$$\mathcal{F} := \mathcal{F}(\overline{S}(A)).$$

The support function, supporting hyperplanes and exposed faces are defined for $u \in A_{\mathrm{sa}} \setminus \{0\}$ and are denoted by

$$\begin{aligned} h(u) &:= h(\overline{S}(A), u) = \max_{\rho \in \overline{S}(A)} \langle u, \rho \rangle, \\ H(u) &:= H(\overline{S}(A), u), \\ F_\perp(u) &:= F_\perp(\overline{S}(A), u). \end{aligned}$$

In addition, $h(0) = 0$. The face of a density matrix $\rho \in \overline{S}(A)$ is denoted

$$F(\rho) := F(\overline{S}(A), \rho).$$

This is the unique face of the state space $\overline{S}(A)$ with ρ in the relative interior.

Proposition 4.18 The state space $\overline{S}(A)$ is a compact convex set of dimension

$$\dim_{\mathbb{R}}\left(\overline{S}(A)\right) = \dim_{\mathbb{R}}(A_{\mathrm{sa}}) - 1 = \dim_{\mathbb{C}}(A) - 1 \tag{4.24}$$

with affine hull

$$\mathrm{aff}\left(\overline{S}(A)\right) = A^1_{\mathrm{sa}} \tag{4.25}$$

and translation vector space

$$\mathrm{lin}\left(\overline{S}(A)\right) = A^0_{\mathrm{sa}}. \tag{4.26}$$

The relative interior consists of the invertible density matrices

$$\mathrm{ri}\left(\overline{S}(A)\right) = S(A). \tag{4.27}$$

The support function evaluated in $u \in A_{\mathrm{sa}}$ is

$$h(u) = \mu_+(u). \tag{4.28}$$

The exposed face of $\overline{S}(A)$ exposed by $u \in A_{\mathrm{sa}} \setminus \{0\}$ is

$$F_\perp(u) = \kappa^{p_+(u)}\left(\overline{S}(A^{p_+(u)})\right). \tag{4.29}$$

The affine hull of this face is

$$\operatorname{aff}(F_\perp(u)) = \kappa^{p_+(u)}\left((A^{p_+(u)})^1_{\mathrm{sa}}\right), \tag{4.30}$$

the translation vector space is

$$\operatorname{lin}(F_\perp(u)) = \kappa^{p_+(u)}\left((A^{p_+(u)})^0_{\mathrm{sa}}\right) \tag{4.31}$$

and the relative interior is

$$\operatorname{ri}(F_\perp(u)) = \kappa^{p_+(u)}\left(S(A^{p_+(u)})\right). \tag{4.32}$$

[Proof on page 207]

Definition 4.19 The **face with support** $p \in \mathcal{P}$ for the state space $\overline{S}(A)$ is defined as

$$\mathbb{F}(p) := \kappa^p\left(\overline{S}(A^p)\right). \tag{4.33}$$

The following is the state space counterpart to Corollary 4.13 on the positive cone, concerning the lattice isomorphism between projectors and faces. We assume the matrix algebra is the direct sum $A = M_{n_1} \oplus \ldots \oplus M_{n_N}$ for a multi-index $n \in \mathbb{N}_0^N$ and $N \in \mathbb{N}$. For a matrix $a \in A$ the matrix $a_i \in M_{n_i}$ denotes the i-th direct summand of a, $i = 1, \ldots, N$.

Corollary 4.20 Every face of $\overline{S}(A)$ is exposed. The allocation of faces to orthogonal projectors

$$\mathbb{F}: \quad \mathcal{P} \to \mathcal{F}, \quad p \mapsto \mathbb{F}(p) \tag{4.34}$$

is an isomorphism of complete lattices. For a projector $p \in \mathcal{P}$ the face with support p has dimension

$$\dim_{\mathbb{R}}(\mathbb{F}(p)) = \dim_{\mathbb{R}}((A^p)_{\mathrm{sa}}) - 1 = \sum_{i=1}^{N} \operatorname{rk}(p_i)^2 - 1. \tag{4.35}$$

The affine hull is

$$\operatorname{aff}(\mathbb{F}(p)) = \kappa^p\left((A^p)^1_{\mathrm{sa}}\right), \tag{4.36}$$

the translation vector space is

$$\operatorname{lin}(\mathbb{F}(p)) = \kappa^p\left((A^p)^0_{\mathrm{sa}}\right) \tag{4.37}$$

and the relative interior is

$$\operatorname{ri}(\mathbb{F}(p)) = \kappa^p\left(S(A^p)\right). \tag{4.38}$$

[Proof on page 209]

With Corollary 4.20 we can set the following definitions.

Definition 4.21 The lattice isomorphism inverse to $\mathbb{F} : \mathcal{P} \to \mathcal{F}$ is denoted

$$s : \quad \mathcal{F} \to \mathcal{P}, \qquad F \mapsto s(F). \tag{4.39}$$

The orthogonal projector $s(F)$ is called the **support projector** of a face $F \in \mathcal{F}$.

Remark 4.22 (a) The lattice isomorphism in Corollary 4.20 may be expressed functionally by $\mathbb{F}(s(F)) = F$ and $s(\mathbb{F}(p)) = p$ for faces $F \in \mathcal{F}$ and for projectors $p \in \mathcal{P}$. In particular by (4.33) we have

$$F = \kappa^{s(F)}(\overline{S}(A^{s(F)})). \tag{4.40}$$

For a density matrix $\rho \in \overline{S}(A)$ we decide the face membership

$$\rho \in F \iff s(\rho) \leq s(F) \tag{4.41}$$

using (4.22). With (4.38) and (4.23) we decide the membership

$$\rho \in \mathrm{ri}(F) \iff s(\rho) = s(F). \tag{4.42}$$

Since $\rho \in \mathrm{ri}(F(\rho))$ the equation (4.42) gives

$$s(F(\rho)) = s(\rho) \tag{4.43}$$

and application of the lattice isomorphism $\mathbb{F}$ gives

$$F(\rho) = \mathbb{F}(s(\rho)). \tag{4.44}$$

By (4.43) and by (4.41) we obtain

$$\rho \in F(\sigma) \quad \iff \quad s(\rho) \leq s(\sigma) \tag{4.45}$$

for two density matrices $\rho, \sigma \in \overline{S}(A)$. The dimension formula (4.35) and (4.44) give

$$\dim(F(\rho)) = \sum_{i=1}^{N} \mathrm{rk}(\rho_i)^2 - 1 \tag{4.46}$$

where we notice $\mathrm{rk}(s(\rho_i)) = \mathrm{rk}(\rho_i)$ for $i = 1, \ldots, N$.

(b) Exposed faces are represented by projectors. For a vector $v \in A_{\mathrm{sa}} \setminus \{0\}$ we have by (4.29) and (4.33)

$$F_\perp(v) = \mathbb{F}(p_+(v)) \tag{4.47}$$

with the maximal eigenprojector $p_+(v)$ of v. This shows

$$s(F_\perp(v)) = p_+(v) \tag{4.48}$$

and we obtain the following commuting diagram.

$$\begin{tikzcd} & \mathcal{F} \arrow[dd, shift left, "\mathbb{F}"] \\ A_{\mathrm{sa}} \setminus \{0\} \arrow[ur, "F_\perp"] \arrow[dr, "p_+"'] & \\ & \mathcal{P} \arrow[uu, shift left, "s"] \end{tikzcd} \tag{4.49}$$

Notice that $\emptyset \in \mathcal{F}$ resp. $0 \in \mathcal{P}$ do not belong to the image of $A^0_{\mathrm{sa}} \setminus \{0\}$ under $F_\perp$ resp. p_+.

(c) The Corollaries 4.13 and 4.20 with isomorphisms between the projector lattice of A and the face lattices of the positive cone or of the state space complement the isomorphism $p \mapsto \mathrm{Im}(p)$ between the projector lattice and the Grassmannian presented in Section 2.3.

Definition 4.23 The **normal cone** (3.36) of the state space at $\rho \in \overline{S}(A)$ is denoted

$$\mathrm{N}(\rho) := \mathrm{N}(\overline{S}(A), \rho). \tag{4.50}$$

The state space and a normal cone is considered a subset of the Euclidean space of self-adjoint matrices in A. The **normal cone lattice** (3.43) of the state space is denoted

$$\mathcal{N} := \mathcal{N}(\overline{S}(A)). \tag{4.51}$$

Proposition 4.24 The normal cone of the state space at $\rho \in \overline{S}(A)$ is

$$\mathrm{N}(\rho) = \{a \in A_{\mathrm{sa}} : p_+(a) \geq s(\rho)\} \tag{4.52}$$

and the relative interior is

$$\mathrm{ri}\,(\mathrm{N}(\rho)) = \{a \in A_{\mathrm{sa}} : p_+(a) = s(\rho)\}. \tag{4.53}$$

[Proof on page 209]

From (4.32) and (4.53) we have the following result.

Corollary 4.25 Let $u \in A_{\mathrm{sa}} \setminus \{0\}$ and $\rho \in \overline{S}(A)$. Equivalent are

$$\begin{array}{ll} \text{(a)} & \rho \in \mathrm{ri}(F_\perp(u)), \\ \text{(b)} & u \in \mathrm{ri}(\mathrm{N}(\rho)), \\ \text{(c)} & s(\rho) = p_+(u). \end{array} \tag{4.54}$$

Remark 4.26 The equivalence (4.54) is a stronger variant of the ubiquitous duality (3.39)

$$x \in F_\perp(C,u) \quad \Longleftrightarrow \quad \langle u, x\rangle = h(C,u) \quad \Longleftrightarrow \quad u \in \mathrm{N}(C,x)$$

which is true for an arbitrary convex set $C \subset \mathbb{R}^n$, for vectors $u \in B(C) \setminus \{0\}$ and points $x \in C$. Some geometric content of the equivalence of (a) and (b) in (4.54) is examined in Section 3.5.

4.4 The face manifold

Beyond the lattice isomorphism between the projector lattice of a matrix algebra and the face lattice of the state space (4.39)

$$s : \mathcal{F} \to \mathcal{P}$$

treated in the previous section, we can prove the new result that the face lattice $\mathcal{F}$ equipped with Hausdorff distance is homeomorphic to the projector lattice. This gives the structure of a partition into compact differentiable manifolds to the face lattice. An easy consequence is that the dimension function on the state space is lower-semicontinuous, a fact which is probably more familiar because it follows from the well-known lower semi-continuity of the rank function.

We assume the matrix algebra is the direct sum $A = M_{n_1} \oplus \ldots \oplus M_{n_N}$ for a multi-index $n \in \mathbb{N}_0^N$ and $N \in \mathbb{N}$. For a matrix $a \in A$ the matrix $a_i \in M_{n_i}$ denotes the i-th direct summand of a, $i = 1, \ldots, N$. The conjugation classes (by unitaries) of the projector lattice $\mathcal{P}$ were described in Lemma 2.38 on page 49. They are given for multi-indices $k \in \mathbb{N}_0^N$ such that $k \leq n$ by the conjugation manifolds

$$\mathcal{P}_k = \mathcal{P}_{k_1}(M_{n_1}) \oplus \cdots \oplus \mathcal{P}_{k_N}(M_{n_N})$$

where $\mathcal{P}_{k_i}(M_{n_i}) := \{p \in \mathcal{P}(M_{n_i}) : \mathrm{rk}(p) = k_i\}$ for $i = 1, \ldots, N$. As a Cartesian product of Grassmannian manifolds, a conjugation manifold (of projectors) is a compact differentiable manifold.

The conjugation manifolds of projectors are characterized as equivalence classes of conjugation by a unitary. The idea is that we can find an analogue structure of the face lattice. Let $v \in A$ be a unitary. Then the isometry (with respect to HS scalar product) of A

$$a \to vav^*$$

restricts to a bijection $\overline{S}(A) \to \overline{S}(A)$. A complete description of "automorphism of state space" is found in Theorem 7.33 in [Va].

Definition 4.27 We call two faces $F, G \in \mathcal{F}$ **conjugate** if for some unitary $v \in A$ we have

$$G = vFv^*. \tag{4.55}$$

The **face manifold** for a multi-index $k \in \mathbb{N}_0^N$ with $k \leq n$ is defined by

$$\mathcal{F}_k := \{\mathbb{F}(p) : p \in \mathcal{P}_k\}. \tag{4.56}$$

Lemma 4.28 The conjugation classes (by unitaries) of the face lattice $\mathcal{F}$ are the face manifolds $\mathcal{F}_k$ for the multi-indices $k \in \mathbb{N}_0^N$ with $k \leq n$. In detail, if $p \in \mathcal{P}$ then $v\mathbb{F}(p)v^* = \mathbb{F}(vpv^*)$. For faces $F, G \in \mathcal{F}$ and a unitary $v \in A$ $G = vFv^* \iff s(G) = vs(F)v^*$ holds. [Proof on page 210]

For topological studies of face manifolds we use the Hausdorff distance.

Definition 4.29 (a) We denote the space of compact subsets of $\mathbb{R}^m$ by $\mathcal{C}(\mathbb{R}^m)$. The **Hausdorff distance** is given by

$$\mathrm{d_H}(C, D) := \max(\sup_{x\in C} \inf_{y\in D} \|x - y\|_2, \sup_{y\in D} \inf_{x\in C} \|x - y\|_2) \tag{4.57}$$

for $C, D \in \mathcal{C}(\mathbb{R}^m)$. We denote $\mathcal{K}(\mathbb{R}^m)$ the space of compact and convex subsets of $\mathbb{R}^m$.

(b) We denote the space of compact subsets of self-adjoint matrices by

$$\mathcal{C}(A_{\mathrm{sa}}) := \{C \subset A_{\mathrm{sa}} : C \text{ is compact }\}$$

and consider the Hausdorff distance

$$\mathrm{d_H}(C, D) = \max(\sup_{a\in C} \inf_{b\in D} \|a - b\|_2, \sup_{a\in D} \inf_{b\in C} \|a - b\|_2) \tag{4.58}$$

for non-empty $C, D \in \mathcal{C}(A_{\mathrm{sa}})$ with respect to the HS norm $\| \cdot \|_2$ on A_{sa}. The space of compact and convex subsets of A_{sa} is denoted $\mathcal{K}(A_{\mathrm{sa}})$.

Remark 4.30 $\mathcal{C}(\mathbb{R}^m)$ is a complete metric space for the Hausdorff distance, see Theorem 1.8.2 in [Sch]. In this way, the space $\mathcal{C}(A_{\mathrm{sa}})$ becomes a complete metric space for the Hausdorff distance defined in (4.58). By Theorem 1.8.5 in [Sch] the space of compact convex subsets $\mathcal{K}(\mathbb{R}^m)$ is a closed subset of the space of compact subsets $\mathcal{C}(\mathbb{R}^m)$ with respect to Hausdorff distance. Thus the space $\mathcal{K}(A_{\mathrm{sa}})$ of compact and convex subsets of self-adjoint matrices is a complete metric space with respect to Hausdorff distance.

Remark 4.31 Theorem 1.8.7 in [Sch] gives conditions for convergence in $\mathcal{K}(\mathbb{R}^m)$: a sequence (K_i) in $\mathcal{K}(\mathbb{R}^m)$ converges to $K \in \mathcal{K}(\mathbb{R}^m)$ if and only if the following two conditions hold.

(a) every point in K is the limit of a sequence (x_i) where $x_i \in K_i$ for $i \in \mathbb{N}$,

(b) the limit of any convergent sequence (x_{i_j}) with $x_{i_j} \in K_{i_j}$ for $j \in \mathbb{N}$ belongs to K.

At (b) of course we have to choose an injective enumeration $\mathbb{N} \to \mathbb{N}$, $j \mapsto i_j$ to define a subsequence.

Proposition 4.32 The lattice isomorphism $\mathcal{P} \to \mathcal{F}$, $p \mapsto \mathbb{F}(p)$ is a homeomorphism for the topology of Hausdorff distance on the face lattice $\mathcal{F}$. Each face manifold $\mathcal{F}_k$ for a multi-index $k \in \mathbb{N}_0^N$ with $k \leq n$ is a compact real analytic manifold. [Proof on page 211]

Example 4.33 The state space $\overline{S}(A)$ of the algebra $A := M_2$ of 2×2-matrices is the Bloch ball (4.5) depicted on page 77. It is a Euclidean three-dimensional ball. The face lattice consists of the face manifolds $\mathcal{F}_0 = \emptyset$, $\mathcal{F}_2 = \overline{S}(A)$ and $\mathcal{F}_1 = \{\rho \in \overline{S}(A) : \mathrm{rk}(\rho) = 1\}$. On the face manifold $\mathcal{F}_1$, the Hausdorff distance simplifies to the Hilbert-Schmidt distance $\mathrm{d_H}(\{\rho\}, \{\sigma\}) = \|\rho - \sigma\|_2$ with $\rho, \sigma \in \overline{S}(A)$ being extreme points of the Bloch ball. In this way, the face manifold $\mathcal{F}_1$ is isometric to the Bloch sphere.

We discuss the dimension function of the state space.

Remark 4.34 **(a)** Recall from Definition 3.12 on page 56 that the dimension function of the state space is the mapping

$$\overline{S}(A) \to \mathbb{N}_0, \quad \rho \mapsto \dim(F(\rho))$$

which assigns to a density matrix ρ the dimension of the face that contains ρ in the relative interior. For $d \in \mathbb{N}_0$ the d-skeleton of $\overline{S}(A)$ is the union of faces

$$\mathrm{skel}(\overline{S}(A), d) = \{\rho \in \overline{S}(A) : \dim(F(\rho)) \leq d\}.$$

(b) A function $f : M \to \mathbb{R}$ for $M \subset \mathbb{R}^m$ is lower semi-continuous if for every convergent sequence (x_i) in M we have (Definition 3.4)

$$f(\lim_{i\to\infty} x_i) \leq \liminf_{i\to\infty} f(x_i).$$

(c) We conclude in Remark 3.13 on page 56 that the dimension function of a closed convex set M in $\mathbb{R}^m$ is lower semi-continuous if and only if all skeletons of M are closed. An example where these two properties fail is given in Figure 5.8 on page 123.

(d) The face manifold $\mathcal{F}_k$ for a multi-index $k \in \mathbb{N}_0^N$ with $k \leq n$ consists of the faces $\mathbb{F}(p) \in \mathcal{F}$ with support p where $p \in \mathcal{P}$ satisfies the rank conditions $\mathrm{rk}(p_i) = k_i$ for $i = 1, \ldots, N$. The dimension of a face $F \in \mathcal{F}_k$ is by (4.35) the constant number

$$\dim(F) = \sum_{i=1}^{N} k_i^2 - 1.$$

(e) For a multi-index $k \in \mathbb{N}_0^N$ with $k \leq n$ we denote the union of all faces in the face manifold $\mathcal{F}_k$ with the usual abbreviation

$$\bigcup \mathcal{F}_k := \bigcup_{F\in\mathcal{F}_k} F.$$

Then with the dimension formula in (d) the d-skeleton of the state space is

$$\mathrm{skel}(\overline{S}(A), d) = \bigcup_k (\bigcup \mathcal{F}_k) \tag{4.59}$$

where the union extends over multi-indices $k \in \mathbb{N}_0^N$ such that $k \leq n$ and $\sum_{i=1}^N k_i^2 - 1 \leq d$.

Proposition 4.35 All skeletons of the state space are compact. The dimension function of the state space is lower semi-continuous. [Proof on page 211]

Alternatively one can use lower semi-continuity of the rank function to prove Proposition 4.35. We include the following proof because it can be used to prove lower semi-continuity Lemma 5.37.

Lemma 4.36 Let $M := \mathbb{C}^{k_1 \times k_2}$ be the space of complex $k_1 \times k_2$-matrices. The rank function $\mathrm{rk} : M \to \mathbb{N}$ is lower semi-continuous on M. [Proof on page 212]

Remark 4.37 The closedness of the d-skeletons is a special property of the state space. It is not clear if lower semi-continuity of the dimension function is inherited by state reflections, which are the orthogonal projections of a state space to a linear subspace. We will approach this problem in Section 5.4.

5 State reflections

In this chapter we study orthogonal projections of the state space in a matrix algebra. We provide new results about its geometric and algebraic nature. Seen through an invertible affine transformation, such a projection is the same as the set of mean values of some observables and the application will be the mean value parametrization for an exponential family in Section 6.3 and its extension in Chapter 7. The projection of a state contains only partial information about a true quantum state; we remember Plato's idea of a "shadow" or "reflection" of reality in Allegory of the Cave.

Definition 5.1 For a matrix algebra A we fix a vector space $V \subset A^0_{\mathrm{sa}}$ of traceless self-adjoint matrices throughout the chapter. The orthogonal projection to V is the linear mapping

$$\pi_V : A_{\mathrm{sa}} \to V \tag{5.1}$$

which satisfies $u - \pi_V(u) \perp V$ for all $u \in A_{\mathrm{sa}}$, see (3.2). Orthogonality is understood with respect to the HS inner product. The compact and convex set

$$\mathrm{sr}_V := \pi_V(\overline{S}(A)) \tag{5.2}$$

is called the **state reflection** on V. For a density matrix $\rho \in \overline{S}(A)$ we call the matrix $\pi_V(\rho) \in \mathrm{sr}_V$ the **mean value** of ρ (with respect to V).

We study the exposed faces of the state reflection sr_V in Section 5.1 and generate examples and shapes. Surprisingly, the face structure of sr_V is richer compared to $\overline{S}(A)$ and non-exposed faces emerge. The introduction of the face lattice is done in Section 5.2 including lattice isomorphisms to projector lattices in the algebra and equivariance assertions of these lattices under an algebra embedding. Using access sequences we can give an algorithm to compute the projector lattices of sr_V in Section 5.3. In the last section topological questions about the projector lattices are formulated.

5.1 Exposed faces and examples

We compute the exposed faces of the state reflection sr_V in terms of maximal projectors of V and generate examples of state reflection shapes. The quantum novelty is documented in Example 5.7 where a conic frustum appears as a state reflection. The projections of the frustum are discussed in detail. Depending on the projection direction non-exposed faces appear for the two-dimensional shape in Example 5.7 (e). At the beginning we consider an example that can be computed without further preparation.

Example 5.2 (State reflection on a face) We consider the translation vector space $\mathrm{lin}(F)$ of the state space face $F := \mathbb{F}(p)$ with support $p \in \mathcal{P} \setminus \{0\}$, see (4.33). An arbitrary matrix $a \in A$ is mapped under orthogonal projection to $\mathrm{lin}(F)$ to the matrix

$$\pi_{\mathrm{lin}(F)}(a) = pap - \mathrm{tr}(pa)\frac{p}{\mathrm{tr}(p)}. \tag{5.3}$$

The projection to the affine hull of the face is

$$\pi_{\mathrm{aff}(F)}(a) = pap + (1 - \mathrm{tr}(pa))\,\frac{p}{\mathrm{tr}(p)}. \tag{5.4}$$

The inequality $|\,\mathrm{tr}(p\rho)| \leq \|p\|\,\mathrm{tr}(\rho) = 1$ (2.12) shows $\pi_{\mathrm{aff}(F)}(\rho) \in F$ for $\rho \in \overline{S}(A)$ and this implies

$$\mathrm{sr}_{\mathrm{lin}(F)} = \pi_{\mathrm{lin}(F)}(\overline{S}(A)) = \pi_{\mathrm{lin}(F)}(F) = F - \frac{p}{\mathrm{tr}(p)}. \tag{5.5}$$

Using the lattice isomorphism (4.39) and (4.40) we can write the state reflection on $\mathrm{lin}(F)$ with the support projector and compression as

$$\mathrm{sr}_{\mathrm{lin}(F)} = \kappa^{s(F)}(\overline{S}(A^{s(F)})) - \frac{s(F)}{\mathrm{tr}(s(F))}. \tag{5.6}$$

Remark 5.3 **(a)** The state reflection on a non-empty face $F \in \mathcal{F}$ is isometric to the state space $\overline{S}(A^{s(F)})$ by (5.6). Since all faces of the state space $\overline{S}(A^{s(F)})$ are exposed by Corollary 4.20, all faces of $\mathrm{sr}_{\mathrm{lin}(F)}$ are exposed.

(b) Faces of a state space are “large”. A non-empty face $F \in \mathcal{F}$ is the projection of the whole state space onto $\mathrm{aff}(F)$ by (5.5). A corresponding property is wrong, for example, for a regular pentagon. It is also wrong for the Kirchhoff polytope. Consider the two short faces of the kite in Figure 9.3 on page 186.

Definition 5.4 The support function (3.30) of sr_V, defined for $v \in V$, is denoted by

$$h_V(v) := h(\mathrm{sr}_V, v) = \sup_{x \in \mathrm{sr}_V} \langle v, x \rangle.$$

If $v \neq 0$, the supporting hyperplane (3.31) of sr_V for v is denoted

$$H_V(v) := H(\mathrm{sr}_V, v) = \{x \in A_{\mathrm{sa}} : \langle v, x \rangle = h_V(v)\}$$

and the exposed face (3.32) of sr_V for v is denoted

$$F_{V,\perp}(v) := F_\perp(\mathrm{sr}_V, v) = \mathrm{sr}_V \cap H_V(v). \tag{5.7}$$

The exposed face lattice (3.33) of sr_V is denoted

$$\mathcal{F}_{V,\perp} := \mathcal{F}_\perp(\mathrm{sr}_V) = \{F_{V,\perp}(v) : v \in V \setminus \{0\}\} \cup \{\emptyset, \mathrm{sr}_V\}. \tag{5.8}$$

The **state space lift** for V of a subset $M \subset A_{\mathrm{sa}}$ is defined by (3.50)

$$L_V(M) := L_V^{\overline{S}(A)}(M) = (M + V^\perp) \cap \overline{S}(A). \tag{5.9}$$

The lifted exposed face lattice (3.51) of sr_V is denoted

$$\mathcal{L}_{V,\perp} := L_V(\mathcal{F}_{V,\perp}) = \{L_V(F) : F \in \mathcal{F}_{V,\perp}\}. \tag{5.10}$$

The **exposed projector lattice** of sr_V is defined by

$$\mathcal{P}_{V,\perp} := s(\mathcal{L}_{V,\perp}) \subset \mathcal{P} \tag{5.11}$$

where $s(F)$ denotes the support projector (4.39) of a face F of the state space. We define the **support projector** of an exposed face $F \in \mathcal{F}_{V,\perp}$ by

$$s_V(F) := s \circ L_V(F). \tag{5.12}$$

The **face reflection** on V with support $p \in \mathcal{P}$ is

$$\mathbb{F}_V(p) := \pi_V \circ \mathbb{F}(p). \tag{5.13}$$

Remark 5.5 **(a)** The support function of sr_V for a vector $v \in V$ is the maximal eigenvalue

$$h_V(v) = \mu_+(v) \tag{5.14}$$

of v. See the summary of support functions in Remark 3.35 on page 65 and recall the support function $h(v) = \mu_+(v)$ of the state space (4.28).

(b) The state reflection sr_V has non-empty interior in V because $\overline{S}(A)$ has a non-empty interior in A^1_{sa}.

(c) The following commuting diagram connects exposed faces of the state reflection sr_V with lifted faces and support projectors.

For non-zero vectors $v \in V$, the formation of an exposed face $F_\perp(v)$ of the state space and $F_{V,\perp}(v)$ of the state reflection is linked by lattice isomorphisms, the state space lift $L_V|_{\mathcal{F}_{V,\perp}} : \mathcal{F}_{V,\perp} \to \mathcal{L}_{V,\perp}$ with inverse $\pi_V|_{\mathcal{L}_{V,\perp}} : \mathcal{L}_{V,\perp} \to \mathcal{F}_{V,\perp}$ the projection to V, see Proposition 3.38. This explains the left triangle in the diagram.

The allocation of a support projector $s : \mathcal{L}_{V,\perp} \to \mathcal{P}_{V,\perp}$ and the association of a face $\mathbb{F} : \mathcal{P}_{V,\perp} \to \mathcal{L}_{V,\perp}$ are inverses to each other by restriction of the lattice isomorphism $\mathcal{F} \to \mathcal{P}$ for the state space (4.49). One has $p_+ = s \circ F_\perp$ and $F_\perp = \mathbb{F} \circ p_+$ on $V \setminus \{0\}$. Here $p_+(v)$ is the maximal projector of a self-adjoint matrix $v \in A_{\mathrm{sa}}$.

The mappings s_V and $\mathbb{F}_V$ fit into the commuting diagram by definition. This way they become lattice isomorphisms.

$$\begin{tikzcd}
 & \mathcal{L}_{V,\perp} \arrow[dd, shift left, "\pi_V"] \arrow[dr, shift left, "s"] & \\
V \setminus \{0\} \arrow[ur, "F_\perp"] \arrow[rr, "p_+"] \arrow[dr, "F_{V,\perp}"'] & & \mathcal{P}_{V,\perp} \arrow[ul, shift left, "\mathbb{F}"] \arrow[dl, shift left, "\mathbb{F}_V"] \\
 & \mathcal{F}_{V,\perp} \arrow[uu, shift left, "L_V"] \arrow[ur, shift left, "s_V"] &
\end{tikzcd} \tag{5.15}$$

The exposed face lattice of sr_V is $\mathcal{F}_{V,\perp} = \{F_{V,\perp}(v) : v \in V \setminus \{0\}\} \cup \{\emptyset, \mathrm{sr}_V\}$. By the above diagram the exposed projector lattice of sr_V is

$$\mathcal{P}_{V,\perp} = p_+(V) \cup \{0\} \tag{5.16}$$

with the maximal projector $p_+(v)$ of a vector $v \in V$. Notice that $p_+(0) = \mathbb{1}$.

(d) We can generate the shape of sr_V using only (5.16) and (5.15). By Remark 3.15 (f) the relative boundary of the state reflection sr_V is the union of proper exposed faces.

(e) The reduction of a convex set to the relative interior commutes with the application of an affine mapping (3.15). Hence one has for an arbitrary projector $p \in \mathcal{P}$ the equality

$$\mathrm{ri}(\mathbb{F}_V(p)) = \pi_V(\mathrm{ri}(\mathbb{F}(p))). \tag{5.17}$$

In particular, for $p = \mathbb{1}$ follows $\mathrm{ri}(\mathrm{sr}_U) = \pi_V(S(A))$ with the space $S(A)$ of invertible density matrices.

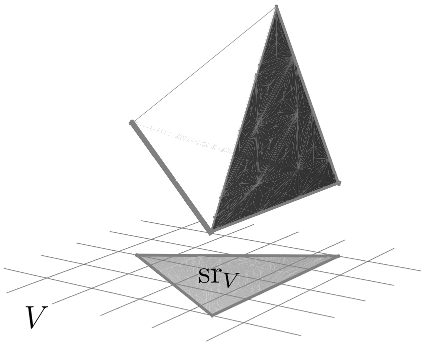


Figure 5.1: The drawing shows a tetrahedron, the state space $\overline{S}(\mathbb{C}^4)$, and a triangle in the plane V, the state reflection sr_V. To fit into a three dimensional configuration, $\overline{S}(\mathbb{C}^4)$ is rearranged perpendicular to V. The dark triangular face of $\overline{S}(\mathbb{C}^4)$ and the gray segment emanating to the left belong to the lifted exposed face lattice of sr_V. The coordinate lines in V relate to its basis vectors $(1,-1,0,0)$ and $(1,1,1,-3)$.

Example 5.6 (An abelian shape) We consider the state space $\overline{S}(\mathbb{C}^4)$ which is the probability simplex spanned by the vectors $\delta_1 = (1,0,0,0)$, $\delta_2 = (0,1,0,0)$, $\delta_3 = (0,0,1,0)$ and $\delta_4 = (0,0,0,1)$, corresponding to the Dirac measures on $\{1,2,3,4\}$, see Example 3.2 on page 51. We discuss the state reflection on $V = \mathrm{Lin}_{\mathbb{R}}\{u,v\}$ for $u := (1,-1,0,0)$ and $v := (1,1,1,-3)$. Here the vectors u and v are chosen perpendicular to $\delta_3 - \frac{1}{2}(\delta_1+\delta_2)$ in $\mathrm{lin}\{\delta_1,\delta_2,\delta_3\}$, see Figure 5.1 for a drawing. Since the state space $\overline{S}(\mathbb{C}^4)$ is a simplex it follows that the state reflection is a polytope, see Ziegler [Zi]. All faces of sr_V are exposed.

We use (5.16) and determine the exposed projector lattice by calculation of maximal projectors, $\mathcal{P}_{V,\perp} = p_+(V) \cup \{0\}$. In this abelian case the eigenvalues of a vector in V are directly at our disposal. For $\alpha \in [0,2\pi)$ we have to find the maximal coefficients of the vector

$$v(\alpha) := u\cos(\alpha) + v\sin(\alpha) = (\sqrt{2}\cos(\alpha - \frac{\pi}{4}),\ \sqrt{2}\sin(\alpha - \frac{\pi}{4}),\ \sin(\alpha),\ -3\sin(\alpha)).$$

The result for the exposed projector lattice of V is

$$\mathcal{P}_{V,\perp} = \{\delta_1,\ \delta_2,\ \delta_4,\ \delta_1+\delta_4,\ \delta_2+\delta_4,\ \mathbb{1}-\delta_4\} \quad \cup \ \{0,\ \mathbb{1}\}.$$

The lifted exposed face lattice consists of the faces $\mathbb{F}(p) = \kappa^p\left(\overline{S}(A^p)\right)$ for $p \in \mathcal{P}_{V,\perp}$. These are the empty set, three extreme points δ_1, δ_2, δ_4, two segments $[\delta_1,\ \delta_4]$ and $[\delta_2,\ \delta_4]$, the triangle $\mathrm{conv}\{\delta_1,\ \delta_2,\ \delta_3\}$ and the state space $\overline{S}(\mathbb{C}^4)$, which is a tetrahedron.

Example 5.7 (Walk through a conic frustum) We study two-dimensional state reflections for the algebra $A := M_2 \oplus \mathbb{C}$. The vector space A_{sa}^0 is the orthogonal direct sum of the vector spaces

$$U := \mathrm{Lin}\,\{\sigma_1 \oplus 0,\ \ \sigma_2 \oplus 0,\ \ \sigma_3 \oplus 0\} \qquad \text{and} \qquad \mathrm{Lin}\{z\} \quad \text{for} \quad z := (-\frac{\mathbb{1}_2}{2}) \oplus 1.$$

Here 0_2 and $\mathbb{1}_2$ denote the zero element and the identity in M_2 and $\widehat{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ is the vector of Pauli matrices $\sigma_1, \sigma_2, \sigma_3$. The vector z connects the centroid $\frac{\mathbb{1}_2}{2} \oplus 0$ to the pure state $0_2 \oplus 1$. On the joining segment lies the trace state

$$\frac{\mathbb{1}}{3} \;=\; \frac{2}{3}\left(\frac{\mathbb{1}_2}{2} \oplus 0\right) \;+\; \frac{1}{3}\,(0_2 \oplus 1)\,.$$

The three points $\frac{\mathbb{1}_2}{2} \oplus 0$, $\frac{\mathbb{1}}{3}$ and $0_2 \oplus 1$ belong to the plane $\mathrm{Lin}\{z, \mathbb{1}\}$. The four-dimensional state space is the convex hull

$$\overline{S}(A) = \overline{S}(M_2 \oplus \mathbb{C}) = \mathrm{conv}\left(\overline{S}(M_2) \oplus 0,\ \ 0_2 \oplus 1\right)$$

by Lemma 4.8 with the three-dimensional Bloch ball (4.5)

$$\overline{S}(M_2) = \left\{\frac{1}{2}(\mathbb{1}_2 + a\widehat{\sigma}) :\ \ |a| \leq 1,\ \ a \in \mathbb{R}^3\right\}$$

introduced in Example 4.6. Since $0_2 \oplus 1 \notin \mathrm{aff}\left(\overline{S}(M_2) \oplus 0\right) = \frac{\mathbb{1}_2}{2} \oplus 0 + \mathrm{U}$, the state space $\overline{S}(A)$ is a frustum (truncated cone) with base $\overline{S}(M_2) \oplus 0$ and apex $0_2 \oplus 1$.

The automorphisms of A are given by unitary conjugation of the first summand M_2, they act on $\overline{S}(A)$ by rotation of the Bloch ball $\overline{S}(M_2)$. The Grassmannian manifold $G_{4,2}$ of two-dimensional subspaces $V \subset A_{\mathrm{sa}}^0$ being of dimension 4, the group $\mathrm{SO}(3)$ acts by (trace preserving) algebra automorphisms on it,

$$\phi : \mathrm{SO}(3) \times G_{4,2} \to G_{4,2}.$$

The orbits are parametrized by the angle $\varphi \in [0, \frac{\pi}{2}]$ between V and z. State reflections in the same orbit are isometric, $\mathrm{sr}_{\phi(\xi,V)} = \xi(\mathrm{sr}_V)$ holds for an algebra automorphism $\xi \in \mathrm{SO}(3)$.

(a) Angular representation. Let $V \subset A_{\mathrm{sa}}^0$ be a two-dimensional vector space. The minimum (non-negatively taken) angle between z and a vector in V,

$$\varphi := \angle(z, V)$$

belongs to $[0, \frac{\pi}{2}]$. We have $\varphi = 0 \iff z \in V$ and $\varphi = \frac{\pi}{2} \iff V \subset U$. Other values of special interest are discussed below in (e) and (f):

- $\varphi = \operatorname{arccot}(\sqrt{\frac{2}{3}})$ $\approx 0.28\pi$ is the simplest algebraic instance of a state reflection with non-exposed faces and
- $\varphi = \frac{\pi}{3}$ has a segment of the frustum barrel perpendicular to V.

We introduce a basis v_1, v_2 for V with φ as a parameter. Since the sum of $\dim(U) = 3$ and $\dim(V) = 2$ is larger than $\dim(A^0_{\mathrm{sa}}) = 4$ there is a non-zero $a \in \mathbb{R}^3$ with

$$v_1 := a\widehat{\sigma} \oplus 0 \in V.$$

A second basis vector of V takes the from

$$\widetilde{v}_2 := \lambda b\widehat{\sigma} \oplus 0 + \mu z \in V$$

for some $b \in \mathbb{R}^3$ and $\lambda, \mu \in \mathbb{R}$. At the expense of λ we choose $\mu \geq 0$ and at the expense of b we choose $\lambda \geq 0$. Under the assumption $a \perp b$ we have $\varphi = \angle(z, V) = \angle(z, \widetilde{v}_2)$. We assume $a \perp b$ adding multiples of a to b if necessary. Then

$$\cos(\varphi) = \cos(\angle(z, \widetilde{v}_2)) = \frac{\mu \|z\|_2}{\|\widetilde{v}_2\|_2} = \frac{\mu}{\sqrt{(\frac{2}{\sqrt{3}}\lambda|b|)^2 + \mu^2}}.$$

The non-degenerate case $\varphi > 0$ implies $\lambda > 0$ and $|b| > 0$. Here we have

$$\cot(\varphi) = \frac{\sqrt{3}}{2} \frac{\mu}{\lambda|b|}. \tag{5.18}$$

This gives with $v_2 := \lambda^{-1}\widetilde{v}_2$ the basis of V

$$v_1 = a\widehat{\sigma} \oplus 0 \qquad \text{and} \qquad v_2 = b\widehat{\sigma} \oplus 0 + \frac{2}{\sqrt{3}}|b| \cot(\varphi) z. \tag{5.19}$$

Conversely, for each angle $\varphi \in (0, \frac{\pi}{2}]$ and two non-zero and perpendicular vectors $a, b \in \mathbb{R}^3$ the vectors v_1, v_2 in (5.19) span a two-dimensional vector space V with angle $\angle(z, V) = \varphi$.

For $\varphi > 0$ the vectors v_1, v_2 in (5.19) are adopted to the spectral analysis in (d). We use the substitution $g := a + b$ and $h := b - a$ and we take the vectors g and h orthonormal. Then

$$v_1 = \frac{1}{2}(g - h)\widehat{\sigma} \oplus 0 \qquad \text{and} \qquad v_2 = \frac{1}{2}(g + h)\widehat{\sigma} \oplus 0 + \sqrt{\frac{2}{3}} \cot(\varphi) z. \tag{5.20}$$

The moduli are $\|v_1\|_2 = 1$ and $\|v_2\|_2 = \frac{1}{\sin(\varphi)}$. For graphical issues let us complete the basis $\{v_1, v_2\}$ of V to an orthogonal basis of $V + \mathbb{R}z$ by a vector v_3 of modulus $\frac{1}{\sin(\varphi)}$

$$v_3 := -\frac{1}{2}\cot(\varphi)(g+h)\widehat{\sigma} \oplus 0 + \sqrt{\frac{2}{3}}z. \tag{5.21}$$

(b) Reduced dimension. Since $\dim(V) = 2$ the state reflection sr_V has an adequate description in dimension two or three. Let $W \subset U = \mathrm{Lin}\{\sigma_1 \oplus 0, \sigma_2 \oplus 0, \sigma_3 \oplus 0\}$ be a vector space of dimension $\dim(W) \geq 1$. Based on the ball $K := (\overline{S}(M_2) \oplus 0) \cap (\frac{\mathbb{1}_2}{2} \oplus 0 + W)$ of radius $\frac{1}{\sqrt{2}}$ with translation vector space $\mathrm{lin}(K) = W$ we consider the frustum

$$C := \mathrm{conv}(K,\ 0_2 \oplus 1) \tag{5.22}$$

having affine hull $\mathrm{aff}(C) = \frac{\mathbb{1}}{3} + W + \mathbb{R}z$ and translation vector space $\mathrm{lin}(C) = W + \mathbb{R}z$. The frustum C is the intersection of its affine hull with the state space.

$$\begin{array}{rcl} \overline{S}(A) \cap \mathrm{aff}(C) & = & \mathrm{conv}\left(\overline{S}(M_2) \oplus 0,\ 0_2 \oplus 1\right) \cap \left(\frac{\mathbb{1}}{3} + W + \mathbb{R}z\right) \\ & = & \frac{1}{3}(\mathbb{1} - z) + \left[\mathrm{conv}\left(\overline{S}(M_2) \oplus 0 - \frac{\mathbb{1}_2}{2} \oplus 0,\ z\right)\ \cap\ (W + \mathbb{R}z)\right] \\ & = & \frac{1}{3}(\mathbb{1} - z) + \mathrm{conv}\left[(\overline{S}(M_2) \oplus 0 - \frac{\mathbb{1}_2}{2} \oplus 0) \cap W,\ z\right] \\ & = & \frac{1}{3}(\mathbb{1} - z) + \mathrm{conv}\left(K - \frac{\mathbb{1}_2}{2} \oplus 0,\ z\right) \\ & = & \mathrm{conv}(K,\ 0_2 \oplus 1) = C. \end{array} \tag{5.23}$$

The third equality follows from the direct sum $U + \mathbb{R}z$, we notice $\overline{S}(M_2) \oplus 0 - \frac{\mathbb{1}_2}{2} \oplus 0 \subset U$.

We show that C is isometric to a state reflection $C = \pi_{\mathrm{aff}(C)}(\overline{S}(A))$. By symmetry of the Bloch ball $\overline{S}(M_2) \oplus 0$ the identity $\pi_W(K) = \pi_W(\overline{S}(M_2) \oplus 0)$ holds and one has $\pi_{\mathbb{R}z}(K) = \pi_{\mathbb{R}z}(\overline{S}(M_2) \oplus 0) = \{-\frac{1}{3}z\}$. This shows that $\pi_{\mathrm{lin}(C)}(K) = \pi_{\mathrm{lin}(C)}(\overline{S}(M_2) \oplus 0)$. Then

$$\begin{array}{rl} \pi_{\mathrm{lin}(C)}(C) = & \pi_{\mathrm{lin}(C)}\left(\mathrm{conv}\left(K,\ 0_1 \oplus 1\right)\right) \\ = & \mathrm{conv}\left(\pi_{\mathrm{lin}(C)}(K),\ \pi_{\mathrm{lin}(C)}(0_2 \oplus 1)\right) \\ = & \mathrm{conv}\left(\pi_{\mathrm{lin}(C)}(\overline{S}(M_2) \oplus 0),\ \pi_{\mathrm{lin}(C)}(0_2 \oplus 1)\right) \\ = & \pi_{\mathrm{lin}(C)}\left(\mathrm{conv}(\overline{S}(M_2) \oplus 0,\ 0_2 \oplus 1)\right) \\ = & \pi_{\mathrm{lin}(C)}\left(\overline{S}(A)\right). \end{array} \tag{5.24}$$

With (5.24) a reduced dimension in the discussion of the state reflection sr_V is obtained provided that $V \subset \mathrm{lin}(C)$. Let $V \subset A^0_{\mathrm{sa}}$ be an arbitrary two-dimensional subspace and take $W \subset U$ such that

$$\pi_U(V) \subset W. \tag{5.25}$$

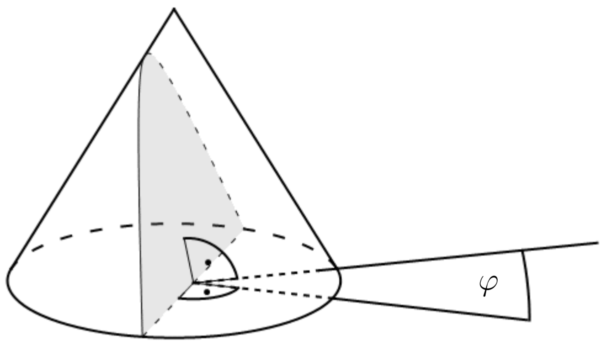


Figure 5.2: The picture shows a conic frustum C associated to a two-dimensional vector space $V \subset (M_2 \oplus \mathbb{C})^0_{\mathrm{sa}}$. The horizontal plane W is a two-dimensional subspace of $U = \mathrm{Lin}\{\sigma_1, \sigma_2, \sigma_3\} \oplus 0$ (with Pauli matrices $\sigma_1, \sigma_2, \sigma_3$) such that $\pi_U(V) \subset W$. The frustum is the intersection $C = \overline{S}(M_2 \oplus \mathbb{C}) \cap (\frac{\mathbb{1}}{3} + W + \mathbb{R}z)$ with vertical symmetry axis $z = (-\frac{\mathbb{1}_2}{2}) \oplus 1$ and angular aperture $\frac{\pi}{6}$. One has $V \subset \mathrm{lin}(C)$ and the section $C \cap (\frac{\mathbb{1}}{3} + V)$ is drawn in gray. The main issue is $\mathrm{sr}_V = \pi_V(C)$ for the orthogonal projection π_V to V. The angle $\varphi = \angle(z, V)$ appears between W and the kernel $V^\perp \cap \mathrm{lin}(C)$ of π_V. Details are proved in Example 5.7 (a) and (b).

The inclusion $W^\perp \cap U \subset (\pi_U(V))^\perp \cap U = V^\perp \cap U$ gives $V + U^\perp \subset W + U^\perp$. In intersection with the space of traceless matrices this is

$$V + \mathbb{R}z \subset \mathrm{lin}(C). \tag{5.26}$$

The choice $W := \pi_U(V)$ in (5.25) gives $V + \mathbb{R}z = \mathrm{lin}(C)$. In general one has by (5.24)

$$\mathrm{sr}_V = \pi_V(C). \tag{5.27}$$

The plane case $\dim(C) = 2$ is discussed in (c). If $\dim(C) = 3$ then K is a disk of diameter $\sqrt{2}$ and $C = \mathrm{conv}(K,\ 0_2 \oplus 1)$ is a rotationally symmetric three-dimensional frustum on the base disk K and with barrel length $\sqrt{2}$, see Figure 5.2. By projection to V the disk K projects to an ellipse $\pi_V(K)$, the degenerate case of a segment included. By (5.27) the state reflection sr_V is the convex hull of the ellipse $\pi_V(K)$ and of the point $\pi_V(0_2 \oplus 1)$.

For $\dim(C) = 3$ we want to introduce a coordinate system for C adopted to V. Thus we use the ONB $\{v_1, v_2 \sin(\varphi), v_3 \sin(\varphi)\}$ from (5.20) and (5.21) for an orthonormal pair $g, h \in \mathbb{R}^3$. If $\varphi > 0$ then $\{v_1, v_2 \sin(\varphi)\}$ is an ONB for V and $\pi_U(V) = \mathrm{Lin}\{g, h\}\widehat{\sigma} \oplus 0$. Otherwise for $\varphi = 0$ the space $\pi_U(V)$ has dimension one. Still we can use the above basis by reducing the cotangent in v_2 and v_3 with the sine and $\{v_1, v_2 \sin(\varphi)\}$ is an ONB for V also for $\varphi = 0$. In either case we put $W := \mathrm{Lin}\{g, h\}\widehat{\sigma} \oplus 0 \subset U$ and have then $\pi_U(V) \subset W$. For normalized $a \in \mathbb{R}^3$ there is a rank-one projector $p_\pm(a\widehat{\sigma}) = \frac{1}{2}(\mathbb{1}_2 \pm a\widehat{\sigma})$. We parametrize the boundary circle of $K = (\overline{S}(M_2) \oplus 0) \cap (\frac{\mathbb{1}_2}{2} + W)$ using the curve

$$\mathbb{R} \to \mathbb{R}^3, \qquad \alpha \mapsto c(\alpha) := g\cos(\alpha) + h\sin(\alpha) \tag{5.28}$$

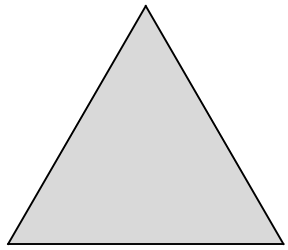


Figure 5.3: The state reflection sr_V on V for $\varphi = 0$ in Example 5.7 (c) is depicted, sr_V is an equilateral triangle.

and we obtain the actual points on the boundary circle of K by

$$\mathbb{R} \to K, \qquad \alpha \mapsto p_+(c(\alpha)\widehat{\sigma}) \oplus 0.$$

The frustum $C = \mathrm{conv}(K,\ 0_2 \oplus 1)$ has translation vector space $\mathrm{lin}(C) = W + \mathbb{R}z$ (5.22) and $\mathrm{Lin}\{v_1, v_2 \sin(\varphi), v_3 \sin(\varphi)\} = W + \mathbb{R}z$. Thus $\{v_1, v_2 \sin(\varphi), v_3 \sin(\varphi)\}$ can be used as a coordinate system for C. The coordinates of the apex of C and of the base circle of C are for $\alpha \in \mathbb{R}$

$$\begin{array}{rrcl} & \langle v_1,\ 0_2 \oplus 1\rangle & = & 0, \\ & \langle v_2 \sin(\varphi),\ 0_2 \oplus 1\rangle & = & \sqrt{\tfrac{2}{3}}\cos(\varphi), \\ & \langle v_3 \sin(\varphi),\ 0_2 \oplus 1\rangle & = & \sqrt{\tfrac{2}{3}}\sin(\varphi), \\ & \langle v_1,\ p_+(c(\alpha)\widehat{\sigma}) \oplus 0\rangle & = & -\tfrac{1}{\sqrt{2}}\sin(\alpha - \tfrac{\pi}{4}), \\ & \langle v_2 \sin(\varphi),\ p_+(c(\alpha)\widehat{\sigma}) \oplus 0\rangle & = & \tfrac{1}{\sqrt{6}}\left(\sqrt{3}\sin(\varphi)\cos(\alpha - \tfrac{\pi}{4}) - \cos(\varphi)\right), \\ \text{and} & \langle v_3 \sin(\varphi),\ p_+(c(\alpha)\widehat{\sigma}) \oplus 0\rangle & = & -\tfrac{1}{\sqrt{6}}\left(\sqrt{3}\cos(\varphi)\cos(\alpha - \tfrac{\pi}{4}) + \sin(\varphi)\right). \end{array} \tag{5.29}$$

(c) The degenerate case $\varphi = 0$. Here we have $V = \mathrm{Lin}\{z,\ a\widehat{\sigma} \oplus 0\}$ for some $a \in \mathbb{R}^3$. We assume that $|a| = 1$ is a unit vector. For the dimension reduction started in (b) we use the one-dimensional space $W := \mathbb{R}a\widehat{\sigma} \oplus 0$. Then K is a segment of length $\sqrt{2}$ and C is an equilateral triangle of edge length $\sqrt{2}$. Here $V = \mathrm{lin}(C)$ so the state reflection sr_V is isometric to C. See Figure 5.3 for a graphic. We go the formal way and calculate the exposed projector lattice $\mathcal{P}_{V,\perp}$ of V with (5.16) by determination of maximal projectors. This is sufficient to be done for the matrices on an ellipse about the origin in the linear span of $0_2 \oplus 1$ and $a\widehat{\sigma} \oplus 0$. For $\alpha \in [0, 2\pi)$ the spectral decomposition is

$$\begin{aligned} &(0_2 \oplus 1)\cos(\alpha) + (a\widehat{\sigma} \oplus 0)\sin(\alpha) \\ &= \sin(\alpha)(p_+(a\widehat{\sigma}) \oplus 0) - \sin(\alpha)(p_-(a\widehat{\sigma}) \oplus 0) + \cos(\alpha)(0_2 \oplus 1) \end{aligned}$$

with rank-one projectors $p_\pm(a\widehat{\sigma}) = \frac{1}{2}(\mathbb{1}_2 \pm a\widehat{\sigma})$. The exposed projector lattice of V is

$$\mathcal{P}_{V,\perp} = \left\{ \begin{array}{l} 0,\ p_+(a\widehat{\sigma}) \oplus 0,\ p_-(a\widehat{\sigma}) \oplus 0,\ 0_2 \oplus 1, \\ p_+(a\widehat{\sigma}) \oplus 1,\ p_-(a\widehat{\sigma}) \oplus 1,\ \mathbb{1}_2 \oplus 0,\ \mathbb{1} \end{array} \right\}.$$

The lifted exposed face lattice of sr_V is by (5.15) the set

$$\mathcal{L}_{V,\perp} = \left\{ \begin{array}{l} \emptyset,\ \{p_+(a\widehat{\sigma}) \oplus 0\},\ \{p_-(a\widehat{\sigma}) \oplus 0\},\ \{0_2 \oplus 1\}, \\ [p_+(a\widehat{\sigma}) \oplus 0,\ 0_2 \oplus 1],\ [p_-(a\widehat{\sigma}) \oplus 0,\ 0_2 \oplus 1], \\ \overline{S}(M_2) \oplus 0,\ \overline{S}(A) \end{array} \right\}. \tag{5.30}$$

The state space face $\mathbb{F}(p) = \kappa^p(A^p)$ of the rank-two projector $p = p_+(a\widehat{\sigma}) \oplus 1$ is a segment because $A^p = \mathbb{C}^2$ is abelian. The analogue is true for $p_-(a\widehat{\sigma}) \oplus 1$. For $p = \mathbb{1}_2 \oplus 0$ the face $\mathbb{F}(p)$ is the Bloch ball $\overline{S}(M_2) \oplus 0$.

(d) Spectral analysis in the region $\varphi > 0$. We modify the basis representation (5.20) of v_1, v_2 and we skip to the symmetrization $v_1 + v_2$ and $v_2 - v_1$ with the matrix $\frac{\cot(\varphi)}{\sqrt{6}}\mathbb{1}$ added. One obtains

$$w_1 := g\widehat{\sigma} \oplus \sqrt{\frac{3}{2}}\cot(\varphi) \qquad \text{and} \qquad w_2 := h\widehat{\sigma} \oplus \sqrt{\frac{3}{2}}\cot(\varphi). \tag{5.31}$$

For detection of maximal projectors, the vectors w_1, w_2 are as good as any other representation of V or of V crooked by the identity $\mathbb{1}$. For $\alpha \in \mathbb{R}$ we put

$$f(\alpha) := \sqrt{3}\cot(\varphi)\cos(\alpha - \frac{\pi}{4}).$$

Then we get the spectral decomposition

$$\begin{aligned} w(\alpha) &:= w_1\cos(\alpha) + w_2\sin(\alpha) \\ &= \ p_+\left(c(\alpha)\widehat{\sigma}\right) \oplus 0 \ - \ p_-\left(c(\alpha)\widehat{\sigma}\right) \oplus 0 \ + \ f(\alpha)(0_2 \oplus 1). \end{aligned} \tag{5.32}$$

This allows to determine the maximal projector of $w(\alpha)$ by maximization of $f(\alpha)$ against one for each $\alpha \in [0, 2\pi)$. In the following parts we distinguish the three regions

(e) $0 < \varphi < \frac{\pi}{3}$,
(f) $\varphi = \frac{\pi}{3}$,
(g) $\frac{\pi}{3} < \varphi \leq \frac{\pi}{2}$.

The state reflection in case (e) has an apex which is a shadow of the apex $0_2 \oplus 1$ of the

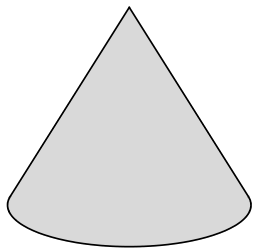


Figure 5.4: The state reflection for $\varphi = \frac{\pi}{6}$ in Example 5.7 (e) is depicted. A crossing from the arc to each of the two segments is a non-exposed extreme point.

state space, decreasing in distinctness with increasing φ and vanishing at (f) where a pure ellipse appears as state reflection. In (g) the elliptical shape turns into a disk at $\varphi = \frac{\pi}{2}$.

(e) The region $0 < \varphi < \frac{\pi}{3}$**.** Most remarkable for this region are two non-exposed extreme points of sr_V. Observe that $0 < \tan(\varphi) < \sqrt{3}$. Thus

$$\arccos\left(\frac{\tan(\varphi)}{\sqrt{3}}\right) \in (0, \frac{\pi}{2})$$

and the distinct angles

$$\alpha_\pm := \frac{\pi}{4} \pm \arccos\left(\frac{\tan(\varphi)}{\sqrt{3}}\right)$$

satisfy $\cos(\alpha_\pm - \frac{\pi}{4}) = \frac{\tan(\varphi)}{\sqrt{3}}$. Using $f(\alpha) = \sqrt{3}\cot(\varphi)\cos(\alpha - \frac{\pi}{4})$ we find modulo 2π for $\alpha \in \mathbb{R}$ that

- $f(\alpha) < 1 \iff \alpha \in (\alpha_+,\ 2\pi + \alpha_-)$,
- $f(\alpha) = 1 \iff \alpha \in \{\alpha_+,\ \alpha_-\}$,
- $f(\alpha) > 1 \iff \alpha \in (\alpha_-,\ \alpha_+)$.

The exposed projector lattice of V is derived from the spectral decomposition (5.32),

$$\begin{aligned} \mathcal{P}_{V,\perp} \;=\; & \{0,\ 0_2 \oplus 1,\ p_+(c(\alpha_-)\widehat{\sigma}) \oplus 1,\ p_+(c(\alpha_+)\widehat{\sigma}) \oplus 1,\ \mathbb{1}\} \\ & \cup\ \{p_+(c(\alpha)\widehat{\sigma}) \oplus 0\ :\ \alpha \in (\alpha_+,\ 2\pi + \alpha_-)\}\,. \end{aligned} \tag{5.33}$$

For $p := p_+(c(\alpha_\pm)\widehat{\sigma}) \oplus 1$ the algebra A^p is abelian, hence the state space face $\mathbb{F}(p) = \kappa^p(A^p)$ is a segment. The lifted exposed face lattice $\mathcal{L}_{V,\perp}$ of the state reflection sr_V is obtained with (5.15). It contains the empty set $\emptyset$, an extreme point for each rank one projector in $\mathcal{P}_{V,\perp}$, the segments

$$s_- := [p_+(c(\alpha_-)\widehat{\sigma}) \oplus 0,\ 0_2 \oplus 1] \qquad \text{and} \qquad s_+ := [p_+(c(\alpha_+)\widehat{\sigma}) \oplus 0,\ 0_2 \oplus 1] \tag{5.34}$$

and the frustum $\overline{S}(A)$. For the drawing of an example see Figure 5.4. Let us denote $\rho_- := p_+\left(c(\alpha_-)\widehat{\sigma}\right) \oplus 0$ and $\rho_+ := p_+\left(c(\alpha_+)\widehat{\sigma}\right) \oplus 0$ for the endpoints of the segments $s_\pm$.

We prove that $\pi_V(\rho_-)$ and $\pi(\rho_+)$ are non-exposed points of sr_V using a normal cone argument. Every vector $u \in A_{\mathrm{sa}} \setminus \{0\}$ is acute normal (3.52) for the state space by Corollary 4.25, that is, $\rho \in \mathrm{ri}(F_\perp(u))$ implies $u \in \mathrm{ri}(\mathrm{N}(\rho))$. Hence by Lemma 3.43 every vector $v \in V \setminus \{0\}$ is acute normal for the state reflection sr_V. Now we can use Remark 3.44 (a): a face F of sr_V is non-exposed if and only if $F = \{x\}$ where x is the endpoint of some one-dimensional face of sr_V but x is not the endpoint of two distinct one-dimensional faces of sr_V.

By the arguments in the previous paragraph the points $\pi_V(\rho_\pm)$ are the only candidates of non-exposed extreme points of sr_V. They are non-exposed, because the segments s_- and s_+ have distinct one-dimensional face reflections $\pi_V(s_-)$ and $\pi_V(s_+)$, as sr_V has non-empty interior in V. The face lattice $\mathcal{F}(\mathrm{sr}_V)$ of sr_V is the disjoint union

$$\mathcal{F}(\mathrm{sr}_V) \;=\; \mathcal{F}_{V,\perp} \;\cup\; \{\{\pi_V(\rho_-)\},\; \{\pi_V(\rho_+)\}\}\,. \tag{5.35}$$

A simple algebra appears at an angle of $\varphi = \mathrm{arccot}(\sqrt{\tfrac{2}{3}}) \approx 0.28\pi$ for $g\widehat{\sigma} = \sigma_1$ and $h\widehat{\sigma} = \sigma_2$. The basis v_1, v_2 of V from (5.20) is then

$$v_1 = \frac{1}{2}(\sigma_1 - \sigma_2) \oplus 0 \qquad \text{and} \qquad v_2 = \frac{1}{2}(\sigma_1 + \sigma_2) \oplus 1 - \frac{\mathbb{1}}{3}$$

and the vectors satisfy

$$v_1 + v_2 = \sigma_1 \oplus 1 - \frac{\mathbb{1}}{3} \qquad \text{and} \qquad v_2 - v_1 = \sigma_2 \oplus 1 - \frac{\mathbb{1}}{3}. \tag{5.36}$$

The vectors for spectral analysis defined in (5.31) are $w_1 = \sigma_1 \oplus 1$ and $w_2 = \sigma_2 \oplus 1$. The state reflection for $\varphi = \mathrm{arccot}(\sqrt{\tfrac{2}{3}})$ is depicted in Figure 4.1 on page 79. For completeness, the exposed projector lattice is

$$\begin{aligned} \mathcal{P}_{V,\perp} \;=\; & \big\{0,\; 0_2 \oplus 1,\; p_+(\sigma_1) \oplus 1,\; p_+(\sigma_2) \oplus 1,\; \mathbb{1}\big\} \\ & \cup \big\{p_+(\sigma_1\cos(\alpha) + \sigma_2\sin(\alpha)) \oplus 0 \;:\; \alpha \in (\tfrac{\pi}{2}, 2\pi)\big\}. \end{aligned} \tag{5.37}$$

The two non-exposed points of the state reflection sr_V are

$$\pi_V(\frac{1}{2}(\mathbb{1}_2 + \sigma_1) \oplus 0) \qquad \text{and} \qquad \pi_V(\frac{1}{2}(\mathbb{1}_2 + \sigma_2) \oplus 0). \tag{5.38}$$

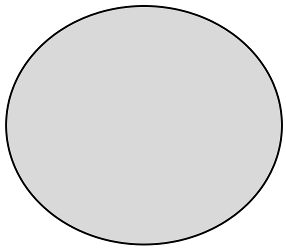


Figure 5.5: The state reflection for $\varphi = \frac{\pi}{3}$ in Example 5.7 (f).

(f) The case $\varphi = \frac{\pi}{3}$. Remarkable for this case is a one-dimensional face in the lifted exposed face lattice $\mathcal{L}_{V,\perp}$ that projects to a point whereas other proper faces in the lattice are extreme points of the state space frustum $\overline{S}(A)$. The one-dimensional face is parallel to $V^\perp$. One has $\cot(\varphi) = \frac{1}{\sqrt{3}}$ and the spectral decomposition (5.32) reveals that

$$\mathcal{P}_{V,\perp} \;=\; \Big\{0,\; p_+(c(\frac{\pi}{4})\widehat{\sigma}) \oplus 1,\; \mathbb{1}\Big\} \;\cup\; \Big\{p_+(c(\alpha)\widehat{\sigma}) \oplus 0 \;:\; \alpha \in [0, 2\pi) \setminus \{\frac{\pi}{4}\}\Big\}. \tag{5.39}$$

The algebra A^p is abelian for $p := p_+(c(\frac{\pi}{4})\widehat{\sigma}) \oplus 1$, hence the face $\mathbb{F}(p) = \kappa^p(A^p)$ is the segment

$$\Big[p_+(c(\frac{\pi}{4})\widehat{\sigma}) \oplus 0,\; 0_2 \oplus 1\Big]. \tag{5.40}$$

It follows from (5.29) that this segment projects to a point under π_V. The remaining faces in the lifted exposed face lattice (apart from $\emptyset$ and the frustum $\overline{S}(A)$) are the extreme points corresponding to elements of the pointed circle of rank-one projectors in $\mathcal{P}_{V,\perp}$:

$$\begin{aligned} \mathcal{L}_{V,\perp} \;=\; & \big\{\emptyset,\; \big[p_+(c(\tfrac{\pi}{4})\widehat{\sigma}) \oplus 0,\; 0_2 \oplus 1\big],\; \overline{S}(A)\big\} \\ & \cup\; \big\{\{p_+(c(\alpha)\widehat{\sigma}) \oplus 0\} \;:\; \alpha \in [0, 2\pi) \setminus \{\tfrac{\pi}{4}\}\big\}. \end{aligned} \tag{5.41}$$

See Figure 5.5 for the drawing of an example. A simple algebra is obtained for $g\widehat{\sigma} = \frac{\sigma_1 + \sigma_2}{\sqrt{2}}$ and $h\widehat{\sigma} = \frac{\sigma_2 - \sigma_1}{\sqrt{2}}$. Here the basis (5.20) of V consists of

$$v_1 = \frac{\sigma_1}{\sqrt{2}} \oplus 0 \qquad \text{and} \qquad v_2 = \frac{1}{\sqrt{2}}\left(\sigma_2 \oplus 1 - \frac{\mathbb{1}}{3}\right). \tag{5.42}$$

For $\alpha \in \mathbb{R}$ we have $c(\alpha)\widehat{\sigma} = \sigma_2 \cos(\alpha - \frac{\pi}{4}) - \sigma_1 \sin(\alpha - \frac{\pi}{4})$ and in particular $c(\frac{\pi}{4})\widehat{\sigma} = \sigma_2$. The exposed projector lattice is

$$\mathcal{P}_{V,\perp} \;=\; \{0,\; p_+(\sigma_2) \oplus 1,\; \mathbb{1}\} \;\cup\; \Big\{p_+(c(\alpha)\widehat{\sigma}) \oplus 0 \;:\; \alpha \in [0, 2\pi) \setminus \{\frac{\pi}{4}\}\Big\}. \tag{5.43}$$

The lifted exposed face lattice is

$$\begin{aligned} \mathcal{L}_{V,\perp} \;=\; & \big\{\emptyset,\; [p_+(\sigma_2) \oplus 0,\; 0_2 \oplus 1],\; \overline{S}(A)\big\} \\ & \cup\; \big\{\{p_+(c(\alpha)\widehat{\sigma}) \oplus 0\} \;:\; \alpha \in [0, 2\pi) \setminus \{\tfrac{\pi}{4}\}\big\}. \end{aligned} \tag{5.44}$$

(g) The region $\frac{\pi}{3} < \varphi \leq \frac{\pi}{2}$. Remarkable is that the lifted exposed face lattice consists of a circle of extreme points apart from the improper faces. One has $0 \leq \cot(\varphi) < \frac{1}{\sqrt{3}}$ and the spectral decomposition (5.32) reveals

$$\mathcal{P}_{V,\perp} \;=\; \{0,\ \mathbb{1}\} \;\cup\; \{p_+\left(c(\alpha)\widehat{\sigma}\right) \oplus 0 \;:\; \alpha \in [0, 2\pi)\}\,.$$

By (5.29) the state reflection is an ellipse with major axis $\sqrt{2}$ and minor axis $\sqrt{2}\sin(\varphi)$.

5.2 The face lattice

The face and projector lattice of the state reflection are extended from the exposed case to the general case. Recall that the face lattices of a convex set are complete.

Definition 5.8 The face lattice (3.22) of the state reflection sr_V is denoted by

$$\mathcal{F}_V := \mathcal{F}(\mathrm{sr}_V). \tag{5.45}$$

The face of a point $x \in \mathrm{sr}_V$ is denoted by

$$F_V(x) := F(\mathrm{sr}_V, x). \tag{5.46}$$

This is the unique face of sr_V which contains x in the relative interior (3.24). We use the state space lift (5.9) $L_V(M) = (M + V^{\perp}) \cap \overline{S}(A)$, $M \subset A_{\mathrm{sa}}$ and define the **lifted face lattice** of sr_V by

$$\mathcal{L}_V := \{L_V(F) : F \in \mathcal{F}_V\}. \tag{5.47}$$

Remark 5.9 **(a)** The state space lift induces by Proposition 3.38 a lattice isomorphism

$$L_V : \mathcal{F}_V \to \mathcal{L}_V \tag{5.48}$$

of complete lattices with inverse $\pi_V : \mathcal{L}_V \to \mathcal{F}_V$. The lifted face lattice $\mathcal{L}_V$ is a subset of the face lattice $\mathcal{F}$ of the state space.

(b) The state reflection sr_V is covered disjointly with relative interiors of faces $F \in \mathcal{F}_V$ by the stratification property of a convex set. These faces are orthogonal projections of faces from the lifted face lattice by (a). Thus

$$\mathrm{sr}_V = \bigcup_{F \in \mathcal{L}_V} \pi_V(\mathrm{ri}(F)) \tag{5.49}$$

is a disjoint union because a linear mapping commutes with reduction to the relative interior of a convex set (3.15).

Definition 5.10 The **projector lattice** of the state reflection sr_V is

$$\mathcal{P}_V := \{s(F) : F \in \mathcal{L}_V\}. \tag{5.50}$$

The **support projector** of a face $F \in \mathcal{F}_V$ is

$$s_V(F) := s \circ L_V(F). \tag{5.51}$$

This definition extends the support projector of an exposed face (5.12). The **support projector** of a point $x \in \mathrm{sr}_V$ is

$$s_V(x) := s_V(F_V(x)). \tag{5.52}$$

Remark 5.11 Allocation of the support projector $s|_{\mathcal{L}_V} : \mathcal{L}_V \to \mathcal{P}_V$ is a lattice isomorphism. It is the restriction of the support (4.39) of a state space face $s : \mathcal{F} \to \mathcal{P}$ to the lifted faces lattice $\mathcal{L}_V$. The inverse is the assignment of a face with a given support $\mathbb{F}|_{\mathcal{P}_V} : \mathcal{P}_V \to \mathcal{L}_V$ (4.34). In combination with the lattice isomorphism (5.48) $L_V : \mathcal{F}_V \to \mathcal{L}_V$ we get the isomorphism

$$s_V = s \circ L_V : \quad \mathcal{F}_V \to \mathcal{P}_V \tag{5.53}$$

providing the support of a face of sr_V and with inverse lattice isomorphism the association of a face reflection (5.13)

$$\mathbb{F}_V|_{\mathcal{P}_V} = \pi_V \circ \mathbb{F}|_{\mathcal{P}_V} : \quad \mathcal{P}_V \to \mathcal{F}_V. \tag{5.54}$$

In particular, for $p \in \mathcal{P}_V$ we have $s_V(\mathbb{F}_V(p)) = p$ and for $F \in \mathcal{F}_V$ we have $\mathbb{F}_V(s_V(F)) = F$.

Remark 5.12 Many of the upcoming statements are written for projectors rather than faces using the lattice isomorphism $\mathbb{F} : \mathcal{P} \to \mathcal{F}$, $p \mapsto \mathbb{F}(p)$ from (4.34). The support projector (5.52) of a point $x \in \mathrm{sr}_V$ satisfies

$$\mathbb{F}_V(s_V(x)) = \mathbb{F}_V(s_V(F_V(x))) = F_V(x). \tag{5.55}$$

The state reflection sr_V is covered by the (disjoint union of) relative interiors of face reflections

$$\mathrm{sr}_V = \bigcup_{p \in \mathcal{P}_V \setminus \{0\}} \mathrm{ri}(\mathbb{F}_V(p)). \tag{5.56}$$

This follows from the stratification of a convex set into relative interiors of faces. A convex subset F of $\overline{S}(A)$ projects to the relative boundary $\mathrm{rb}(\mathrm{sr}_V)$ if and only if $F \subset G$ for a proper face $G \in \mathcal{L}_V$. This is proved in Lemma 3.40. For a projector $p \in \mathcal{P}$ this is

$$\mathbb{F}_V(p) \subset \mathrm{rb}(\mathrm{sr}_V) \quad \iff \quad p \leq q \text{ for a proper } q \in \mathcal{P}_V. \tag{5.57}$$

For an individual state $\rho \in \overline{S}(A)$ we have with (4.41)

$$\pi_V(\rho) \in \mathrm{rb}(\mathrm{sr}_V) \quad \iff \quad s(\rho) \leq q \text{ for a proper } q \in \mathcal{P}_V. \tag{5.58}$$

Remark 5.13 (Structural summary for state reflections) We discuss the commuting diagram depicted in Figure 5.6 on page 116 except the arrow $V \setminus \{0\} \overset{\text{a. s.}}{\rightsquigarrow} \mathcal{P}_V$ to which the following section is dedicated. A good entrance to the diagram is the upper front triangle with the lattice isomorphism $s : \mathcal{F} \to \mathcal{P}$ defined on the face lattice of the state space and assigning the support projector (4.49). The triangle also includes the exposed face description for non-zero vectors in A^0_{sa}.

Further structure emerges by projection of the state space $\overline{S}(A)$ to V. The state reflection $\mathrm{sr}_V = \pi_V(\overline{S}(A))$ appears. The exposed face lattice $\mathcal{F}_{V,\perp}$ and the face lattice $\mathcal{F}_V$ of sr_V are located in the bottom row of the diagram. The state space lift L_V (5.48) acts as a lattice isomorphism on the face lattice $\mathcal{F}_V$ and associates a face of the state space to each face of sr_V. The lifted face lattice $\mathcal{L}_V = L_V(\mathcal{F}_V)$ appears. The exposed face lattice $\mathcal{F}_{V,\perp} \subset \mathcal{F}_V$ as a subset is carried by the same isomorphism L_V and gives raise to the lifted exposed face lattice $\mathcal{L}_{V,\perp} = L_V(\mathcal{F}_{V,\perp})$. In addition, a support projector is associated to a face $F \in \mathcal{F}_V$ by the formula

$$s_V(F) = s \circ L_V(F)$$

such that $s_V : \mathcal{F}_V \to \mathcal{P}_V$ is a lattice isomorphism (5.53). The inverse is the lattice isomorphism $\mathbb{F}_V|_{\mathcal{P}_V} : \mathcal{P}_V \to \mathcal{F}_V$ that associates a face reflection (5.54). Restriction to exposed faces gives the lattice isomorphism

$$s_V|_{\mathcal{F}_{V,\perp}} : \mathcal{F}_{V,\perp} \to \mathcal{P}_{V,\perp}$$

with inverse $\mathbb{F}_V|_{\mathcal{P}_{V,\perp}}$. In addition to the isomorphisms $\mathcal{F}_{V,\perp} \to \mathcal{P}_{V,\perp}$ and $\mathcal{F}_{V,\perp} \to \mathcal{L}_{V,\perp}$ there is a description of proper exposed faces by their exposing vector (5.15).

Remark 5.14 (Invariance under embedding) The state reflection sr_V is invariant under an algebra embedding. We prove below that for a direct sum matrix algebra $B \subset A$ and provided that $V \subset B$ one has

$$\pi_V(\overline{S}(A)) = \pi_V(\overline{S}(B)).$$

In principle we can treat the general case of a subalgebra $B \subset A$ by setting $\overline{S}(B) := \overline{S}(A) \cap B$ for the state space of B. However, we stay with the direct sum representation of algebras and do not enter the discussion.

For the case that $B = A^p \simeq pAp$ is a compression (2.27) of A for a projector $p \in \mathcal{P}$, we can prove slightly more. We use the *-isomorphism $\kappa^p : A^p \to pAp$ and obtain equivariance of all face reflections under κ^p in Lemma 5.17.

An invariance discussion for the lifted face lattices $\mathcal{L}_{V,\perp}$ and $\mathcal{L}_V$ and for the projector lattices $\mathcal{P}_{V,\perp}$ and $\mathcal{P}_V$ is postponed to Remark 5.35 in the following section where access sequences are available.

Proposition 5.15 If $B \subset A$ is a direct sum of full matrix algebras and if $V \subset B$ then $\mathrm{sr}_V = \pi_V(\overline{S}(B))$. [Proof on page 212]

We introduce extended notation for the development of access sequences in the following section and for the lemma below.

Remark 5.16 **(a)** For a projector $p \in \mathcal{P}$ we have defined in (2.27) the compression

$$A^p = M_{\mathrm{rk}(p_1)} \oplus \cdots \oplus M_{\mathrm{rk}(p_N)}.$$

We denote the projector lattice of A^p by $\mathcal{P}(A^p) = \{p \in (A^p) : p^2 = p^* = p\}$. The state space is denoted $\overline{S}(A^p)$. Given a linear subspace $W \subset (A^p)^0_{\mathrm{sa}}$ we write sr_W for the state reflection $\pi_W(\overline{S}(A^p))$ and we denote the lattices of sr_W as follows. We write $\mathcal{F}_W$ for the face lattice and $\mathcal{F}_{W,\perp}$ for the exposed face lattice. Moreover we apply the general notation $\mathcal{F}_\perp(C)$ for the exposed face lattice of a convex set $C \subset \mathbb{R}^m$. In particular $\mathcal{F}(\mathrm{sr}_V) = \mathcal{F}_V$ and $\mathcal{F}_\perp(\mathrm{sr}_V) = \mathcal{F}_{V,\perp}$.

The face of the state space $\overline{S}(A^p)$ with support $r \in \mathcal{P}(A^p)$ is denoted $\mathbb{F}(r)$. The face reflection on W with support r is denoted $\mathbb{F}_W(r)$. The omitted reference to the algebra A^p in the previous notation is intended. The corresponding projector is determined from the context. Notice that p can not be recovered from the matrix algebra A^p.

(b) In the analogue way as $\kappa^p : A^p \to pAp$ is defined in (2.28) we choose for a projector $r \in \mathcal{P}(A^p)$ a trace-preserving *-isomorphism

$$\kappa^r_{\mathbb{1}^p} : (A^p)^r \to r(A^p)r.$$

The composition $\kappa^p \circ \kappa^r_{\mathbb{1}^p}$ is a trace-preserving *-isomorphism, for $q := \kappa^p(r)$,

$$\kappa^p \circ \kappa^r_{\mathbb{1}^p} : (A^p)^r \to qAq.$$

Observe the equality of matrix algebras $(A^p)^r = A^q$. The isomorphism κ^q is another trace-preserving *-isomorphism from $(A^p)^r$ to qAq so modulo a trace-preserving *-automorphism of A^q we have $\kappa^p \circ \kappa^r_{\mathbb{1}^p} = \kappa^q$. This automorphism leaves the state space $\overline{S}((A^p)^r) = \overline{S}(A^q)$ invariant, so

$$\kappa^p(\mathbb{F}(r)) = \kappa^p\big(\kappa^r_{\mathbb{1}^p}(\overline{S}(A^q))\big) = \kappa^q(\overline{S}(A^q)) = \mathbb{F}(q). \qquad (5.59)$$

Lemma 5.17 Let $p \in \mathcal{P}$ be a non-zero projector, $V \subset \operatorname{lin}(\mathbb{F}(p))$ and put $W := (\kappa^p)^{-1}(V)$. For a projector $q \in \mathcal{P}(A^p)$ we have $\kappa^p(\mathbb{F}_W(q)) = \mathbb{F}_V(\kappa^p(q))$. Furthermore, $\kappa^p(\mathrm{sr}_W) = \mathrm{sr}_V$.
[Proof on page 213]

5.3 Geometry of access sequences

We use access sequences and characterize the projector lattice $\mathcal{P}_V$. This closes the gap in the structural summary of state reflections (Figure 5.6). The analysis is based on an affine transformation of a face reflection to a state reflection in a compressed algebra. We can show that $\mathcal{P}_V$ (except possibly the identity element in it) depends only on V and not on the algebra.

Definition 5.18 Recall for a non-zero projector $p \in \mathcal{P}$ the *-monomorphism $\kappa^p : A^p \to A$ (2.28). This embeds the compression $A^p \cong pAp$ into A. The **traceless compression** by p is the mapping

$$\varsigma^p : A_{\mathrm{sa}} \to (A^p)^0_{\mathrm{sa}}, \qquad \varsigma^p := (\kappa^p)^{-1} \circ \pi_{\operatorname{lin}(\mathbb{F}(p))} \qquad (5.60)$$

for the orthogonal projection $\pi_{\operatorname{lin}(\mathbb{F}(p))}$ to the translation vector space of the face $\mathbb{F}(p)$ described in (5.3).

Remark 5.19 For non-zero $p \in \mathcal{P}$ the vector space $\varsigma^p(V)$ has the same dimension as the face reflection $\mathbb{F}_V(p) = \pi_V(\mathbb{F}(p))$ defined in (5.13),

$$\dim(\varsigma^p(V)) = \dim(\pi_{\operatorname{lin}(\mathbb{F}(p))}(V)) = \dim(\pi_V(\operatorname{lin}(\mathbb{F}(p)))) = \dim(\mathbb{F}_V(p)). \qquad (5.61)$$

The reason is that for linear spaces $X, Y \subset \mathbb{R}^m$ with ONB $\{x_i\}_i$ of X and ONB $\{y_j\}_j$ of Y, the dimension of $\pi_X(Y)$ is the rank of the matrix $\langle x_i, y_j\rangle_{i,j}$. The rank is invariant under transposition of a matrix so $\pi_Y(X)$ has the same dimension as $\pi_X(Y)$.

Proposition 5.20 For $p \in \mathcal{P} \setminus \{0\}$ there is an affine isomorphism $\vartheta^p : \mathrm{aff}(\mathbb{F}_V(p)) \to \varsigma^p(V)$ such that the following diagrams commute.

$$
\begin{tikzcd}
(A^p)^1_{\mathrm{sa}} \arrow[r, shift left, "\kappa^p"] \arrow[d, "\pi_{\varsigma^p(V)}"'] & \mathrm{aff}(\mathbb{F}(p)) \arrow[l, shift left] \arrow[d, "\pi_V"] \\
\varsigma^p(V) \arrow[r, shift left] & \mathrm{aff}(\mathbb{F}_V(p)) \arrow[l, shift left, "\vartheta^p"]
\end{tikzcd}
\qquad
\begin{tikzcd}
\overline{S}(A^p) \arrow[r, shift left, "\kappa^p"] \arrow[d, "\pi_{\varsigma^p(V)}"'] & \mathbb{F}(p) \arrow[l, shift left] \arrow[d, "\pi_V"] \\
\mathrm{sr}_{\varsigma^p(V)} \arrow[r, shift left] & \mathbb{F}_V(p) \arrow[l, shift left, "\vartheta^p"]
\end{tikzcd}
\qquad
\begin{tikzcd}
S(A^p) \arrow[r, shift left, "\kappa^p"] \arrow[d, "\pi_{\varsigma^p(V)}"'] & \mathrm{ri}(\mathbb{F}(p)) \arrow[l, shift left] \arrow[d, "\pi_V"] \\
\mathrm{ri}(\mathrm{sr}_{\varsigma^p(V)}) \arrow[r, shift left] & \mathrm{ri}(\mathbb{F}_V(p)) \arrow[l, shift left, "\vartheta^p"]
\end{tikzcd}
$$

The map ϑ^p is expanding and it is isometric if and only if $\pi_{\mathrm{lin}(\mathbb{F}(p))}(V) \subset V$. The projections $\pi_{\varsigma^p(V)}$ and π_V are onto the given ranges in the diagrams. [Proof on page 213]

Remark 5.21 **(a)** The idea of Proposition 5.20 is to zoom in on a face reflection $\mathbb{F}_V(p)$ by transformation to the state reflection $\mathrm{sr}_{\varsigma^p(V)}$ on the traceless compression $\varsigma^p(V)$.

(b) The case $\pi_{\mathrm{lin}(F)}(V) \subsetneq V$ for a non-empty lifted face $F \in \mathcal{L}_V$ is not unusual, see Example 5.22. In fact, the metric properties of the affine isomorphism $\vartheta^{s(F)} : \pi_V(F) \to \mathrm{sr}_{\varsigma^{s(F)}(V)}$ in Proposition 5.20 are meaningless for the following analysis.

(c) The commuting diagram in Proposition 5.20 uses maximal domains. In a conceivable extension from $(A^p)^1_{\mathrm{sa}}$ to $(A^p)_{\mathrm{sa}}$ the identity $\mathbb{1}^p$ in A^p maps to zero under projection by $\pi_{\varsigma^p(V)}$. The counterpart $p = \kappa^p(\mathbb{1}^p)$ on the side of the algebra A will not map to zero under the projection by π_V unless for unimportant cases: if $\pi_V(p) = 0$ then

$$\pi_V(\frac{p}{\mathrm{tr}(p)}) = \pi_V(\frac{\mathbb{1}}{\mathrm{tr}(\mathbb{1})}) = 0.$$

If $p \neq \mathbb{1}$ then by the disjoint cover (5.56) of the state reflection by relative interiors of face reflections the projector p is not a member of $\mathcal{P}_V$.

Example 5.22 (Strict expansion) We consider the example in Figure 5.1 on page 97. The algebra is $A = \mathbb{C}^4$ and the state space $\overline{S}(\mathbb{C}^4)$ is the probability simplex spanned by the vectors $\delta_1 = (1,0,0,0)$, $\delta_2 = (0,1,0,0)$, $\delta_3 = (0,0,1,0)$ and $\delta_4 = (0,0,0,1)$ corresponding to the Dirac measures on $\{1,2,3,4\}$. The state reflection is examined on $V := \mathrm{Lin}_{\mathbb{R}}\{u, v\}$ for $u := \delta_1 - \delta_2$ and $v := \delta_1 + \delta_2 + \delta_3 - 3\delta_4$. The projector $p := \delta_1 + \delta_4$ is an exposed projector for sr_V because $4u - v = (3, -5, -1, 3)$. The corresponding lifted face is the segment $\mathbb{F}(p) = [\delta_1,\ \delta_4]$ with translation vector space

$$\mathrm{lin}(\mathbb{F}(p)) = \mathbb{R}(\delta_1 - \delta_4).$$

Since

$$\pi_V(\delta_1 - \delta_4) \;=\; \frac{1}{6}(5,-1,2,-6) \;\neq\; 0$$

we have $\pi_{\mathrm{lin}(\mathbb{F}(p))}(V) = \mathbb{R}(\delta_1 - \delta_4)$. This is an example where $\pi_{\mathrm{lin}(\mathbb{F}(p))}(V) \subsetneq V$, otherwise $\pi_V(\delta_1 - \delta_4)$ should be $\delta_1 - \delta_4$ or zero. The reason why $\pi_{\mathrm{lin}(\mathbb{F}(p))}(V) \subsetneq V$ is that the face $\mathbb{F}(p) = [\delta_1,\ \delta_4]$ has a slope with respect to V which disappears under compression. Indeed, by definition of ς^p in (5.60)

$$\varsigma^p(V) = (\kappa^p)^{-1} \circ \pi_{\mathrm{lin}(\mathbb{F}(p))}(V) = (\kappa^p)^{-1}(\mathbb{R}(\delta_1 - \delta_4)) = \mathbb{R}(1,-1)$$

holds. So $\varsigma^p(V) = (A^p)^0_{\mathrm{sa}} = \mathbb{R}(1,-1)$ and by Proposition 5.20 the projection $\pi_{\varsigma^p(V)}|_{(A^p)^1_{\mathrm{sa}}}$ is the translation $x \mapsto x - \frac{1}{2}(1,1)$, while the projection $\pi_{\varsigma^p(V)}|_{(A^p)^0_{\mathrm{sa}}}$ is the identical map. In particular,

$$\pi_{\varsigma^p(V)}\left((\kappa^p)^{-1}(\delta_1 - \delta_4)\right) = (1,-1)$$

has norm $\|(1,-1)\|_2 = \sqrt{2}$ while $\|\pi_V(\delta_1 - \delta_4)\|_2 = \sqrt{\frac{11}{6}}$. Since $\pi_{\varsigma^p(V)} \circ (\kappa^p)^{-1} = \vartheta^p \circ \pi_V$, the affine isomorphism ϑ^p is expanding by the factor $\sqrt{\frac{12}{11}}$.

For a face reflection the isomorphism ϑ^p is an explicit isometry.

Corollary 5.23 If $F \in \mathcal{F}$ is a non-empty face, $V := \mathrm{lin}(F)$ and $p \leq s(F)$ is a non-zero projector in $\mathcal{P}$ then $(\vartheta^p)^{-1} = \kappa^p - \frac{s(F)}{\mathrm{tr}(s(F))} + \frac{p}{\mathrm{tr}(p)}$. [Proof on page 215]

Remark 5.24 Proposition 5.20 is very general and ambiguous. Not all face reflections $\mathbb{F}_V(p)$ for $p \in \mathcal{P}$ are faces of sr_V. For instance, the face reflections $\mathbb{F}_V(1,0,1,0)$ and $\mathbb{F}_V(1,1,1,0)$ in Example 5.6 on page 97 have both dimension one but $\mathbb{F}_V(1,0,1,0) \subsetneq \mathbb{F}_V(1,1,1,0)$. The first task is to mediate between face lattices of $\mathbb{F}_V(p)$ and face lattices of $\mathrm{sr}_{\varsigma^p(V)}$.

Corollary 5.25 Let $p \in \mathcal{P} \setminus \{0\}$, $q \in \mathcal{P}$ such that $q \leq p$ and put $r := (\kappa^p)^{-1}(q)$. Then $\vartheta^p\left(\mathbb{F}_V(q)\right) = \mathbb{F}_{\varsigma^p(V)}(r)$ and $\mathbb{F}_V(q)$ is an (exposed) face of $\mathbb{F}_V(p)$ if and only if $\mathbb{F}_{\varsigma^p(V)}(r)$ is an (exposed) face of $\mathrm{sr}_{\varsigma^p(V)}$. [Proof on page 216]

Remark 5.26 Still, a face of sr_V can be the face reflection for various support projectors. In Example 5.6 one has $F := \mathbb{F}_V(1,1,1,0) = \mathbb{F}_V(1,1,0,0) \in \mathcal{F}_V$ and the support projector

(5.51) of F is $s_V(F) = (1,1,1,0)$. This ambiguity is removed by lifted faces. The second task is to mediate for a non-zero projector $p \in \mathcal{P}_V$ between the lifted faces in $\mathcal{L}_V$ that project to the face reflection $\mathbb{F}_V(p)$ and between the lifted face lattice $\mathcal{L}_{\varsigma^p(V)}$ of the state reflection $\mathrm{sr}_{\varsigma^p(V)}$. By Proposition 3.39 a face $F \in \mathcal{F}$ belongs to the lifted face lattice of sr_V if and only if $L_V(F) = F$, that is

$$F \in \mathcal{L}_V \quad \iff \quad (F + V^\perp) \cap \overline{S}(A) = F. \tag{5.62}$$

Indeed, the analysis will be done with projectors. We compare the projectors dominated by p in the lattice $\mathcal{P}_V$ with the projector lattice of $\mathrm{sr}_{\varsigma^p(V)}$. In terms of a projector $p \in \mathcal{P}$

$$p \in \mathcal{P}_V \quad \iff \quad (\mathbb{F}(p) + V^\perp) \cap \overline{S}(A) = \mathbb{F}(p) \tag{5.63}$$

holds by the lattice isomorphism $s : \mathcal{F} \to \mathcal{P}$ restricted to $\mathcal{L}_V \to \mathcal{P}_V$, see Remark 5.11.

Proposition 5.27 For non-zero $p \in \mathcal{P}$ we have $\kappa^p\left(\varsigma^p(V)^\perp\right) = \left(V^\perp \cap \mathrm{lin}(\mathbb{F}(p))\right) + \mathbb{R}p$. For non-zero $p \in \mathcal{P}_V$ and a subset $M \subset \mathbb{F}(p)$ we have

$$\left(M + V^\perp\right) \cap \overline{S}(A) = \left[M + \kappa^p\left(\varsigma^p(V)^\perp\right)\right] \cap \mathbb{F}(p).$$

[Proof on page 216]

Corollary 5.28 Let $p \in \mathcal{P}_V \setminus \{0\}$ and $q \in \mathcal{P}$ such that $q \leq p$. Then $q \in \mathcal{P}_V$ if and only if $q \in \kappa^p(\mathcal{P}_{\varsigma^p(V)})$. [Proof on page 217]

Definition 5.29 An **access sequence of faces** for a convex set $C \subset \mathbb{R}^k$ is a sequence of faces $(F_1, \ldots, F_m)$ of C such that $C \supsetneq F_1 \supsetneq \cdots \supsetneq F_m \supsetneq \emptyset$ and such that $F_1 \in \mathcal{F}_\perp(C)$ and $F_i \in \mathcal{F}_\perp(F_{i-1})$ for $i = 2, \ldots, m$. An **access sequence of projectors** for sr_V is a sequence of projectors $(p_1, \ldots, p_m)$ in $\mathcal{P}$ such that $\mathbb{1} \gneq p_1 \gneq \cdots \gneq p_m \gneq 0$ and such that $p_1 \in \mathcal{P}_{V,\perp}$ and $(\kappa^{p_{i-1}})^{-1}(p_i) \in \mathcal{P}_{\varsigma^{p_{i-1}}(V),\perp}$ for $i = 2, \ldots, m$.

As the analogue of a state reflection in the context of a Borel measure on $\mathbb{R}^k$ (with statistic), Csiszár and Matúš [Cs05] use the concept of convex support and they introduced the concept of convex core of a measure. The convex geometry of these sets is studied by access sequences of faces. The idea of an access sequence is also used by Grünbaum [Grü]. He defines a **poonem** of a convex set $C \subset \mathbb{R}^k$ as a member of an access sequence of faces for C. The equivalence of the concepts face and poonem is the statement of the lemma below.

Lemma 5.30 Every proper face F of a convex set $C \subset \mathbb{R}^k$ belongs to an access sequence of faces for C. Every ordered pair $F \subset G$ of proper faces of C belongs to an access sequence of faces for C. [Proof on page 217]

Theorem 2 For $m \in \mathbb{N}$ let $(F_1, \dots, F_m) \subset \mathcal{F}_V$ be a sequence of faces and set $p_i := s_V(F_i)$ for $i = 1, \dots, m$. Then $(F_1, \dots, F_m)$ is an access sequence of faces for sr_V if and only if $(p_1, \dots, p_m) \subset \mathcal{P}_V$ is an access sequence of projectors for sr_V. [Proof on page 217]

Corollary 5.31 A proper projector $p \in \mathcal{P}$ belongs to $\mathcal{P}_V$ if and only if p belongs to an access sequence of projectors for sr_V. [Proof on page 218]

Corollary 5.32 Every pair of proper projectors $p, q \in \mathcal{P}_V$ which are comparable, that is $p \leq q$ or $q \leq p$, belongs to an access sequence of projectors for sr_V. [Proof on page 218]

The following remark completes the summary on state reflections in Figure 5.6.

Remark 5.33 (Calculation of projector lattices with access sequences) Let us summarize what we need to calculate a projector lattice. By Corollary 5.31 a proper projector $p \in \mathcal{P}$ belongs to $\mathcal{P}_V$ if and only if p belongs to an access sequence of projectors for sr_V. The task is to calculate the access sequences. This may be done in ascending sequence length. A sequence of length one is a proper exposed projector. All of these can be calculated from V by (5.16)

$$\mathcal{P}_{V,\perp} = p_+(V) \cup \{0\}.$$

Every proper projector $p \in \mathcal{P}_{V,\perp}$ (of rank ≥ 2) may give rise to a number of access sequences for sr_V of length two

$$(p, \kappa^p(r))$$

where $r \in \mathcal{P}_{\varsigma^p(V),\perp}$ hence $r \in p_+(\varsigma^p(V))$. The traceless compression (5.60)

$$\varsigma^p(V) = (\kappa^p)^{-1} \circ \pi_{\mathrm{lin}(\mathbb{F}(p))}(V)$$

is calculated by orthogonal projection of a vector $v \in V$ to $\mathrm{lin}(\mathbb{F}(p))$ using (5.3)

$$\pi_{\mathrm{lin}(\mathbb{F}(p))}(v) = pvp - \mathrm{tr}(pv)\frac{p}{\mathrm{tr}(p)}.$$

In this second step we are only concerned with projectors $\kappa^p(r) \lneq p$. We may skip the *-isomorphism κ^p and calculate maximal projectors for elements of $\pi_{\mathrm{lin}(\mathbb{F}(p))}(V)$ directly.

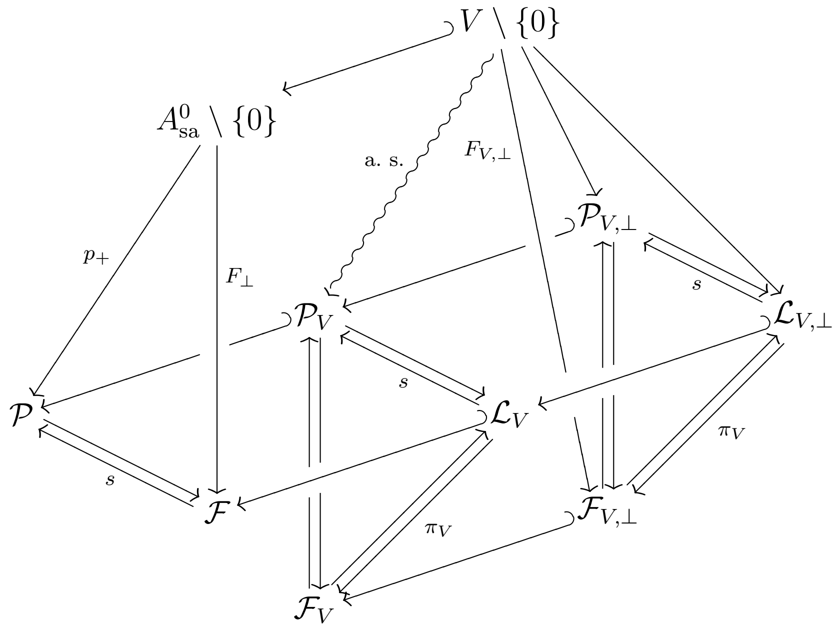


Figure 5.6: The commuting diagram summarizes the structure of a state reflection. The arrows with a curved tail denote embeddings. The sets and mappings are explained in Remark 5.13 on page 109 except the arrow $V \setminus \{0\} \overset{\text{a. s.}}{\rightsquigarrow} \mathcal{P}_V$ which is not a mapping but an algorithm, see Remark 5.33.

To simplify calculus, we may add any multiple of $\mathbb{1}$ to a matrix $v \in V$ before evaluation of $\pi_{\mathrm{lin}(\mathbb{F}(p))}(v)$. This can also be useful when an exponential family is given by a statistics of self-adjoint matrices, see Remark 7.19. Iteratively, we can calculate any access sequence of projectors for sr_V and determine the projector lattice $\mathcal{P}_V$. A handicap with calculation of projector lattices will of course be the eigenvalue discussion for the non-abelian case in higher dimensions.

Example 5.34 We use the method explained in Remark 5.33 and calculate the projector lattice for two simple examples.

(a) Consider the case (5.36) with $V = \mathrm{Lin}_{\mathbb{R}}\{(\sigma_1 \oplus 1) - \frac{\mathbb{1}}{3},\ (\sigma_2 \oplus 1) - \frac{\mathbb{1}}{3}\}$ and exposed projector lattice (5.37)

$$\mathcal{P}_{V,\perp} \;=\; \big\{0,\ 0_2 \oplus 1,\ p,\ q,\ \mathbb{1}\big\} \;\cup\; \big\{p_+(\sigma_1 \cos(\alpha) + \sigma_2 \sin(\alpha)) \oplus 0 \;:\; \alpha \in (\tfrac{\pi}{2}, 2\pi)\big\}.$$

The proper exposed projectors of rank ≥ 2 are $p := p_+(\sigma_1) \oplus 1$ and $q := p_+(\sigma_2) \oplus 1$. We get with (4.6)

$$\begin{aligned} p(\sigma_1 \oplus 1)p &= [p_+(\sigma_1)\,\sigma_1\,p_+(\sigma_1)] \oplus 1 = p_+(\sigma_1) \oplus 1 = p, \\ p(\sigma_2 \oplus 1)p &= [p_+(\sigma_1)\,\sigma_2\,p_+(\sigma_1)] \oplus 1 = 0_2 \oplus 1. \end{aligned}$$

Then by (5.3) we obtain $\pi_{\mathrm{lin}(\mathbb{F}(p))}(\sigma_1 \oplus 1) = 0$ and $\pi_{\mathrm{lin}(\mathbb{F}(p))}(\sigma_2 \oplus 1) = \frac{1}{2}(-p_+(\sigma_1) \oplus 1)$, so $\pi_{\mathrm{lin}(\mathbb{F}(p))}(V) = \mathbb{R}[p_+(\sigma_1) \oplus (-1)]$. Only the two maximal projectors $p_+(\sigma_1) \oplus 0$ and $0_2 \oplus 1$ appear for elements of $\pi_{\mathrm{lin}(\mathbb{F}(p))}(V)$ and only the first one is new. For q one has with the analogue arguments as before

$$\begin{aligned} q(\sigma_1 \oplus 1)q &= [p_+(\sigma_2)\,\sigma_1\,p_+(\sigma_2)] \oplus 1 = 0_2 \oplus 1, \\ q(\sigma_2 \oplus 1)q &= [p_+(\sigma_2)\,\sigma_2\,p_+(\sigma_2)] \oplus 1 = p_+(\sigma_2) \oplus 1 = q \end{aligned}$$

and

$$\mathcal{P}_V = \mathcal{P}_{V,\perp} \ \cup \ \big\{p_+(\sigma_1) \oplus 0,\ \ p_+(\sigma_2) \oplus 0\big\}.$$

(b) Consider the case (5.42) with $V = \mathrm{Lin}_{\mathbb{R}}\{\sigma_1 \oplus 0,\ (\sigma_2 \oplus 1) - \frac{\mathbb{1}}{3}\}$ with exposed projector lattice (5.43)

$$\mathcal{P}_{V,\perp} \ = \ \{0,\ p,\ \mathbb{1}\} \ \cup \ \Big\{p_+(\sigma_2 \cos(\alpha - \frac{\pi}{4}) - \sigma_1 \sin(\alpha - \frac{\pi}{4})) \oplus 0 \ : \ \alpha \in [0, 2\pi) \setminus \{\frac{\pi}{4}\}\Big\}.$$

The only proper exposed projector of rank ≥ 2 is $p := p_+(\sigma_2) \oplus 1$. We get with (4.6)

$$\begin{aligned} p(\sigma_1 \oplus 0)p &= [p_+(\sigma_2)\,\sigma_1\,p_+(\sigma_2)] \oplus 0 = 0_2 \oplus 0 = 0, \\ p(\sigma_2 \oplus 1)p &= [p_+(\sigma_2)\,\sigma_2\,p_+(\sigma_2)] \oplus 1 = p_+(\sigma_2) \oplus 1 = p. \end{aligned}$$

So the traceless compression (5.60) of V is $\varsigma^p(V) = 0$ and this gives $\mathcal{P}_V = \mathcal{P}_{V,\perp}$.

Remark 5.35 (Invariance under embedding) We have seen in Proposition 5.15 the independence of a state reflection given an algebra $B \subset A$ with $V \subset B$,

$$\mathrm{sr}_V = \pi_V(\overline{S}(B))$$

if B is a direct sum of full matrix algebras. The face lattice $\mathcal{F}_V$ and exposed face lattice $\mathcal{F}_{V,\perp}$ of the state reflection will be the same whether calculated in A or B. By Remark 5.33 the projector lattices $\mathcal{P}_V$ and $\mathcal{P}_{V,\perp}$ are equal for the two algebras, except possibly the identity that can be different.

However, the lifted face lattices may be different. An example is calculated in Example 5.7 (c) on page 102. One can consider the vector space $V \subset M_2 \oplus \mathbb{C}$ spanned by $-\frac{\mathbb{1}_2}{2} \oplus 1$ and $\sigma_3 \oplus 0$. The lifted face for the projector $p := \mathbb{1}_2 \oplus 0$ is the Bloch ball

$$M_2 \oplus 0.$$

When the vector space V is considered as a subset of the abelian algebra $\mathbb{C}^3$ of diagonal matrices, then V is the space of real triples that sum up to zero. Now the lifted face for the projector p is the segment

$$[(1,0,0),\ (0,1,0)],$$

which is a face of the triangle $\overline{S}(\mathbb{C}^3)$.

If the algebra B is a compression of A by a non-zero projector $p \in \mathcal{P}$ the situation is better. We have $\mathcal{P}_V \setminus \{\mathbb{1}\} = \kappa^p(\mathcal{P}_{(\kappa^p)^{-1}(V)}) \setminus \{p\}$ and by (5.59) we get for $r \in \mathcal{P}_{(\kappa^p)^{-1}(V)}$ the equality $\mathbb{F}(\kappa^p(r)) = \kappa^p(\mathbb{F}(r))$.

5.4 Topology of projector lattices

We start to discuss the topology of the face lattices $\mathcal{P}_V$ and $\mathcal{P}_{V,\perp}$. We ask if $\mathcal{P}_V$ and $\mathcal{P}_{V,\perp}$ have the same closure—this question is linked to the topology of an exponential families (Remark 7.25). We can not give an answer and are led to further unsolved questions about closedness of the skeletons of a state reflection, symmetrizations of the state space and open mappings.

In a classical case $A = \mathbb{C}^n$ for some $n \in \mathbb{N}$ the projector lattices are finite and one has the equality $\mathcal{P}_{V,\perp} = \mathcal{P}_V \cong \mathcal{F}_V$ because the state reflection as a polytope has only exposed faces, see Remark 8.5 (c). The lattice is finite because $\mathcal{P}_V \subset \mathcal{P}$ and $\mathcal{P}$ is isomorphic to the power set of $\{1,\dots,n\}$, see Remark 2.36 (b). For a matrix algebra A the projector lattice $\mathcal{P}$ is a union of compact real analytical manifolds (Lemma 2.38) and the projector lattices $\mathcal{P}_{V,\perp}$ and $\mathcal{P}_V$ can be distinct. In Example 5.34 (a) the exposed projector lattice $\mathcal{P}_{V,\perp}$ is not closed while

$$\mathcal{P}_V \;=\; \overline{\mathcal{P}_{V,\perp}} \;=\; \mathcal{P}_{V,\perp} \;\cup\; \big\{p_+(\sigma_1) \oplus 0,\ p_+(\sigma_2) \oplus 0\big\}.$$

In example (b) the lattice $\mathcal{P}_{V,\perp} = \mathcal{P}_V$ is not closed and $\overline{\mathcal{P}_V} \;=\; \mathcal{P}_V \;\cup\; \big\{p_+(\sigma_2) \oplus 0\big\}$. These examples suggest a first question.

Question 5 Do the projector lattices $\mathcal{P}_{V,\perp}$ and $\mathcal{P}_V$ have the same closure? In other words, does $\mathcal{P}_V \subset \overline{\mathcal{P}_{V,\perp}}$ hold?

We can try and prove a consequence of Question 5. By (5.56) one has the disjoint cover by relative interiors of face reflections

$$\mathrm{sr}_V = \bigcup_{p \in \mathcal{P}_V \setminus \{0\}} \mathrm{ri}(\mathbb{F}_V(p)).$$

So an affirmative of Question 5 implies an affirmative of the following question.

Question 6 Do the relative interiors of face reflections with support projector in $\overline{\mathcal{P}_{V,\perp}}$ cover the state reflection,

$$\mathrm{sr}_V \subset \bigcup_{p \in \overline{\mathcal{P}_{V,\perp}}} \mathrm{ri}(\mathbb{F}_V(p))\,?$$

We prove Question 6 in Theorem 3 but under the additional assumption that the skeletons (3.28) of the state reflection sr_V are closed. This leads us to the next question.

Question 7 Are the skeletons of the state reflection sr_V closed?

We approach Question 6 using dimension functions of faces in dependence of their support projectors. Let us compare to the dimension of a face of the state space. The face dimension for the state space $\mathcal{P} \to \mathbb{N}_0$, $p \mapsto \dim(\mathbb{F}(p))$ is a locally constant function by Lemma 2.38 and (4.35). In contrast, the face dimension for the state reflection $\mathcal{P} \to \mathbb{N}_0$, $p \mapsto \dim(\mathbb{F}_V(p))$ is not continuous:

Example 5.36 We consider Example 5.7 (f) and the family of rank two projectors for $\alpha \in [0, 2\pi)$

$$p_\alpha := p_+(c(\alpha)\widehat{\sigma}) \oplus 1.$$

For each $\alpha \in [0, 2\pi)$ the compressed algebra $A^{p_\alpha} = \mathbb{C}^2$ is abelian so the face $\mathbb{F}(p_\alpha) = [p_+(c(\alpha)\widehat{\sigma}) \oplus 0,\ 0_2 \oplus 1]$ is a segment with one-dimensional translation vector space $\mathrm{lin}(\mathbb{F}(p_\alpha))$ generated by $p_+(c(\alpha)\widehat{\sigma}) \oplus (-1)$. We show that

$$\dim(\mathbb{F}_V(p_\alpha)) = \begin{cases} 0 & \text{if } \alpha = \frac{\pi}{4}, \\ 1 & \text{otherwise.} \end{cases}$$

This follows by discussion of the coefficients in (5.29). A convex geometric aspect argument for $\alpha \neq \frac{\pi}{4}$ is the following. The projector $p := p_+(c(\alpha)\widehat{\sigma}) \oplus 0$ belongs to the lifted exposed projector lattice $\mathcal{P}_{V,\perp}$ (5.39). If we had $\dim(\mathbb{F}_V(p_\alpha)) = 0$ then $\mathrm{lin}(\mathbb{F}(p_\alpha)) \subset V^\perp$ and by (5.63) follows

$$0_2 \oplus 1 \in (\mathbb{F}(p) + V^\perp) \cap \overline{S}(A) = \mathbb{F}(p).$$

But this is a contradiction to the fact that p is an extreme point of $\overline{S}(A)$.

Lemma 5.37 The function $\mathcal{P} \to \mathbb{N}_0$, $p \mapsto \dim(\mathbb{F}_V(p))$ is lower semi-continuous.

[Proof on page 218]

Remark 5.38 (A wrong short argument) Lower semi-continuity of the function $\mathcal{P} \to \mathbb{N}_0$, $p \mapsto \dim(\mathbb{F}_V(p))$, which was proved in Lemma 5.37, does not straight forward imply lower semi-continuity of the dimension function

$$\mathrm{sr}_V \to \mathbb{N}_0, \qquad x \mapsto \dim(F_V(x)).$$

Surely from a sequence $x_i \subset \mathrm{sr}_V$ with limit $x := \lim_{i\to\infty} x_i$ we can select a subsequence with converging support projectors $p := \lim_{i\to\infty} s_V(x_i)$. Then with (5.55) we have

$$\dim(\mathbb{F}_V(p)) \leq \liminf_{i\to\infty} \dim(\mathbb{F}_V(s_V(x_i))) = \liminf_{i\to\infty} \dim(F_V(x_i)).$$

However the example (5.39) teaches us that the case $s_V(x) \gneq p$ may well happen and $\dim(F_V(x)) = \dim(\mathbb{F}_V(s_V(x))) > \dim(\mathbb{F}_V(p))$ is possible.

To establish an affine lifting coordinate system we compare balls and simplices.

Remark 5.39 (Balls and simplices) Let $k \in \mathbb{N}$, $\epsilon > 0$ and x belong to a finite dimensional Euclidean space. Into a closed k-dimensional ball with center x and radius ϵ we can inscribe a regular k-dimensional simplex of edge length $\epsilon\sqrt{\frac{2(k+1)}{k}}$. Into the simplex we can inscribe a closed k-dimensional ball with center x and with radius $\frac{\epsilon}{k}$. A three-dimensional example is depicted in Figure 5.7. A simple proof can use the state space $\overline{S}(\mathbb{C}^{k+1})$ as a model simplex with the trace state $\frac{\mathbb{1}}{\mathrm{tr}(\mathbb{1})}$ as centroid. Observe that the smallest sphere about the simplex has radius $\|p - \frac{\mathbb{1}}{\mathrm{tr}(\mathbb{1})}\|_2$ and the largest ball inside the simplex has radius $\|\frac{q}{\mathrm{tr}(q)} - \frac{\mathbb{1}}{\mathrm{tr}(\mathbb{1})}\|_2$ where $p \in \mathbb{C}^{k+1}$ is an orthogonal projector of rank one and $q \in \mathbb{C}^{k+1}$ is an orthogonal projector of rank k.

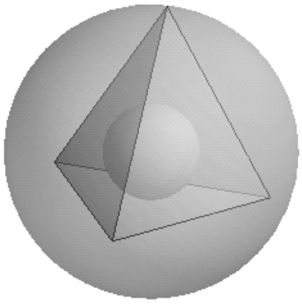


Figure 5.7: The regular three-dimensional simplex of edge length $\frac{2\sqrt{2}}{\sqrt{3}}$ is depicted "between" concentric spheres of radius 1 and $\frac{1}{3}$. The small sphere is incident with the centroids of the four facets of the tetrahedron. The large sphere is incident with the four extreme points of the tetrahedron.

Lemma 5.40 If a sequence $(x_i) \subset \mathrm{sr}_V$ converges to $x \in \mathrm{sr}_V$ then $x \in \mathrm{ri}(\mathbb{F}_V(p))$ for some $p \in \overline{\{s(x_i) : i \in \mathbb{N}\}}$ or $\lim\limits_{i\to\infty} \left(\min_{y\in \mathrm{rb}(F_V(x_i))} \|x_i - y\|_2\right) = 0$. [Proof on page 219]

Remark 5.41 (Two arguments for Theorem 3) **(a)** Recall (3.28) for $d \geq 0$ the definition of the d-skeleton of a convex compact subset $C \subset \mathbb{R}^n$

$$\mathrm{skel}(C, d) = \{x \in C : \dim(F(C, x)) \leq d\}.$$

The d-skeleton of C is the union of all faces of C having dimension less or equal d. By Remark 3.13 (a) the skeletons of C are closed if and only if the dimension function

$$C \to \mathbb{N}_0, \qquad x \mapsto \dim(F(C, x))$$

is lower semi-continuous.

(b) The theorem of Straszewicz [St] says that an extreme point of C (point in the 0-skeleton) is the limit of a sequence of exposed points of C. A generalization due to Asplund [Asp] says that a point in the d-skeleton of C is the limit of a sequence (x_i) where each point x_i belongs to an exposed face of C with dimension less or equal d.

Theorem 3 If the skeletons of sr_V are closed then $\mathrm{sr}_V = \bigcup_{p \in \overline{\mathcal{P}_{V,\perp}}} \mathrm{ri}\,(\mathbb{F}_V(p))$.
[Proof on page 221]

Corollary 5.42 If $\mathcal{P}_V$ is closed and the skeletons of sr_V are closed then $\overline{\mathcal{P}_{V,\perp}} = \mathcal{P}_V$.
[Proof on page 221]

We start the discussion of Question 7 about the closedness of skeletons of a state reflection by an approach with symmetrizations and stable convex sets.

Definition 5.43 Let $V \subset \mathbb{R}^m$ be a vector space. The **reflection** at V is the linear mapping

$$r_V : \mathbb{R}^m \to \mathbb{R}^m \qquad \text{defined by} \qquad r_V|_V = \mathrm{Id}\,|_V \qquad \text{and} \qquad r_V|_{V^\perp} = -\,\mathrm{Id}\,|_{V^\perp}.$$

A subset $M \subset \mathbb{R}^m$ is **symmetric** at V if $M = r_V(M)$. Let $H \subset \mathbb{R}^m$ be a linear hyperplane. The **Steiner symmetrization** of a compact convex set $C \subset \mathbb{R}^m$ at H is [Lei]

$$S_H(C) := \bigcup_g \left(\frac{1}{2}(g \cap C) + \frac{1}{2}(g \cap r_H(C)) \right). \tag{5.64}$$

Here g is an arbitrary line perpendicular to H.

Remark 5.44 (Symmetrizations) Let $C \subset \mathbb{R}^m$ be an arbitrary compact convex set.
(a) The Steiner symmetrization of C at a linear hyperplane H is created as follows. The lines perpendicular to H cut segments from C which are moved parallel until their midpoints lie on H. By Theorem 18.1 in [Lei], the Steiner symmetrization of C at H is a compact convex subset of $\mathbb{R}^m$ symmetric at H.

(b) For a vector space V in higher dimensions $d := \mathrm{codim}(V) > 1$ the generalization of the Steiner symmetrization is the following [Wil]. At each point $v \in V$ we consider the section $(v+V^\perp)\cap C$. This is known to be a compact convex subset of $v+V^\perp$ and the finite d-dimensional volume $V^d((v+V^\perp)\cap C)$ is defined. Consider the closed ball $B(v) \subset V^\perp$ centered at zero and having the same d-dimensional volume $V^d(B(v)) = V^d((v+V^\perp)\cap C)$. The union

$$S_V(C) := \bigcup_{v \in V} (v + B(v))$$

is a compact convex set symmetric at V with m-dimensional volume $V^m(S_V(C)) = V^m(C)$.

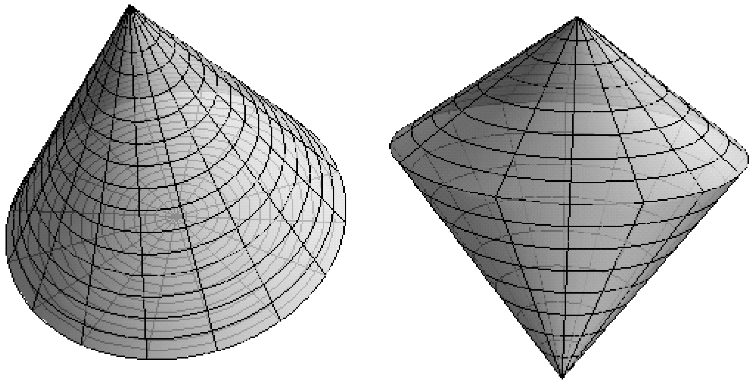


Figure 5.8: Depicted are the frustum C (left) discussed in Example 5.45 with Steiner symmetrization $S_V(C)$ at the plane V (right). The horizontal is V and the vertical is $V^\perp \cap \mathrm{lin}(C)$. The horizontal ellipse about $S_V(C)$ consists of extreme points except at the intersection with the vertical segment in front. The 0-skeleton of $S_V(C)$ is not closed.

Example 5.45 To illustrate Steiner symmetrization in the context of state spaces we consider Example 5.7 (f) based on the algebra $A = M_2 \oplus \mathbb{C}$ with four-dimensional state space. We define V by multiples of the basis vectors from (5.42) on page 106

$$v_1 = \sigma_1 \oplus 0 \qquad \text{and} \qquad v_2 = \sigma_2 \oplus 1 - \frac{\mathbb{1}}{3}.$$

Instead of the four dimensional state space we consider for reasons of intuitive geometric assistance the three dimensional frustum C with apex $a := 0_2 \oplus 1$ and with base circle (5.42) consisting of pure states for $\alpha \in [0, 2\pi)$

$$b(\alpha) := p_+(c(\alpha)\widehat{\sigma}) \oplus 0 = \frac{1}{2}\Big(\mathbb{1}_2 + \sigma_2 \cos(\alpha - \frac{\pi}{4}) - \sigma_1 \sin(\alpha - \frac{\pi}{4})\Big) \oplus 0 \tag{5.65}$$

The state reflection $\mathrm{sr}_V = \pi_V(\overline{S}(A))$ is also the projection to V of C (5.27). The kernel $V^\perp \cap \mathrm{lin}(C)$ of the projection π_V has the slope $\varphi = \frac{\pi}{3}$ to the base of the frustum C and to its translation vector space $\mathrm{Lin}\{\sigma_1, \sigma_2\} \oplus 0$, see Figure 5.2 on page 101. In the coordinate system (5.29) spanning $\mathrm{lin}(C)$ and having V as the x-y plane we have

$$\begin{aligned} a &= \Big(0,\ \tfrac{1}{\sqrt{6}},\ \tfrac{1}{\sqrt{2}}\Big), \\ b(\alpha) &= \Big(-\tfrac{1}{\sqrt{2}}\sin(\alpha - \tfrac{\pi}{4}),\ \tfrac{1}{2\sqrt{6}}\big(3\cos(\alpha - \tfrac{\pi}{4}) - 1\big),\ -\tfrac{1}{2\sqrt{2}}\big(\cos(\alpha - \tfrac{\pi}{4}) + 1\big)\Big). \end{aligned}$$

The segment $s := \big[b(\frac{\pi}{4}),\ a\big] = \big[\frac{1}{2}(\mathbb{1}_2 + \sigma_2) \oplus 0,\ 0_2 \oplus 1\big]$ of C is perpendicular to V (5.40). It has the coordinates

$$s = \left[(0, \frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{2}}),\ (0, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{2}})\right]. \tag{5.66}$$

The right angle $V \perp s$ corresponds to the kernel $V^\perp \cap \mathrm{lin}(C)$ of π_V parallel to s.

The Steiner symmetrization $S_V(C)$ of C at V is a compact and convex set by Remark 5.44 (b). Hence by Minkowski Theorem we just need to discuss the extreme points of $S_V(C)$. These are the symmetrizations of the apex a and of the base points $b(\alpha)$ for $\alpha \in [0, 2\pi)$. The segment s including the apex a and the base point $b(\frac{\pi}{4})$ is already symmetric in the above coordinates. The symmetrizations of other base points $b(\alpha)$ belong to the ellipse

$$e := \{(x,y,0) \in \mathbb{R}^3 : \frac{x^2}{\left(\frac{1}{\sqrt{2}}\right)^2} + \frac{\left(y + \frac{1}{2\sqrt{6}}\right)^2}{\left(\frac{\sin(\varphi)}{\sqrt{2}}\right)^2} = 1\} \tag{5.67}$$

with $\sin(\varphi) = \frac{1}{2}\sqrt{3}$. We have

$$S_V(C) = \mathrm{conv}(e \ \cup \ \{(0, \frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{2}}),\ (0, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{2}})\}). \tag{5.68}$$

The frustum and the symmetrization are depicted in Figure 5.8. The 0-skeleton of the Steiner symmetrization $S_V(C)$ is not closed. Points on the ellipse e are extreme points of the Steiner symmetrization $S_V(C)$ with exception of $(0, \frac{1}{\sqrt{6}}, 0)$. This point belongs to the relative interior of the segment s. Notice that the well-known **skew cone** [Pa]

$$\mathrm{conv}(\{(x,y,0) : (x-1)^2 + y^2 \leq 1\} \cup \{(0,0,1), (0,0,-1)\})$$

is affinely isomorphic to $S_V(C)$. This is the standard example for a convex set which fails to have a closed 0-skeleton.

Definition 5.46 A convex set $C \subset \mathbb{R}^m$ is **stable** if the **midpoint map**

$$C \times C \to C, \qquad (x,y) \mapsto \frac{1}{2}(x+y)$$

is open in the relative topology.

Remark 5.47 (a) The connection to our problem is Theorem 2.3 in [Pa]. Let $C \subset \mathbb{R}^m$ be a compact convex set. Among others the theorem states equivalence of the assertions

- C is stable,
- all skeletons of C are closed.

We apply this equivalence to the study of projection maps and of their images.

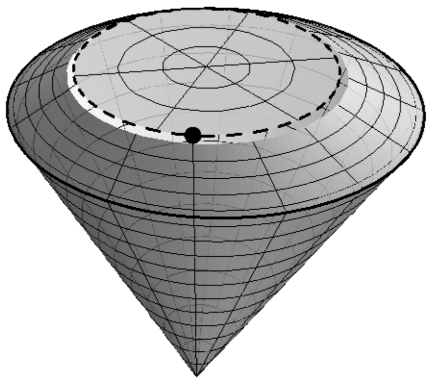


Figure 5.9: The drawing shows the truncation $\{(x, y, z) \in S_V(C) : z \leq 0.3\}$ of the Steiner symmetrization $S_V(C)$ from Example 5.45. The horizontal is V and the vertical is $V^\perp \cap \operatorname{lin}(C)$. The projection of the open set $U := \{(x, y, z) \in S_V(C) : z > 0.3\}$ to V is depicted under translation $(0, 0, 0.3)$ as the top ellipse. This ellipse does not contain its own relative boundary (dashed) except for $(0, \frac{1}{\sqrt{6}}, 0) \in \pi_V(U)$ (thick point). The boundary curve e (thick) of the compact set $\pi_V(S_V(C))$ intersects $\pi_V(U)$ exactly at $(0, \frac{1}{\sqrt{6}}, 0)$ so the projection image $\pi_V(U)$ is not open.

(b) In Example 5.45 the Steiner symmetrization $S_V(C)$ has a 0-skeleton which is not closed. This implies by (a) that $S_V(C)$ is not stable. To disprove stability directly, we use $0 < z_0 < \frac{1}{\sqrt{2}}$ the non-empty open sets $U_\pm := \{(x, y, z) \in S_V(C) : \pm z > z_0\}$. Then the image of the open set $U_+ \times U_-$ under the midpoint map

$$\{\frac{1}{2}(b_+ + b_-) : b_+ \in U_+, b_- \in U_-\}$$

contains $(0, \frac{1}{\sqrt{6}}, 0)$ but no other point of the ellipse $e \subset S_V(C)$ (5.67).

Lemma 5.48 Let $C \subset \mathbb{R}^m$ be a stable convex set symmetric at a vector space $V \subset \mathbb{R}^m$. Then the projection $\pi_V|_C$ is open and the image $\pi_V(C)$ is stable. [Proof on page 221]

Remark 5.49 **(a)** We practice the concepts in use and give an alternative proof to Remark 5.47 (b) that the Steiner symmetrization $S_V(C)$ (5.68) is not stable: as discussed in Figure 5.9 the projection mapping $\pi_V|_{S_V(C)}$ is not open. Then by Lemma 5.48 the symmetric set $S_V(C)$ is not stable.

(b) In the affirmative we can use Lemma 5.48 to prove for a convex set $C \subset \mathbb{R}^m$ and a vector space $V \subset \mathbb{R}^m$ that the projection $\pi_V(C)$ is stable. We have to decide whether there exists a stable convex set $\widetilde{C} \subset \mathbb{R}^m$ symmetric at V and such that $\pi_V(C) = \pi_V(\widetilde{C})$. This may be difficult. If C is stable by itself then we can give an alternative criterion at least in the hyperplane case $\operatorname{codim}(V) = 1$.

Definition 5.50 A **symmetrization map** for $M \subset \mathbb{R}^m$ at a linear space $V \subset \mathbb{R}^m$ is an injective mapping $s : M \to \mathbb{R}^m$ such that $s(M)$ is symmetric at V and such that s preserves fibers, that is $\pi_V \circ s = \pi_V$. Then $s(M)$ is called a **symmetrization** of M at V.

Example 5.51 For a compact convex set $C \subset \mathbb{R}^m$ and a hyperplane $H \subset \mathbb{R}^m$ we can use Steiner symmetrization (5.64) to define a symmetrization map. If $h \in H^\perp$ is normalized we obtain the Steiner symmetrization $s_H(C) = s(C)$ of C from the mapping defined for $x \in C$ by $s(x) := x - \frac{1}{2}(\max_{y \in (x+H^\perp)\cap C}\langle h, y\rangle + \min_{y \in (x+H^\perp)\cap C}\langle h, y\rangle)h$.

Proposition 5.52 Let $C \subset \mathbb{R}^m$ be a stable convex set with a homeomorphic symmetrization map at a linear hyperplane $H \subset \mathbb{R}^m$. Then the projection $\pi_H|_C$ is open and the image $\pi_H(C)$ is stable. [Proof on page 222]

Corollary 5.53 Let $H \subset A^0_{\mathrm{sa}}$ be a linear hyperplane. If there exists a continuous symmetrization map of the state space $\overline{S}(A)$ at H then the state reflection sr_H is stable.
[Proof on page 224]

Remark 5.54 **(a)** We can show that any symmetrization map s for the frustum C in Example 5.45 at V is discontinuous. Otherwise by compactness the mapping s would be a homeomorphism. Since the frustum is stable, Proposition 5.52 shows that the projection mapping $\pi_V|_C$ is open. This is disproved in Figure 5.9.

(b) The symmetrization map s in Example 5.51 for the frustum C at the vector space V in Example 5.45 has a discontinuity at the bottom point $(0, \frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{2}})$ of the segment (5.66)

$$\left[b(\frac{\pi}{4}) \oplus 0,\ 0_2 \oplus 1\right] = \left[(0, \frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{2}}),\ (0, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{2}})\right].$$

The vector space V corresponds to the x-y-plane so the segment is fixed under symmetrization while the pointed base circle (5.65) of C

$$\{b(\alpha) : \alpha \in [0, 2\pi) \setminus \{\frac{\pi}{4}\}\}$$

is moved to the pointed ellipse (5.67) $e \setminus \{0, \frac{1}{\sqrt{6}}, 0\}$. The frustum C and symmetrization $S_V(C)$ are depicted in Figure 5.8.

(c) The discontinuity discovered in (b) persists under restriction of the symmetrization map s to the union of proper lifted faces of sr_V (5.44) which cover exactly the base circle and segment used in (b). It follows that a symmetrization map for the state space $\overline{S}(A)$ at V has a discontinuity at $b(\frac{\pi}{4}) = \frac{1}{2}(\mathbb{1}_2 + \sigma_2) \oplus 0$.

(d) We stop the discussion here because we are not prepared with the necessary examples so far. Stability of a convex set is an issue for dimension three and higher. Indeed, if $C \subset \mathbb{R}^m$ is a compact convex set then the $(m-1)$- and $(m-2)$-skeleton is closed. The $(m-1)$-skeleton is the relative boundary of C or the set C itself. For the $(m-2)$-skeleton we notice that a point in the relative interior of a $(m-1)$-dimensional face F of C belongs to an open ball of dimension m that meets only $\mathrm{ri}(F)$ and $\mathrm{ri}(C)$ and no other points of C. The relative interior of a facet can not be approximated by points in the $(m-2)$-skeleton.

6 Exponential families

This chapter is a collection and customization of known facts about exponential families in a matrix algebra. The swallow and Staffelberg family in Section 6.1 seem strange examples of quantum exponential families though they are really the first choice—judging by their simple definition. In Section 6.2 we recall from the literature connections among relative entropy, the BKM-metric and charts for the manifold of invertible density matrices. In Section 6.3 we write the mean value chart for exponential families; this was not done before in this general form.

6.1 Examples and illustrations

We start with a definition of exponential families and with examples.

Definition 6.1 The **normalized exponential** is the real-analytic function

$$\exp_1 : \quad A_{\mathrm{sa}} \to S(A), \quad \theta \mapsto \frac{\exp(\theta)}{\mathrm{tr}(\exp(\theta))} \tag{6.1}$$

with the matrix exponential $e^{\theta} = \exp(\theta) = \sum_{k=0}^{\infty} \frac{\theta^k}{k!}$. The image $\mathcal{E}$ of a non-empty affine subspace of A_{sa} under the normalized exponential $\exp_1$ is an **exponential family** in A. If $\mathcal{E}' \subset \mathcal{E}$ is an exponential family in A, then $\mathcal{E}'$ is an **exponential sub-family** of $\mathcal{E}$.

The restriction $\exp_1|_{A^0_{\mathrm{sa}}}$ to the traceless matrices is a diffeomorphism. The inverse is the **traceless logarithm**

$$\ln_0 : \quad S(A) \to A^0_{\mathrm{sa}}, \quad \rho \mapsto \ln(\rho) - \mathrm{tr}(\ln(\rho))\widehat{\mathbb{1}}, \tag{6.2}$$

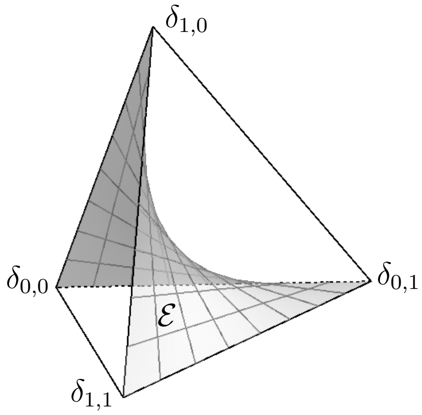


Figure 6.1: The exponential family $\mathcal{E}$ of factorizable probability distribution on the Cartesian product set $\{0,1\} \times \{0,1\}$ is depicted inside the probability simplex on the set. The four Dirac measures are the corners of the simplex.

which makes $S(A)$ a real-analytic manifold. Here $\widehat{\mathbb{1}} = \frac{\mathbb{1}}{\mathrm{tr}(\mathbb{1})}$ denotes the trace state.

The initial idea for this thesis was the study of multi-information in a quantum setting. A common approach to the problem uses the factorizable family discussed in Section 8.3. Here we present the simplest non-trivial example.

Example 6.2 (The factorizable family) We consider the algebra $A = \mathbb{C}^2 \otimes \mathbb{C}^2 \simeq \mathbb{C}^4$. An ONB of A is given by the vectors $\delta_{i,j} := \delta_i \otimes \delta_j$ for $i,j \in \{0,1\}$ and $\delta_0 := (1,0)$, $\delta_1 := (0,1)$, corresponding to the Dirac measures on $\{0,1\} \times \{0,1\}$. The state space is $\overline{S}(A) = \mathrm{conv}(\delta_{0,0}, \delta_{0,1}, \delta_{1,0}, \delta_{1,1})$. For $u := (\delta_0 - \delta_1) \otimes \mathbb{1}_2$ and $v := \mathbb{1}_2 \otimes (\delta_0 - \delta_1)$ we parametrize the exponential family $\mathcal{E} := \exp_1(\mathrm{Lin}\{u,v\})$ by

$$\mathbb{R}^2 \to \mathcal{E}, \qquad (\lambda, \mu) \mapsto \exp_1(\lambda u + \mu v).$$

Using Lemma 8.11 this is

$$\begin{aligned} \exp_1(\lambda u + \mu v) &= \left(\tfrac{e^{\lambda}}{e^{\lambda}+e^{-\lambda}} \delta_0 + \tfrac{e^{-\lambda}}{e^{\lambda}+e^{-\lambda}} \delta_1 \right) \otimes \left(\tfrac{e^{\mu}}{e^{\mu}+e^{-\mu}} \delta_0 + \tfrac{e^{-\mu}}{e^{\mu}+e^{-\mu}} \delta_1 \right) \\ &= \sum_{i,j=0}^{1} x_{i,j}(\lambda,\mu) \delta_{i,j} \end{aligned} \tag{6.3}$$

with coefficients $x_{i,j} = x_{i,j}(\lambda,\mu) = \frac{e^{(-1)^i \lambda}}{e^{\lambda}+e^{-\lambda}} \frac{e^{(-1)^j \mu}}{e^{\mu}+e^{-\mu}}$ for $i,j = 0,1$. The coefficients satisfy the relations

$$x_{0,0} + x_{0,1} + x_{1,0} + x_{1,1} = 1 \qquad \text{and} \qquad x_{0,0}x_{1,1} = x_{0,1}x_{1,0}.$$

From these relations it follows that probability distributions in $\mathcal{E}$ are factorizable in the following sense. The probability of the elementary event $(i,j) \in \{0,1\} \times \{0,1\}$ is $P(i,j) = x_{i,j}$. The first marginal of an elementary event $i \in \{0,1\}$ is $P_I(i) = x_{i,0} + x_{i,1}$ and the second marginal of an elementary event $j \in \{0,1\}$ is $P_{II}(j) = x_{0,j} + x_{1,j}$. Then for $(i,j) \in \{0,1\} \times \{0,1\}$

$$P(i,j) = P_I(i) P_{II}(j)$$

holds. The exponential family $\mathcal{E}$ is a piece of a hyperbolic paraboloid. The unitary transformation

$$S := \frac{1}{2}\begin{pmatrix} \sqrt{2} & 0 & 1 & 1 \\ 0 & \sqrt{2} & -1 & 1 \\ 0 & -\sqrt{2} & -1 & 1 \\ -\sqrt{2} & 0 & 1 & 1 \end{pmatrix} \qquad \text{respectively} \qquad T := \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \end{pmatrix}$$

yields that a coordinate quadruple $(x,y,z,c)^{\mathrm{t}} := S^{\mathrm{t}}(x_{0,0}, x_{0,1}, x_{1,0}, x_{1,1})^{\mathrm{t}}$ of $S^{\mathrm{t}}(\mathcal{E})$ respectively $(x,y,z,c)^{\mathrm{t}} := T^{\mathrm{t}}(x_{0,0}, x_{0,1}, x_{1,0}, x_{1,1})^{\mathrm{t}}$ of $T^{\mathrm{t}}(\mathcal{E})$ satisfies $z = x^2 - y^2$ respectively $z = 2xy$ and $c = \frac{1}{2}$. Here $^{\mathrm{t}}$ denotes the transposed of a vector or matrix. The space $T^{\mathrm{t}}(\mathcal{E})$ is the image of $(-\frac{1}{2}, \frac{1}{2}) \times (-\frac{1}{2}, \frac{1}{2})$ under $(x,y) \mapsto (x, y, 2xy, \frac{1}{2})^{\mathrm{t}}$. A hyperbolic paraboloid is covered by two families of transverse straight lines. The canonical parametrization (6.3) can be used to generate the two straight line families. One of them is the collection for $\lambda \in \mathbb{R}$ of exponential families with canonical parameter space

$$\Theta(\lambda) := \{\exp_1(\lambda u + \mu v) : \mu \in \mathbb{R}\}.$$

The variable factor of (6.3), the mapping

$$\mathbb{R} \to S(\mathbb{C}^2), \qquad \mu \mapsto \left(\frac{e^{\mu}}{e^{\mu} + e^{-\mu}}\delta_0 + \frac{e^{-\mu}}{e^{\mu} + e^{-\mu}}\delta_1\right)$$

parametrizes the relative open segment $]\delta_0, \delta_1[$. Since the tensor product is linear in each factor, the exponential sub-family $\exp_1(\Theta(\lambda))$ is a relative open segment included in $\mathcal{E}$. The analogous construction with the second parameter μ fixed yields the transverse family.

Example 6.3 (A frustum family) In the algebra $A := M_2 \oplus \mathbb{C}$ we consider a two-dimensional vector space $W \subset \mathrm{Lin}\{\sigma_1, \sigma_2, \sigma_3\}$ the vector $z := (-\frac{\mathbb{1}_2}{2}) \oplus 1$ and

$$C := \overline{S}(A) \cap \left(\frac{\mathbb{1}}{3} + W \oplus 0 + \mathbb{R}z\right).$$

By (5.23) we have $\mathrm{aff}(C) = \frac{\mathbb{1}}{3} + W \oplus 0 + \mathbb{R}z$ and $\mathrm{lin}(C) = W \oplus 0 + \mathbb{R}z$ and a convex hull description

$$C = \mathrm{conv}\left[\left(\overline{S}(M_2) \cap \left(\frac{\mathbb{1}_2}{2} + W\right)\right) \oplus 0,\ 0_2 \oplus 1\right]$$

with the Bloch ball $\overline{S}(M_2)$. The previous equation shows us that C is a frustum with base $\left(\overline{S}(M_2) \cap \left(\frac{\mathbb{1}_2}{2} + W\right)\right) \oplus 0$ and apex $0_2 \oplus 1$. The relative interior of the state space $\overline{S}(A)$ is the space $S(A)$ of invertible density matrices (4.27). We notice that $\mathrm{aff}(C)$ shares the point $\frac{\mathbb{1}}{3}$ with $S(A)$. Hence by (3.17) we have

$$\mathrm{ri}(C) = S(A) \cap \mathrm{aff}(C) = \{\rho \in C : \rho \text{ is invertible}\}\,. \tag{6.4}$$

Now the logarithm is defined on the relative interior of C. Invariance of functional calculus in M_2 (4.7) and the inclusion $\mathrm{ri}(C) \subset (\mathbb{R}\mathbb{1}_2 + W) \oplus \mathbb{R}$ give $\ln(\mathrm{ri}(C)) \subset (\mathbb{R}\mathbb{1}_2 + W) \oplus \mathbb{R}$. Hence the image under the traceless logarithm is included in

$$\ln_0(\mathrm{ri}(C)) \subset W \oplus 0 + \mathbb{R}z = \mathrm{lin}(C).$$

Similarly, for the exponential function $\exp(\mathrm{lin}(C)) \subset (\mathbb{R}\mathbb{1}_2 + W) \oplus \mathbb{R}$ holds. By the trace one condition on images of the normalized exponential we have

$$\exp_1(\mathrm{lin}(C)) \subset \mathrm{aff}\left(\left(\frac{\mathbb{1}_2}{2} + W\right) \oplus 0,\ \ 0_2 \oplus 1\right).$$

By positivity of images of the exponential function we get

$$\exp_1(\mathrm{lin}(C)) \subset \mathrm{conv}\left[\left(\overline{S}(M_2) \cap \left(\frac{\mathbb{1}_2}{2} + W\right)\right) \oplus 0,\ \ 0_2 \oplus 1\right] = C.$$

Images of the exponential are invertible and (6.4) gives $\exp_1(\mathrm{lin}(C)) \subset \mathrm{ri}(C)$. We have proved that $\mathrm{ri}(C)$ is the exponential family

$$\mathrm{ri}(C) = \exp_1(\mathrm{lin}(C)) = \exp_1(\mathrm{aff}(C)). \tag{6.5}$$

The last equality holds by invariance of $\exp_1$ under addition of multiples of $\mathbb{1}$. The frustum C is very versatile. By (5.24) it is also the projection of a state space, $C = \pi_{\mathrm{aff}(C)}\left(\overline{S}(M_2 \oplus \mathbb{C})\right)$. The drawing of a shape of C is included with Example 6.4. The coordinates of the frustum C are calculated at (5.29).

The new features of a quantum exponential family, discussed in Chapter 7, can be demonstrated with one of the following two examples.

Example 6.4 (Two quantum families) We define $\mathcal{E}$ as the image under the normalized exponential $\exp_1$ for the domain

$$\mathrm{Lin}\{\sigma_1 \oplus 1,\ \sigma_2 \oplus 1\} \qquad \text{respectively} \qquad \mathrm{Lin}\{\sigma_1 \oplus 0,\ \sigma_2 \oplus 1\}$$

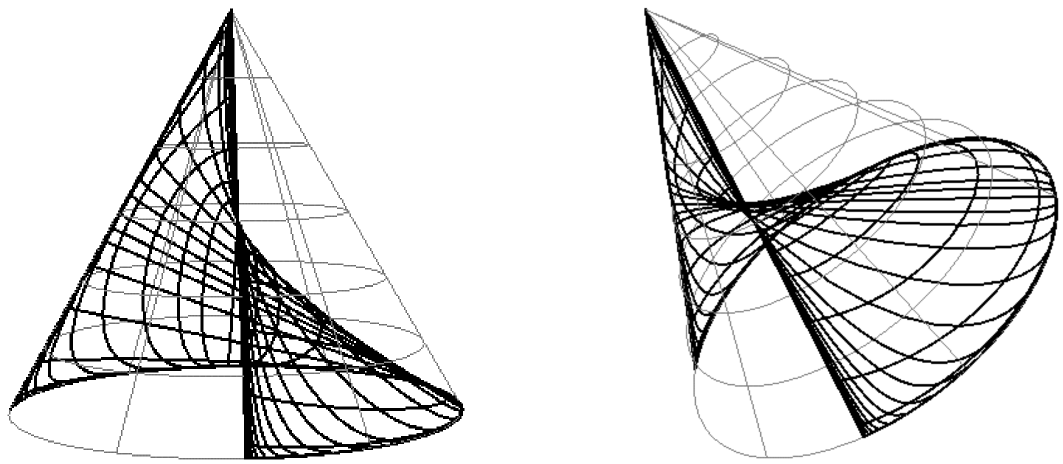


Figure 6.2: The exponential family $\mathcal{E} = \exp_1(\mathrm{Lin}\{\sigma_1 \oplus 1,\ \sigma_2 \oplus 1\})$ (**swallow family**) is depicted isometrically in two different views. Here we use the Pauli matrices $\sigma_1 = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ and $\sigma_2 = \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right)$. A wire frame consisting of one-dimensional exponential sub-families of $\mathcal{E}$ is used to indicate the shape. The frustum about $\mathcal{E}$ is the closure $\overline{S}(M_2 \oplus \mathbb{C}) \cap \mathrm{aff}(\mathcal{E})$ of the exponential family $\exp_1(\mathrm{aff}(\mathcal{E}))$. More details are provided in Example 6.4.

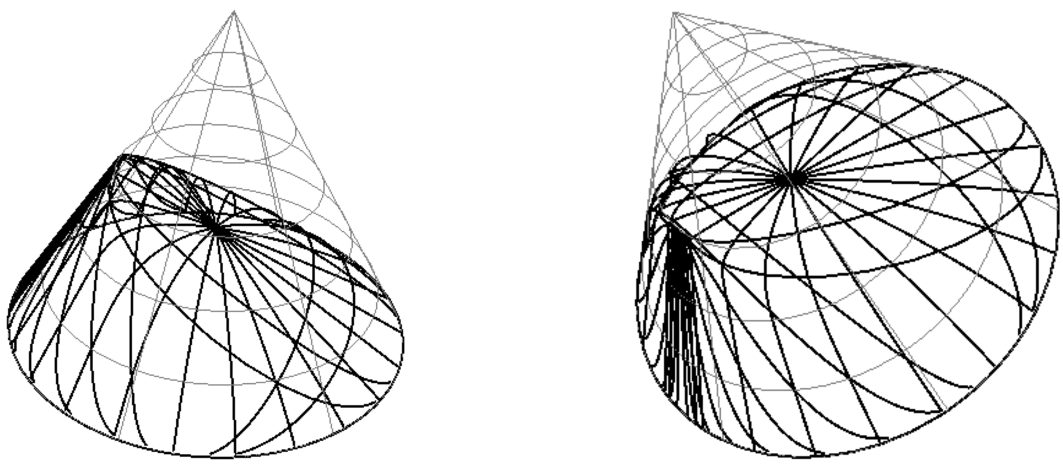


Figure 6.3: The exponential family $\mathcal{E} = \exp_1(\mathrm{Lin}\{\sigma_1 \oplus 0,\ \sigma_2 \oplus 1\})$ (**Staffelberg family**) is depicted isometrically in two different views. Other circumstances are the same as in Figure 6.2.

with Pauli matrices $\sigma_1 = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ and $\sigma_2 = \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right)$. To explain the generation of their shapes drawn in Figure 6.2 and Figure 6.3 we notice that $\exp_1$ is invariant under the additive group $\mathbb{R}\mathbb{1}$. We can use for $\mathcal{E}$ the domain V defined by

$$\mathrm{Lin}\{\sigma_1 \oplus 1 - \tfrac{\mathbb{1}}{3},\ \sigma_2 \oplus 1 - \tfrac{\mathbb{1}}{3}\} \qquad \text{respectively} \qquad \mathrm{Lin}\{\sigma_1 \oplus 0,\ \sigma_2 \oplus 1 - \tfrac{\mathbb{1}}{3}\}.$$

The space V was introduced in (5.36) respectively in (5.42). It has an angle $\varphi = \angle(V, z)$ with $z = -\frac{\mathbb{1}_2}{2} \oplus 1$ equal to $\varphi = \operatorname{arccot}(\sqrt{\frac{2}{3}})$ respectively equal to $\varphi = \frac{\pi}{3}$. We project the family $\mathcal{E}$ on the ONB $\{v_1, v_2 \sin(\varphi), v_3 \sin(\varphi)\}$ from (5.20) and (5.21). In the first case we use the parametrization

$$\mathbb{R}^2 \to \mathcal{E}, \qquad (\lambda, \mu) \mapsto \rho(\lambda, \mu) := \exp_1(\lambda(\sigma_1 \oplus 1) + \mu(\sigma_2 \oplus 1)).$$

With $t := \sqrt{\lambda^2 + \mu^2}$ and $T_1 := 2\cosh(t) + e^{\lambda+\mu}$ the coefficients are

$$\begin{aligned} \langle v_1,\ \rho(\lambda, \mu)\rangle &= \tfrac{\sinh(t)(\lambda-\mu)}{T_1 t}, \\ \langle v_2 \sin(\varphi),\ \rho(\lambda, \mu)\rangle &= \tfrac{1}{\sqrt{15}T_1}\left(\tfrac{3\sinh(t)(\lambda+\mu)}{t} - 2\cosh(t) + 2e^{\lambda+\mu}\right), \\ \text{and} \quad \langle v_3 \sin(\varphi),\ \rho(\lambda, \mu)\rangle &= \tfrac{\sqrt{2}}{\sqrt{5}T_1}\left(-\tfrac{\sinh(t)(\lambda+\mu)}{t} - \cosh(t) + e^{\lambda+\mu}\right). \end{aligned}$$

In the second case we use the parametrization

$$\mathbb{R}^2 \to \mathcal{E}, \qquad (\lambda, \mu) \mapsto \sigma(\lambda, \mu) := \exp_1(\lambda(\sigma_1 \oplus 0) + \mu(\sigma_2 \oplus 1)).$$

With $T_2 := 2\cosh(t) + e^{\mu}$ and t as before, the coefficients are

$$\begin{aligned} \langle v_1,\ \sigma(\lambda, \mu)\rangle &= \tfrac{\sqrt{2}\sinh(t)\lambda}{T_2 t}, \\ \langle v_2 \sin(\varphi),\ \sigma(\lambda, \mu)\rangle &= \tfrac{1}{\sqrt{6}T_2}\left(\tfrac{3\sinh(t)\mu}{t} - \cosh(t) + e^{\mu}\right), \\ \text{and} \quad \langle v_3 \sin(\varphi),\ \sigma(\lambda, \mu)\rangle &= \tfrac{1}{\sqrt{2}T_2}\left(-\tfrac{\sinh(t)\mu}{t} - \cosh(t) + e^{\mu}\right). \end{aligned}$$

This coordinate system is the best choice for $\mathcal{E}$. The vectors v_1 and $v_2 \sin(\varphi)$ span the parameter space V (5.19). Together with $v_3 \sin(\varphi)$ the vectors span $V + \mathbb{R}z$ (5.21). For $U := \mathrm{Lin}\{\sigma_1 \oplus 0, \sigma_2 \oplus 0, \sigma_3 \oplus 0\}$ we put $W := \pi_U(V) = \mathrm{Lin}\{\sigma_1 \oplus 0, \sigma_2 \oplus 0\}$. Then by (5.26) follows $V + \mathbb{R}z = W + \mathbb{R}z$ and by (6.5) we have for the frustum $C := \overline{S}(M_2 \oplus \mathbb{C}) \cap \left(\frac{\mathbb{1}}{3} + W + \mathbb{R}z\right)$ the inclusion

$$\mathcal{E} \subset \exp_1(\mathrm{lin}(C)) = \exp_1(\mathrm{aff}(C)) = \mathrm{ri}(C).$$

One has $\mathrm{lin}(C) = W + \mathbb{R}z$ so $v_1, v_2 \sin(\varphi), v_3 \sin(\varphi)$ is an ONB for $\mathrm{lin}(C)$. This is the smallest space to describe $\mathcal{E}$ isometrically, because $\mathrm{aff}(\mathcal{E}) = \mathrm{aff}(C)$ as will be shown below. In Figure 6.2 and Figure 6.3, C is drawn about the exponential family $\mathcal{E}$.

Together with the trace state $\rho(0,0)=\sigma(0,0)=\exp_1(0)=\frac{1}{3}(1,1,1)^{\mathrm{t}}$ we have the affinely independent coordinate triples with respect to $\{v_1, v_2\sin(\varphi), v_3\sin(\varphi)\}$

$$\begin{array}{rclcrcl}\rho(\ln(2),0) & = & \frac{1}{6}(1,\sqrt{\frac{5}{3}},0)^{\mathrm{t}}, & & \sigma(\ln(2),0) & = & \frac{1}{14}(3\sqrt{2},-\frac{1}{\sqrt{6}},-\frac{1}{\sqrt{2}})^{\mathrm{t}},\\ \rho(0,\ln(2)) & = & \frac{1}{6}(-1,\sqrt{\frac{5}{3}},0)^{\mathrm{t}}, & \text{resp.} & \sigma(0,\ln(2)) & = & (0,\frac{\sqrt{2}}{3\sqrt{3}},0)^{\mathrm{t}},\\ \rho(0,-\ln(2)) & = & \frac{1}{4}(1,-\sqrt{\frac{5}{3}},0)^{\mathrm{t}}, & & \sigma(0,-\ln(2)) & = & (0,-\frac{1}{\sqrt{6}},0)^{\mathrm{t}}.\end{array}$$

6.2 Geometry of relative entropy

Charts for the space $S(A)$ of invertible density matrices are introduced and their connection to the geometry of relative entropy is discussed including the BKM metric and the Pythagorean theorem of relative entropy. This section is an excerpt from the literature.

The real-analytic manifold $S(A)$ of invertible density matrices is diffeomorphic to the space A^0_{sa} of traceless self-adjoint matrices under the real-analytic map $\ln_0$ (6.2)

$$S(A)\cong A^0_{\mathrm{sa}}.$$

Throughout the section we fix a state $\sigma\in S(A)$ and put $\theta:=\ln_0(\sigma)\in A^0_{\mathrm{sa}}$ as well as $Q:=\ln(\sigma)\in A_{\mathrm{sa}}$. We fix tangent vectors $u,v\in \mathrm{T}_\sigma\, S(A)$ at σ.

Definition 6.5 The global chart $(S(A),\ln_0)$ is the **canonical chart** of $S(A)$. The representation of u in the canonical chart is denoted $u^{(\Theta)}$ and $u^{(\Theta)}$ is the **canonical representation** of u. The canonical representation of the tangent space $\mathrm{T}_\sigma\, S(A)$ is denoted $\mathrm{T}^{(\Theta)}_\theta\, S(A)=A^0_{\mathrm{sa}}$ and it is called the **canonical tangent space** of $S(A)$ (at σ).

Definition 6.6 **Free energy** [Ru] is the real analytic function

$$F:\qquad A_{\mathrm{sa}}\to\mathbb{R},\qquad a\mapsto \ln(\mathrm{tr}(e^a)). \tag{6.6}$$

Remark 6.7 **(a)** Free energy has for $a\in A$ and $\lambda\in\mathbb{R}$ the equivariant functional equation $F(a+\lambda\mathbb{1})=F(a)+\lambda$.

(b) In a generalization of the well-known formula $\frac{\partial}{\partial t}e^{at}=ae^{at}$ the derivative of the exponential function for $a, b \in A$ is

$$\frac{\partial}{\partial t}|_{t=0} e^{a+tb} = \int_0^1 e^{ya} b e^{(1-y)a} \, \mathrm{d}\, y, \tag{6.7}$$

where a and b may not commute, see page 127 in [Lie]. This formula can be proved by polynomial expansion. For $r, s \in \mathbb{N}_0$ the integral $\int_0^1 y^r (1-y)^s \, \mathrm{d}\, y = \frac{r!s!}{(r+s+1)!}$ may be used for the calculation.

(c) The derivative (6.7) can be used to differentiate free energy. For arbitrary $a \in A_{\mathrm{sa}}$ we have with $\sigma = \exp_1(\theta)$

$$\frac{\partial}{\partial t}|_{t=0} F(\theta + ta) = \langle \sigma, a \rangle. \tag{6.8}$$

(d) In this section the free energy is considered a function on the manifold $S(A)$ and its Hessian form is used below as a definition for the famous Bogoliubov-Kubo-Mori Riemannian metric. For $u, v \in \mathrm{T}_\sigma\, S(A)$ one has in canonical representation

$$\mathrm{D}^2\, F(u,v)|_\sigma = \frac{\partial^2}{\partial s\, \partial t}|_{s=t=0} F(\theta + s u^{(\Theta)} + t v^{(\Theta)}) = \langle \frac{\partial}{\partial t}|_{t=0} \exp_1(\theta + t v^{(\Theta)}), u^{(\Theta)} \rangle. \tag{6.9}$$

Definition 6.8 The ***m*-chart** is the identity mapping $(S(A), \sigma \mapsto \sigma)$. The representation of u in the m-chart is the ***m*-representation** of u denoted by $u^{(m)}$. The m-representation of the tangent space $\mathrm{T}_\sigma\, S(A)$ is denoted $\mathrm{T}_\sigma^{(m)}\, S(A)$.

Remark 6.9 **(a)** Notice that $\mathrm{T}_\sigma^{(m)}\, S(A) = A_{\mathrm{sa}}^0$. The chart change for a tangent vector from the canonical to the m-representation can be calculated with (6.7). The result is

$$u^{(m)} = \frac{\partial}{\partial t}|_{t=0} \exp_1(\theta + u^{(\Theta)}) = \int_0^1 \sigma^y u^{(\Theta)} \sigma^{1-y} \, \mathrm{d}\, y - \langle \sigma, u^{(\Theta)} \rangle \sigma. \tag{6.10}$$

(b) Using (6.9) and (6.10) we get for the Hessian form of the free energy for $u, v \in \mathrm{T}_\sigma\, S(A)$

$$\mathrm{D}^2\, F(u,v)|_\sigma = \langle u^{(\Theta)}, v^{(m)} \rangle = \int_0^1 \mathrm{tr}(u^{(\Theta)} \sigma^y v^{(\Theta)} \sigma^{1-y}) \, \mathrm{d}\, y - \langle \sigma, u^{(\Theta)} \rangle \langle \sigma, v^{(\Theta)} \rangle. \tag{6.11}$$

(c) In the classical case, that is for commuting matrices $\sigma, u^{(\Theta)}, v^{(\Theta)}$, the canonical representation of the Hessian of the free energy reduces to the covariance

$$\mathrm{D}^2\, F(u,v)|_\sigma = \mathbb{E}_\sigma \left[u^{(\Theta)} v^{(\Theta)} \right] - \mathbb{E}_\sigma \left[u^{(\Theta)} \right] \mathbb{E}_\sigma \left[v^{(\Theta)} \right]$$

of $u^{(\Theta)}$ and $v^{(\Theta)}$ with respect to the measurement probabilities of σ, cf. (1.23).

Definition 6.10 The ***e*-representation** $(S(A), \ln)$ of $S(A)$ is given by the logarithm $\ln : S(A) \to A_{\text{sa}}$. It is not a chart. The associated representation of u is the ***e*-representation** of u and it is denoted $u^{(e)}$.

Remark 6.11 **(a)** The e-representation is half-way between the canonical representation and the m-representation of $S(A)$. We write two chart changes for tangent representations. Since $\ln(\exp_1(\theta)) = \theta - F(\theta)\mathbb{1}$ we get from (6.8)

$$u^{(e)} = \frac{\partial}{\partial t}|_{t=0} \ln \exp_1(\theta + tu^{(\Theta)}) = u^{(\Theta)} - \langle \sigma, u^{(\Theta)} \rangle \mathbb{1}. \tag{6.12}$$

From (6.7) follows with $Q = \ln(\sigma)$

$$u^{(m)} = \frac{\partial}{\partial t}|_{t=0} \exp(Q + tu^{(e)}) = \int_0^1 \sigma^y u^{(e)} \sigma^{1-y} \, \mathrm{d}\, y. \tag{6.13}$$

(b) The Hessian form of the free energy (6.11) transforms with (6.12) to the expression $\mathrm{D}^2 \, F(u,v)|_\sigma = \langle u^{(\Theta)}, v^{(m)} \rangle = \langle u^{(e)}, v^{(m)} \rangle$ because $\mathrm{tr}(v^{(m)}) = 0$. With (6.13) we get the e-representation of the Hessian of the free energy

$$\mathrm{D}^2 \, F(u,v)|_\sigma = \langle u^{(e)}, v^{(m)} \rangle = \int_0^1 \mathrm{tr}(u^{(e)} \sigma^y v^{(e)} \sigma^{1-y}) \, \mathrm{d}\, y. \tag{6.14}$$

We notice, for an abelian algebra one obtains $\mathrm{D}^2 \, F(u,v)|_\sigma = \mathbb{E}_\sigma \left[u^{(e)} v^{(e)} \right]$.

(c) The Hessian of the free energy is positive definite throughout A^0_{sa}. We obtain from the e-representation (6.14) for non-zero $u \in \mathrm{T}_\sigma \, S(A)$

$$\begin{aligned} \mathrm{D}^2 \, F(u,u)|_\sigma &= \textstyle\int_0^1 \mathrm{tr} \left(\sigma^{\frac{y}{2}} u^{(e)} \sigma^{\frac{1-y}{2}} \right)^* \left(\sigma^{\frac{y}{2}} u^{(e)} \sigma^{\frac{1-y}{2}} \right) \mathrm{d}\, y \\ &= \textstyle\int_0^1 \| \sigma^{\frac{y}{2}} u^{(e)} \sigma^{\frac{1-y}{2}} \|_2^2 \, \mathrm{d}\, y > 0 \end{aligned} \tag{6.15}$$

since σ is invertible. This implies that free energy F in canonical representation is a strictly convex function on A^0_{sa}.

Von Neumann entropy is a measure of disorder for a density matrix. It is a fundamental quantity in information theory describing compression rates for quantum source coding, see Remark 1.1.

Definition 6.12 The **von Neumann entropy** is defined for $\rho \in \overline{S}(A)$ by

$$S(\rho) := -\operatorname{tr}(\rho \ln(\rho)). \tag{6.16}$$

Remark 6.13 Von Neumann entropy is continuous and strictly concave on the state space $\overline{S}(A)$ (for a finite dimensional algebra). In infinite dimensions continuity generalizes to the weaker property of insensitivity, see page 237 in [Weh].

Quantum relative entropy is a fundamental quantity in information theory. It is a natural distance measure between density matrices describing asymptotic errors in hypothesis testing, see Remark 1.1.

Definition 6.14 Relative entropy is defined for two density matrices $\rho, \tau \in \overline{S}(A)$ by

$$S(\rho, \tau) := \begin{cases} \operatorname{tr}(\rho(\ln(\rho) - \ln(\tau))) & \text{if } s(\rho) \leq s(\tau), \\ \infty & \text{otherwise} \end{cases} \tag{6.17}$$

with $s(\rho), s(\tau)$ denoting the support projectors. For technical reasons we use also the notation with the first argument fixed

$$S_\rho(\tau) := S(\rho, \tau). \tag{6.18}$$

Remark 6.15 For singular states one uses the convention $0 \ln(0) = 0$ justified by invariance of relative entropy under algebra embedding. For example, this rule is used in part (a) of the proof to Theorem 6.

The distance-like properties of relative entropy for $\rho, \tau \in \overline{S}(A)$

$$S(\rho, \tau) \geq 0 \qquad \text{and} \qquad S(\rho, \tau) = 0 \iff \rho = \tau, \tag{6.19}$$

are proved in [Weh], pages 232 and 250. Relative entropy is joint convex, for $\rho_i, \tau_i \in \overline{S}(A)$, $i = 1, 2$ and $t \in [0, 1]$

$$S\left(t\rho_1 + (1-t)\rho_2,\ t\tau_1 + (1-t)\tau_2\right) \leq tS(\rho_1, \tau_1) + (1-t)S(\rho_2, \tau_2)$$

holds. This is proved on page 250 in [Weh] and the proof depends on Lieb's concavity theorem [Lie]. It is proved on page 251 in [Weh] that the relative entropy is lower semi-continuous on $\overline{S}(A) \times \overline{S}(A)$.

Example 6.16 Relative entropy is not continuous on the domain of finiteness. We can consider the algebra $A := \mathbb{C}^2$. For sequences of density matrices $\rho_n := (1-\frac{1}{n}, \frac{1}{n})$ and $\sigma_n := (1-s_n, s_n)$ with limit $\lim_{n\to\infty} s_n = 0$ we have

$$S(\lim_{n\to\infty} \rho_n, \lim_{n\to\infty} \sigma_n) = S((1,0),(1,0)) = 0$$

and[1]

$$\lim_{n\to\infty} S(\rho_n, \sigma_n) = -\lim_{n\to\infty} \frac{\ln(s_n)}{n} + o(1).$$

With the special choices of sequences $s_n^0 := e^{-\sqrt{n}}$, $s_n^\alpha := e^{-\alpha n}$ for $\alpha > 0$ and $s_n^\infty := e^{-n^2}$ every non-negative real number including $+\infty$ is a limit of $S(\rho_n, \sigma_n)$ for $n \to \infty$.

Remark 6.17 Relative entropy, von Neumann entropy and free energy are connected. For $\rho \in \overline{S}(A)$ and $\theta = \ln_0(\sigma)$ we have

$$S(\rho, \sigma) = -S(\rho) - \langle \rho, \theta \rangle + F(\theta). \tag{6.20}$$

We deduce from (6.8) the derivative for invertible second argument in the canonical chart for the relative entropy

$$\mathrm{D}\, S_\rho(u)|_\sigma = \langle \sigma - \rho, u^{(\Theta)} \rangle. \tag{6.21}$$

The Hessian form is

$$\mathrm{D}^2\, S_\rho(u,v)|_\sigma = \mathrm{D}^2\, F(u,v)|_\sigma. \tag{6.22}$$

The BKM-metric is known to be a Riemannian metric (see the proof (6.15) that it is positive definite). The acronym BKM stands for Bogoliubov, Kubo and Mori. In the classical case it reduces to Fisher's metric. It is a fundamental concept in quantum information theory, see Remark 1.2.

Definition 6.18 The **BKM-metric** on the real-analytic manifold $S(A)$ is defined by [Pe94]

$$\langle u, v \rangle_\sigma := \mathrm{D}^2\, S_\rho(u,v)|_\sigma = \mathrm{D}^2\, F(u,v)|_\sigma. \tag{6.23}$$

[1] The equality $f(n) = o(1)$ for a real valued function f means that $\lim_{n\to\infty} f(n) = 0$.

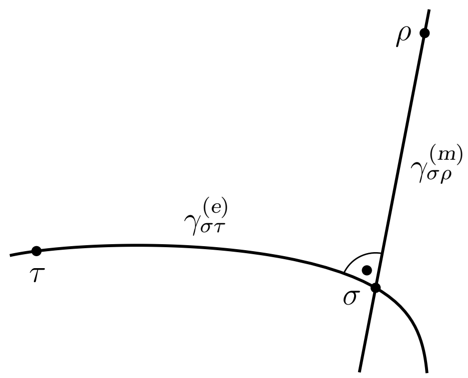


Figure 6.4: For $\rho \in \overline{S}$ and $\sigma, \tau \in S(A)$ the Pythagorean theorem of relative entropy $S(\rho, \sigma) + S(\sigma, \tau) = S(\rho, \tau)$ holds if the e-geodesic $\gamma^{(e)}_{\sigma\tau}$ and the m-geodesic $\gamma^{(m)}_{\sigma\rho}$ meet perpendicularly with respect to the BKM-metric.

Remark 6.19 Relative entropy and the BKM-metric are connected geometrically [Pe94]. In addition to $\sigma \in S(A)$ we choose two more density matrices $\rho \in \overline{S}(A)$ and $\tau \in S(A)$ invertible. The ***m*-geodesic** linking σ to ρ is the curve for $t \in [0, 1]$

$$\gamma^{(m)}_{\sigma\rho}(t) = \sigma + t(\rho - \sigma). \tag{6.24}$$

The ***e*-geodesic** linking σ to τ is the curve for $t \in [0, 1]$

$$\gamma^{(e)}_{\sigma\tau}(t) = \exp_1(\ln_0(\sigma) + t(\ln_0(\tau) - \ln_0(\sigma))). \tag{6.25}$$

We denote the tangent vectors at σ for the curves by $u, v \in T_\sigma S(A)$ respectively. We obtain

$$u^{(m)} = \frac{\partial}{\partial t}|_{t=0}\, \gamma^{(m)}_{\sigma\rho}(t) = \rho - \sigma \qquad \text{and} \qquad v^{(\Theta)} = \frac{\partial}{\partial t}|_{t=0}\, \gamma^{(\Theta)}_{\sigma\tau}(t) = \ln_0(\tau) - \ln_0(\sigma).$$

The BKM-metric (6.23) evaluated under the two vectors is (6.11)

$$\langle u, v\rangle_\sigma = \langle u^{(m)}, v^{(\Theta)}\rangle = \langle \rho - \sigma, \ln_0(\tau) - \ln_0(\sigma)\rangle = \langle \rho - \sigma, \ln(\tau) - \ln(\sigma)\rangle$$

so that (see, e.g., Petz [Pe08])

$$S(\rho, \sigma) + S(\sigma, \tau) = S(\rho, \tau) + \langle u, v\rangle_\sigma.$$

The perpendicular case $\langle u, v\rangle_\sigma = 0$ is known as the **Pythagorean theorem** of relative entropy,

$$S(\rho, \sigma) + S(\sigma, \tau) = S(\rho, \tau). \tag{6.26}$$

An example of the Pythagorean theorem of relative entropy is shown in Figure 6.4.

6.3 The mean value chart

We introduce the mean value chart for an exponential family in a matrix algebra. The geometry of the mean value chart is described by cylinder and domain for the exponential family. This is a generalization from probability theory [Bar, Cs03] to finite-level quantum systems. A set of mean values

$$\langle a_1, \rho \rangle, \dots, \langle a_k, \rho \rangle$$

of observables $a_1, \dots, a_k$ was used before by Wichmann [Wic] to parametrize maximum entropy ensembles $\rho \in \overline{S}(A)$. Instead of the mean values we argue with the orthogonal projection of ρ to a vector space, that is, with a state reflection. The two are equivalent since linked by a linear mapping (Remark 6.32), our choice is coordinate free. In this section we do not go far beyond Wichmann's reasoning. In place of maximum entropy arguments we use the Pythagorean theorem of relative entropy to prove injectivity of the mean value chart (Lemma 6.22). This gives us a wider range of applications. Our original contribution in this section is a description of the domain of an exponential family in terms of a projector lattice (Corollary 6.31).

Definition 6.20 Let $\mathcal{E}$ be an exponential family in A. The affine space $\Theta := \ln_0(\mathcal{E}) \subset A^0_{\text{sa}}$ is the **canonical parameter space** of $\mathcal{E}$. The chart $(\mathcal{E}, \ln_0)$ is the **canonical chart** for $\mathcal{E}$. The inverse $\exp_1|_\Theta : \Theta \to \mathcal{E}$ is the **canonical parametrization** of $\mathcal{E}$.

We can use the canonical representation of $S(A)$ in Definition 6.5. Then the tangent space $\mathrm{T}_\sigma \, \mathcal{E}$ at $\sigma \in \mathcal{E}$ is represented as the translation vector space of Θ. For $\theta := \ln_0(\sigma) \in \Theta$ we have $\mathrm{T}^{(\Theta)}_\theta \, \mathcal{E} = \mathrm{lin}(\Theta) \subset A^0_{\text{sa}}$.

Definition 6.21 The **canonical tangent space** of $\mathcal{E}$ is $U := \mathrm{lin}(\Theta)$. The **cylinder** on $\mathcal{E}$ is $\mathrm{Cyl} := \mathcal{E} + U^\perp$. The **domain** of $\mathcal{E}$ is $\mathrm{Dom} := \overline{S}(A) \cap \mathrm{Cyl}$. The **traceless complement** of a vector space $V \subset A_{\text{sa}}$ is defined by $V^{\perp,0} := V^\perp \cap A^0_{\text{sa}} = \{v \in V^\perp : \mathrm{tr}(v) = 0\}$.

Unless otherwise specified we use Θ, U, Cyl resp. Dom as the canonical parameter space, canonical tangent space, cylinder resp. domain of an exponential family $\mathcal{E}$ in a matrix algebra A throughout and without reference to $\mathcal{E}$.

Lemma 6.22 For $a \in \mathrm{Cyl}$ the intersection $\mathcal{E} \cap (a + U^{\perp})$ has exactly one element. If $\rho \in \mathrm{Dom}$ then the intersection $\mathcal{E} \cap (\rho + U^{\perp,0})$ is transverse with respect to the BKM-metric. The relative entropy S_ρ has a unique minimum on $\mathcal{E}$ at $\mathcal{E} \cap (\rho + U^{\perp})$, identifying a one-element set with the element. [Proof on page 224]

Definition 6.23 Using Lemma 6.22 the **normal projection** for $\mathcal{E}$ is defined by

$$N : \mathrm{Cyl} \to \mathcal{E}, \qquad a \mapsto \mathcal{E} \cap (a + U^{\perp}), \tag{6.27}$$

identifying a one-element set with the element. The **mean value chart** for $\mathcal{E}$ is defined by restriction to $\mathcal{E}$ of the orthogonal projection to the canonical tangent space

$$\pi := \pi_U|_{\mathcal{E}} \ : \quad \mathcal{E} \to U. \tag{6.28}$$

The **mean value parametrization** for $\mathcal{E}$ is defined by restriction of the normal projection

$$M := N|_{\mathrm{Cyl}\cap U} \ : \quad \mathrm{Cyl} \cap U \to \mathcal{E}. \tag{6.29}$$

Remark 6.24 Existence of the normal projection for $\mathcal{E}$ gives the correct impression that the projection on the canonical tangent space U defines a suitable chart for $\mathcal{E}$. The image of $\mathcal{E}$ on U is included in the relative interior of the state reflection sr_U

$$\pi_U(\mathcal{E}) \subset \mathrm{ri}(\mathrm{sr}_U)$$

because $\mathcal{E}$ is a subset of the relative interior of the state space and because an affine map commutes with reduction to the relative interior (3.15). We will analyze $\mathcal{E}$ in the convex geometric framework of a state reflection developed in Chapter 5.

Lemma 6.25 Let $(x_i) \subset A_{\mathrm{sa}}$ be a sequence diverging in modulus $\lim_{i\to\infty} \|x_i\| = \infty$ and let u be an accumulation point of $(\frac{x_i}{\|x_i\|})$. For $i \in \mathbb{N}$ we put $s_i := x_i - \mu_+(x_i)\mathbb{1}$ with the maximal eigenvalue $\mu_+(x_i)$ of x_i. Then every accumulation point z of (e^{s_i}) satisfies $s(z) \leq p_+(u)$ and every accumulation point ρ of $(\exp_1(s_i))$ satisfies $s(\rho) \leq p_+(u)$.
[Proof on page 225]

The following corollary and lemma are used in [Wic] without proof. For completeness the proofs are included here.

Corollary 6.26 If $\rho \in \overline{\mathcal{E}} \setminus \mathcal{E}$ then $s(\rho) \leq p$ for a proper exposed projector $p \in \mathcal{P}_{U,\perp}$.
[Proof on page 226]

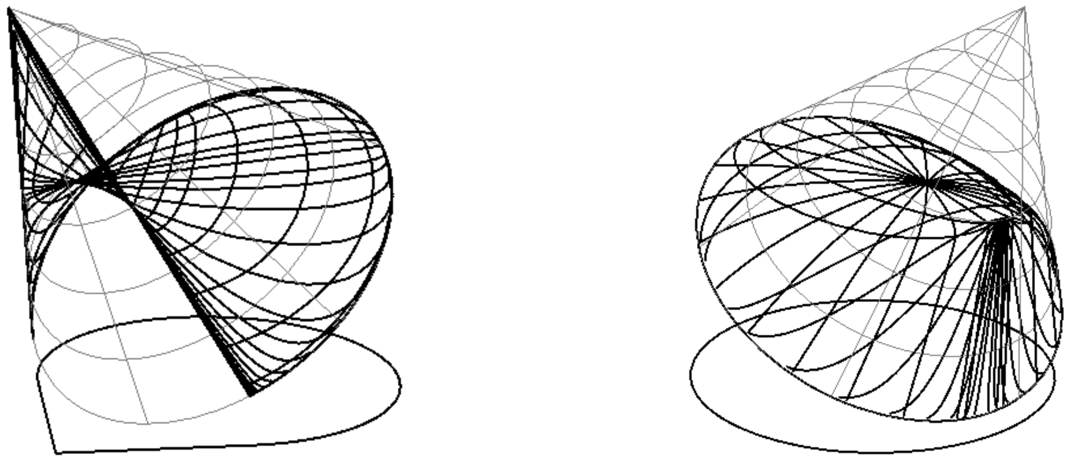


Figure 6.5: The swallow family (left) and Staffelberg family (right) are depicted including the boundary curve of the state reflection on the canonical tangent space. More details are provided in Example 6.29.

Remark 6.27 Every proper projector in $\overline{\mathcal{P}_U}$ is the support projector of a point $\rho \in \overline{\mathcal{E}} \setminus \mathcal{E}$. We can proof this in Lemma 7.24. Whether there are further projectors that can be a support projector of a point $\rho \in \overline{\mathcal{E}} \setminus \mathcal{E}$ is not clear at the moment. The topology of $\mathcal{P}_U$ will be related to the topology of an exponential family in Section 7.2.

Lemma 6.28 Let $K \subset \mathbb{R}^m$ be bounded and $L \subset \mathbb{R}^n$ be connected. Let $f : \overline{K} \to \overline{L}$ be continuous and $f(K)$ be open. If $f(K) \cap L \neq \emptyset$ and $f(\overline{K} \setminus K) \cap L = \emptyset$ then $f(K) \supset L$.
[Proof on page 226]

Theorem 4 The mean value chart $(\mathcal{E}, \pi)$ is a chart with range $\pi(\mathcal{E}) = \mathrm{ri}(\mathrm{sr}_U)$. Furthermore one has $\pi_U(\overline{\mathcal{E}}) = \mathrm{sr}_U$ and $\pi_U(\overline{\mathcal{E}} \setminus \mathcal{E}) = \mathrm{rb}(\mathrm{sr}_U)$. [Proof on page 226]

Example 6.29 The exponential families in Figure 6.5 are depicted together with their state reflections. Using a translation along $\frac{\mathbb{1}}{3}$ and an adjustment orthogonal to the canonical tangent space U the state reflection $\mathrm{sr}_U \subset U$ is moved into the drawing frame of the picture. This is possible because U belongs to the translation vector space of the exponential family, see Example 6.4. The state reflections alone are also depicted in Figure 4.1, lower middle, and in Figure 5.5.

Corollary 6.30 The mean value parametrization $M : \mathrm{ri}(\mathrm{sr}_U) \to \mathcal{E}$ is the inverse diffeomorphism to the mean value chart π and the normal projection of $\mathcal{E}$ is $N = M \circ \pi_U|_{\mathrm{Cyl}}$. [Proof on page 227]

The projector lattice $\mathcal{P}_U$ of the state reflection sr_U can be used to describe the domain.

Corollary 6.31 The cylinder on $\mathcal{E}$ is $\mathrm{Cyl} = \mathrm{ri}(\mathrm{sr}_U) + U^\perp = S(A) + U^\perp$, the domain of $\mathcal{E}$ is $\mathrm{Dom} = \{\rho \in \overline{S}(A) : s(\rho) \not\leq p \text{ for all proper } p \in \mathcal{P}_U\}$. [Proof on page 227]

Remark 6.32 (Statistics and convex support) When a finite number of self-adjoint matrices $a_0, \dots, a_k \in A_{\mathrm{sa}}$ are given then we can consider the exponential family

$$\mathcal{E} := \exp_1\left(a_0 + \mathrm{Lin}\{a_1, \dots, a_k\}\right).$$

The tuple $a_1, \dots, a_k$ is the **statistic** of $\mathcal{E}$ and with $\pi_{A^0_{\mathrm{sa}}}(a_i) = a_i - \frac{\mathrm{tr}(a_i)}{\mathrm{tr}(\mathbb{1})}\mathbb{1}$ for $i = 1, \dots, k$ the canonical tangent space of $\mathcal{E}$ is given by

$$U = \pi_{A^0_{\mathrm{sa}}}\left(\mathrm{Lin}\{a_1, \dots, a_k\}\right) = \mathrm{Lin}\left\{\pi_{A^0_{\mathrm{sa}}}(a_i)\right\}_{i=1}^k.$$

Let us introduce the mean value mapping

$$m : A_{\mathrm{sa}} \to \mathbb{R}^k, \qquad a \mapsto \left(\langle a_1, a\rangle, \dots, \langle a_k, a\rangle\right) \tag{6.30}$$

and consider the set of mean values $m(\mathcal{E}) := \{m(\rho) : \rho \in \mathcal{E}\} \subset \mathbb{R}^k$. We can prove that the following diagram commutes.

$$\begin{tikzcd}
\mathcal{E} \arrow[r, shift left, "\pi_U"] \arrow[dr, "m"'] & \mathrm{ri}(\mathrm{sr}_U) \arrow[l, shift left] \arrow[d, shift left, "\alpha"] \\
& m(\mathcal{E}) \arrow[u, shift left]
\end{tikzcd} \tag{6.31}$$

The projection $\pi_U|_{\mathcal{E}} : \mathcal{E} \to \mathrm{ri}(\mathrm{sr}_U)$ is simply the mean value chart in Theorem 4 which is a diffeomorphism. With the fitting mapping $\alpha : U \to \mathbb{R}^k$, $u \mapsto \left\{\langle a_i, u\rangle + \frac{\mathrm{tr}(a_i)}{\mathrm{tr}(\mathbb{1})}\right\}_{i=1}^k$, the labeled arrows commute. Truely, for $a \in A^1_{\mathrm{sa}}$ and $i \in \{1, \dots, k\}$ we have

$$\langle a_i, a\rangle - \frac{\mathrm{tr}(a_i)}{\mathrm{tr}(\mathbb{1})} = \langle a_i - \frac{\mathrm{tr}(a_i)}{\mathrm{tr}(\mathbb{1})}\mathbb{1}, a\rangle = \langle \pi_U(a_i), a\rangle = \langle a_i, \pi_U(a)\rangle.$$

The affine mapping α is invertible on $\alpha(U)$ because for an ONB $\{x_i\}_{i=1}^{\dim(U)}$ of U

$$\dim(m(\mathcal{E})) = \dim(\mathrm{Im}(\alpha)) = \mathrm{rk}\left(\langle x_i, a_j\rangle\right) = \dim\left(\pi_U\left(\mathrm{Lin}\{a_1, \dots, a_k\}\right)\right) = \dim(U)$$

holds. Hence we have proved that $m(\mathcal{E})$ and $\mathrm{ri}(\mathrm{sr}_U)$ are isomorphic under the affine isomorphism α. The closure $\overline{m(\mathcal{E})}$ is called **convex support** of the statistics $a_1, \dots, a_k$. It generalizes the well-known convex support in the context of probability distributions, see Section 1.3. In particular we have proved that

$$\mathcal{E} \to m(\mathcal{E}), \qquad \rho \mapsto \{\langle a_i, \rho\rangle\}_{i=1}^k \tag{6.32}$$

is a diffeomorphism from $\mathcal{E}$ to the relative interior of convex support.

Using the derivative of free energy (6.8) we can give a condition on the real coefficients $\lambda_1, \dots, \lambda_k$ whether a density matrix $\exp_1(a_0 + \sum_{i=1}^k \lambda_i a_i)$ in $\mathcal{E}$ has certain mean values. Let $(\xi_1, \dots, \xi_k) \in m(\mathcal{E})$. Then for $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ we have

$$m\left(\exp_1(a_0 + \sum_{i=1}^k \lambda_i a_i)\right) = (\xi_1, \dots, \xi_k) \qquad \Longleftrightarrow \qquad \begin{array}{l} \frac{\partial}{\partial\lambda_j} F(a_0 + \sum_{i=1}^k \lambda_i a_i) = \xi_j \\ \text{for } j = 1, \dots, k. \end{array} \tag{6.33}$$

We know from the diffeomorphism $\exp_1|_{A^0_{\mathrm{sa}}} : A^0_{\mathrm{sa}} \to S(A)$ that points in $a_0+\mathrm{Lin}\{a_1, \dots, a_k\}$ that parametrize the same density matrix in $\mathcal{E}$ differ by a multiple of $\mathbb{1}$. The coefficients $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ with equal image $\exp_1(a_0 + \sum_{i=1}^k \lambda_i a_i)$ on $\mathcal{E}$ form affine subspaces of $\mathbb{R}^k$. The inverse (6.33) of the mean m has unique solutions if and only if $\{a_1, \dots, a_k, \mathbb{1}\}$ is a linearly independent set. Only in this case $\dim(U) = k$ holds.

7 Entropy distance

In this chapter we describe the *rI*-closure and the entropy distance for an exponential family $\mathcal{E}$ in a matrix algebra A. Applications are recalled in Section 1.3 for the abelian case of probability distributions. The main application in the non-abelian quantum case will be the description of multi-information in Section 8.3. We denote the canonical parameter space $\Theta = \ln_0(\mathcal{E})$. The canonical tangent space $U = \mathrm{lin}(\Theta)$ is the translation vector space of the affine space Θ.

Definition 7.1 The **entropy distance** from $\mathcal{E}$ is defined with relative entropy S by

$$S_{\mathcal{E}} : \qquad \overline{S}(A) \to \mathbb{R}, \qquad \rho \mapsto \inf_{\sigma \in \mathcal{E}} S(\rho, \sigma). \tag{7.1}$$

The **reverse information closure** or ***rI*-closure** of $\mathcal{E}$ is the extension

$$\mathrm{cl}_{rI}(\mathcal{E}) := \{\rho \in \overline{S}(A) : S_{\mathcal{E}}(\rho) = 0\}. \tag{7.2}$$

We recall that entropy distance is a bounded (notice $\mathcal{E} \neq \emptyset$) non-negative function on the state space. One has the global bounds $0 \leq S_{\mathcal{E}}(\rho) \leq \inf_{\sigma \in \mathcal{E}} \| \ln(\sigma) \| < \infty$ for all $\rho \in \overline{S}(A)$. Here we use the inequality $|\operatorname{tr}(\rho \ln(\sigma))| \leq \| \ln(\sigma) \| \operatorname{tr}(\rho)$ (2.12) and spectral norm $\| \cdot \|$.

Let us give a short overview of the three sections. In Section 7.1 we can extend the normal projection $N : \mathrm{Cyl} \to \mathcal{E}$ from the cylinder Cyl to the closure $\overline{\mathrm{Cyl}} = \overline{S}(A) + U^{\perp}$ as the mapping

$$N^{\mathrm{cmb}} : \overline{S}(A) + U^{\perp} \to \mathrm{cl}_{rI}(\mathcal{E}), \quad x \mapsto (x + U^{\perp}) \cap \mathrm{cl}_{rI}(\mathcal{E})$$

identifying a one-element set with the element and such that for arbitrary $\rho \in \overline{S}(A)$ we have $S_{\mathcal{E}}(\rho) = S(\rho, N^{\mathrm{cmb}}(\rho))$. The extension depends on asymptotics of e-geodesics.

In Section 7.2 we prove that the Staffelberg family has an *rI*-closure strictly included into the topological closure of the family. The topology of an exponential family is related to the topology of an associated projector lattice. From the examples we are inspired to formulate a number of question.

In Section 7.3 we present further new results on entropy distance. There exist exponential families where entropy distance is discontinuous. We can prove the equivalence of continuity to the *rI*-closure being topologically closed and we conjecture the positions of possible discontinuities. Local maximizers of entropy distance have a rank bound. The bound can improve quadratically in the quantum case compared to the classical case.

7.1 The reverse information closure

The progress in this section is the following. We define the combinatorial extension $\mathcal{E}^{\mathrm{cmb}}$ of $\mathcal{E}$ (7.4). This is a union of exponential families in compressed algebras. The mean value chart for $\mathcal{E}$ extends to the combinatorial mean value chart $\pi^{\mathrm{cmb}} = \pi_U|_{\mathcal{E}^{\mathrm{cmb}}}$, which is by Proposition 7.11 the bijection

$$\pi^{\mathrm{cmb}} : \mathcal{E}^{\mathrm{cmb}} \to \mathrm{sr}_U$$

defined by association of mean values with respect to the canonical tangent space U of $\mathcal{E}$. A left-inverse to π^{cmb} is the combinatorial normal projection (7.14)

$$N^{\mathrm{cmb}} : \mathrm{sr}_U + U^{\perp} \to \mathcal{E}^{\mathrm{cmb}}, \qquad x \mapsto (x + U^{\perp}) \cap \mathcal{E}^{\mathrm{cmb}}$$

identifying a one-element set with the element. The combinatorial extension of $\mathcal{E}$, though not a compact set, is the set in the state space $\overline{S}(A)$ where relative entropy (6.17) distance from $\mathcal{E}$ is minimized. We can establish in Theorem 6 for arbitrary $\rho \in \overline{S}(A)$ the unique minimum $N^{\mathrm{cmb}}(\rho)$ on $\mathcal{E}^{\mathrm{cmb}}$

$$S(\rho, N^{\mathrm{cmb}}(\rho)) = \inf_{\sigma \in \mathcal{E}} S(\rho, \sigma).$$

The minimum is found by curves in $\mathcal{E}^{\mathrm{cmb}}$ that are piecewise *e*-geodesic, the pieces corresponding to an access sequence for the state reflection sr_U (Definition 5.29).

Definition 7.2 The **compression** of $\mathcal{E}$ by a non-zero projector $p \in \mathcal{P}$ is defined as the exponential family in A^p

$$\mathcal{E}_p := \exp_1(\varsigma^p(\Theta)) \tag{7.3}$$

with the traceless compression $\varsigma^p : A_{\mathrm{sa}} \to (A^p)^0_{\mathrm{sa}}$, $a \mapsto (\kappa^p)^{-1} \circ \pi_{\mathrm{lin}(\mathbb{F}(p))}(a)$ defined in (5.60). In analogy to Definition 6.21 the cylinder on $\mathcal{E}_p$ is denoted Cyl_p and the domain of $\mathcal{E}_p$ is denoted Dom_p. We denote the normal projection for $\mathcal{E}_p$ (6.27) by N_p. The mean value parametrization for $\mathcal{E}_p$ (6.29) is denoted M_p and the mean value chart for $\mathcal{E}_p$ (6.28) is denoted $\pi_p := \pi_{\varsigma^p(U)}|_{\mathcal{E}_p}$. The **combinatorial extension** of $\mathcal{E}$ is the union

$$\mathcal{E}^{\mathrm{cmb}} := \bigcup_{p \in \mathcal{P}_U \setminus \{0\}} \kappa^p (\mathcal{E}_p) \tag{7.4}$$

of compressions $\kappa^p(\mathcal{E}_p)$ for non-zero projectors p in the projector lattice $\mathcal{P}_U$ with U the canonical tangent space of $\mathcal{E}$. The **combinatorial mean value chart** is

$$\pi^{\mathrm{cmb}} := \pi_U|_{\mathcal{E}^{\mathrm{cmb}}}. \tag{7.5}$$

Remark 7.3 Let $p \in \mathcal{P}$ be a non-zero projector. The canonical tangent space (Definition 6.21) of the compression $\mathcal{E}_p$ is the traceless compression

$$\mathrm{lin}(\varsigma^p(\Theta)) = \varsigma^p(\mathrm{lin}(\Theta)) = \varsigma^p(U). \tag{7.6}$$

The dimension of $\mathcal{E}_p$ is the dimension of a face reflection, $\dim(\mathcal{E}_p) = \dim(\varsigma^p(U)) = \dim(\mathbb{F}_U(p))$, (5.61). The canonical parametrization introduced in Definition 6.20 of $\mathcal{E}_p$ is the diffeomorphism $\exp_1 : \varsigma^p(\Theta) \to \mathcal{E}_p$. The pullback of the canonical parametrization under the mapping $\varsigma^p|_\Theta : \Theta \to \varsigma^p(\Theta)$ followed by a pushforward under κ^p is the surjective mapping

$$\Theta \to \kappa^p(\mathcal{E}_p), \quad \theta \mapsto \kappa^p\left(\exp_1\left(\varsigma^p(\theta)\right)\right) = \kappa^p\left(\exp_1\left((\kappa^p)^{-1}(p\theta p)\right)\right) = \frac{pe^{p\theta p}}{\mathrm{tr}(pe^{p\theta p})}. \tag{7.7}$$

Using the pullback under the invertible mapping $(\kappa^p)^{-1}$ instead of $\varsigma^p = (\kappa^p)^{-1} \circ \pi_{\mathrm{lin}(\mathbb{F}(p))}$ one obtains the diffeomorphism

$$\pi_{\mathrm{lin}(\mathbb{F}(p))}(\Theta) \to \kappa^p(\mathcal{E}_p), \quad \widetilde{\theta} \mapsto \frac{pe^{\widetilde{\theta}}}{\mathrm{tr}(pe^{\widetilde{\theta}})}. \tag{7.8}$$

We see that $\kappa^p(\mathcal{E}_p)$ has an existence independent of a particular choice of κ^p. The mean value chart (Theorem 4) for $\mathcal{E}_p$ is the diffeomorphism

$$\pi_p : \mathcal{E}_p \to \mathrm{ri}(\mathrm{sr}_{\varsigma^p(U)}) \tag{7.9}$$

with inverse the mean value parametrization M_p. Then by the third diagram in Proposition 5.20 one has the diffeomorphism

$$\pi_U|_{\kappa^p(\mathcal{E}_p)} : \kappa^p(\mathcal{E}_p) \to \mathrm{ri}(\mathbb{F}_U(p)). \tag{7.10}$$

By Corollary 6.31 the cylinder on $\mathcal{E}^p$ is

$$\mathrm{Cyl}_p = \mathcal{E}_p + \varsigma^p(U)^\perp = \mathrm{ri}(\mathrm{sr}_{\varsigma^p(U)}) + \varsigma^p(U)^\perp = S(A^p) + \varsigma^p(U)^\perp. \tag{7.11}$$

The normal projection is the differentiable map

$$N_p = M_p \circ \pi_{\varsigma^p(U)}|_{\mathrm{Cyl}_p} \tag{7.12}$$

defined for $a \in \mathrm{Cyl}_p$ by $N_p(a) = \mathcal{E}_p \cap (a + \varsigma^p(U)^\perp)$ identifying a one-element set with the element (6.27). With $p = \mathbb{1}$ we come back to the exponential family $\mathcal{E}$ in A.

We generalize the definition of an e-geodesic (6.25) to unbounded parameter intervals in canonical parametrization.

Definition 7.4 For $\theta, v \in A_{\mathrm{sa}}^0$ and a real interval I (possibly unbounded) we call the curve $I \to S(A)$, $t \mapsto \exp_1(\theta + tv)$ an **e-geodesic**. The image $\{\exp_1(\theta + tv) : t \in I\}$ is also called an e-geodesic.

Lemma 7.5 Let $\theta, u \in A_{\mathrm{sa}}$ and u with maximal eigenvalue $\mu_+(u) = 0$. Then for the kernel projector p of u we have $\lim_{\lambda\to\infty} e^{\theta+\lambda u} = pe^{p\theta p}$. [Proof on page 228]

Lemma 7.6 Let $\theta, u \in A_{\mathrm{sa}}$ and put $p := p_+(u)$ the maximal projector of u. Then

$$\lim_{\lambda\to\infty} \exp_1(\theta + \lambda u) = \frac{pe^{p\theta p}}{\mathrm{tr}(pe^{p\theta p})}.$$

With the maximal eigenvalue $\mu_+(u)$ of u one has for the free energy

$$\lim_{\lambda\to\infty} \left(F(\theta + \lambda u) - \lambda\mu_+(u)\right) = \ln\left(\mathrm{tr}\left(pe^{p\theta p}\right)\right) = F\left((\kappa^p)^{-1}(p\theta p)\right).$$

[Proof on page 229]

In the following theorem we find a connection between the convex geometric concept of non-exposed face and the differential geometric object of e-geodesic.

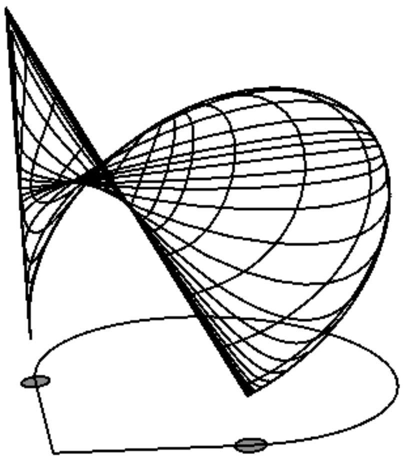


Figure 7.1: The swallow family $\mathcal{E}$ is depicted including the boundary curve of the state reflection sr_U below the family. Two points of sr_U (indicated by circles) are non-exposed faces. Unlike all other points of sr_U they are not the mean value with respect to U of a state in the closure of an e-geodesic in $\mathcal{E}$. The wire frame of $\mathcal{E}$ in the picture consists of e-geodesics.

Theorem 5 A state $\rho \in \overline{S}(A)$ belongs to the closure of an e-geodesic included in $\mathcal{E}$ if and only if ρ belongs to the compression $\kappa^p(\mathcal{E}_p)$ for a non-zero projector $p \in \mathcal{P}_{U,\perp}$.
[Proof on page 229]

Corollary 7.7 A point $x \in \mathrm{sr}_U$ is the mean value with respect to U of a density matrix in the closure of an e-geodesics in $\mathcal{E}$ if and only if x belongs to the relative interior of an exposed face of sr_U. [Proof on page 230]

Example 7.8 We have seen in Example 5.34 that the swallow family depicted in Figure 7.1 has two non-exposed projectors. The corresponding non-exposed extreme points of the state reflection are not covered under projection π_U by the closure of an e-geodesics in $\mathcal{E}$ by Corollary 7.7.

We now start the main construction for the description of the *rI*-closure. This is a piecewise e-geodesic extension of $\mathcal{E}$, which is in bijection to the state reflection under the projection π_U.

Proposition 7.9 For a non-zero projector $p \in \mathcal{P}_U$ the following diagram commutes.

$$\begin{tikzcd}
\mathrm{Dom}_p \arrow[r, shift left, "\kappa^p"] \arrow[d, phantom, "\cup"] \arrow[ddd, bend right=60, "\pi_{\varsigma^p(U)}"'] & \overline{S}(A) \cap \big(\mathrm{ri}(\mathbb{F}_U(p)) + U^{\perp}\big) \arrow[l, shift left] \arrow[d, phantom, "\cup"] \arrow[ddd, bend left=60, "\pi_U"] \\
S(A^p) \arrow[r, shift left, "\kappa^p"] \arrow[d, phantom, "\cup"] \arrow[dd, bend right, "\pi_{\varsigma^p(U)}"'] & \mathrm{ri}(\mathbb{F}(p)) \arrow[l, shift left] \arrow[d, phantom, "\cup"] \arrow[dd, bend left, "\pi_U"] \\
\mathcal{E}_p \arrow[r, shift left, "\kappa^p"] \arrow[d, shift right, "\pi_p"'] & \kappa^p(\mathcal{E}_p) \arrow[l, shift left] \arrow[d, shift left, "\pi_U"] \\
\mathrm{ri}(\mathrm{sr}_{\varsigma^p(U)}) \arrow[r, shift left] \arrow[u, shift right, "M_p"'] & \mathrm{ri}(\mathbb{F}_U(p)) \arrow[l, shift left] \arrow[u, shift left]
\end{tikzcd}$$

[Proof on page 230]

Remark 7.10 By (5.56) one has the disjoint cover $\mathrm{sr}_U = \bigcup_{p \in \mathcal{P}_U \setminus \{0\}} \mathrm{ri}(\mathbb{F}_U(p))$ by face reflections. Since $\mathrm{sr}_U = \pi_U\big(\overline{S}(A)\big)$ there exists for an arbitrary state $\rho \in \overline{S}(A)$ a unique non-zero projector $p \in \mathcal{P}_U$ such that $\rho \in \overline{S}(A) \cap \big(\mathrm{ri}(\mathbb{F}_U(p)) + U^{\perp}\big)$. With Proposition 7.9 this is $\rho \in \kappa^p(\mathrm{Dom}_p)$. One has the disjoint cover

$$\overline{S}(A) = \bigcup_{p \in \mathcal{P}_U \setminus \{0\}} \kappa^p(\mathrm{Dom}_p). \tag{7.13}$$

Proposition 7.11 For $\rho \in \overline{S}(A)$ the projector $p := \bigwedge\{q \in \mathcal{P}_U : q \geq s(\rho)\}$ is the unique projector in $\mathcal{P}_U$ such that $\rho \in \kappa^p(\mathrm{Dom}_p)$. The combinatorial mean value chart $\pi^{\mathrm{cmb}} : \mathcal{E}^{\mathrm{cmb}} \to \mathrm{sr}_U$ is a bijection. The inverse map is $x \mapsto (x + U^{\perp}) \cap \mathcal{E}^{\mathrm{cmb}}$ identifying a one-element set with the element. [Proof on page 230]

Justified by Proposition 7.11 we define a generalization of the normal projection to $\mathcal{E}$.

Definition 7.12 The **combinatorial normal projection** to $\mathcal{E}^{\mathrm{cmb}}$ is

$$N^{\mathrm{cmb}} : \mathrm{sr}_U + U^{\perp} \to \mathcal{E}^{\mathrm{cmb}}, \qquad x \mapsto (x + U^{\perp}) \cap \mathcal{E}^{\mathrm{cmb}} \tag{7.14}$$

identifying a one-element set with the element. As the inverse to the combinatorial mean value chart $\pi^{\mathrm{cmb}} : \mathcal{E}^{\mathrm{cmb}} \to \mathrm{sr}_U$ we define the **combinatorial mean value parametrization** by

$$M^{\mathrm{cmb}} := N^{\mathrm{cmb}}|_{\mathrm{sr}_U}. \tag{7.15}$$

Remark 7.13 **(a)** By Proposition 7.9 and Proposition 7.11 for every density matrix $\rho \in \overline{S}(A)$ one has

$$s(N^{\mathrm{cmb}}(\rho)) = \bigwedge\{q \in \mathcal{P}_U : q \geq s(\rho)\}. \tag{7.16}$$

In particular $s(N^{\mathrm{cmb}}(\rho)) \geq s(\rho)$. The support projector $s(N^{\mathrm{cmb}}(\rho))$ is the unique projector in $\mathcal{P}_U$ such that

$$\rho \in \kappa^{s(N^{\mathrm{cmb}}(\rho))}(\mathrm{Dom}_{s(N^{\mathrm{cmb}}(\rho))}). \tag{7.17}$$

This projector is by construction of the combinatorial extension $\mathcal{E}^{\mathrm{cmb}}$ also the unique projector in $\mathcal{P}_U$ such that

$$N^{\mathrm{cmb}}(\rho) \in \kappa^{s(N^{\mathrm{cmb}}(\rho))}(\mathcal{E}_{s(N^{\mathrm{cmb}}(\rho))}). \tag{7.18}$$

(b) Combinatorial mean value chart (7.5) and mean value parametrization (7.15) are inverses to each other by Proposition 7.11

$$\pi^{\mathrm{cmb}} \circ M^{\mathrm{cmb}} = \mathrm{Id}\,|_{\mathrm{sr}_U} \qquad \text{and} \qquad M^{\mathrm{cmb}} \circ \pi^{\mathrm{cmb}} = \mathrm{Id}\,|_{\mathcal{E}^{\mathrm{cmb}}}. \tag{7.19}$$

For a non-zero projector $p \in \mathcal{P}_U$ one has the restriction $\pi^{\mathrm{cmb}}|_{\kappa^p(\mathcal{E}_p)} = \pi_U|_{\kappa^p(\mathcal{E}_p)}$. Then by (7.10) there is a diffeomorphism $\pi_U|_{\kappa^p(\mathcal{E}_p)} : \kappa^p(\mathcal{E}_p) \to \mathrm{ri}(\mathbb{F}_U(p))$ and therefore

$$M^{\mathrm{cmb}}|_{\mathrm{ri}(\mathbb{F}_U(p))} = \left(\pi_U|_{\kappa^p(\mathcal{E}_p)}\right)^{-1}. \tag{7.20}$$

For $x \in \overline{S}(A) + U^{\perp} = \mathrm{sr}_U + U^{\perp}$ one has $\pi^{\mathrm{cmb}} \circ N^{\mathrm{cmb}}(x) = \pi_U\left((x + U^{\perp}) \cap \mathcal{E}^{\mathrm{cmb}}\right) = \pi_U(x)$. Application of the combinatorial mean value parametrization gives

$$N^{\mathrm{cmb}} = M^{\mathrm{cmb}} \circ \pi_U|_{\overline{S}(A)+U^{\perp}}. \tag{7.21}$$

Then with $\rho \in \kappa^p(\mathrm{Dom}_p)$ and $\rho_p := (\kappa^p)^{-1}(\rho)$ we have by (7.12) and Proposition 7.9

$$\begin{aligned} \kappa^p \circ N_p(\rho_p) &= \kappa^p \circ M_p \circ \pi_{\varsigma^p(U)}(\rho_p) = \left(\pi_U|_{\kappa^p(\mathcal{E}_p)}\right)^{-1} \circ \pi_U \circ \kappa^p(\rho_p) \\ &= M^{\mathrm{cmb}} \circ \pi_U(\rho) = N^{\mathrm{cmb}}(\rho). \end{aligned} \tag{7.22}$$

Lemma 7.14 If $\rho \in \mathrm{Dom}$ then the relative entropy S_ρ has a unique minimum on $\mathcal{E}^{\mathrm{cmb}}$ at $N(\rho) = N^{\mathrm{cmb}}(\rho) \in \mathcal{E}$. [Proof on page 231]

Lemma 7.15 Let $\theta, u \in A^0_{\mathrm{sa}}$, $u \neq 0$ and let $p := p_+(u)$ be the maximal projector of u. Then for $\rho \in \mathbb{F}(p)$ and $\lambda \in \mathbb{R}$ we have $\frac{\mathrm{d}}{\mathrm{d}\,\lambda} S_\rho\left(\exp_1(\theta + \lambda u)\right) < 0$. In the limit for large λ we have $S_\rho\left(\lim_{\lambda\to\infty} \exp_1(\theta + \lambda u)\right) = \lim_{\lambda\to\infty} S_\rho\left(\exp_1(\theta + \lambda u)\right)$. [Proof on page 231]

Lemma 7.16 Let $p \in \mathcal{P}_U$ and $q \in \kappa^p\left(\mathcal{P}_{\varsigma^p(U),\perp}\right)$ be non-zero projectors with $q \lneq p$. Then for each point $\widetilde{\sigma} \in \mathcal{E}_p$ there exists an e-geodesic $g_{\widetilde{\sigma}} : \mathbb{R} \to \mathcal{E}_p$ passing through $\widetilde{\sigma}$ such that

$$\kappa^q(\mathcal{E}_q) = \kappa^p\left(\left\{\lim_{\lambda\to\infty} g_{\widetilde{\sigma}}(\lambda) : \widetilde{\sigma} \in \mathcal{E}_p\right\}\right).$$

For a state $\rho \in \mathbb{F}(q)$ the relative entropy $S_\rho\left(\kappa^p\left(g_{\widetilde{\sigma}}(\lambda)\right)\right)$ is strictly monotone decreasing in $\lambda \in \mathbb{R}$ and $\inf\limits_{\tau \in g_{\widetilde{\sigma}}} S_\rho(\kappa^p(\tau)) = S_\rho\left(\kappa^p\left(\lim\limits_{\lambda\to\infty} g_{\widetilde{\sigma}}(\lambda)\right)\right)$. [Proof on page 232]

Now we can prove that the linear geometry of entropy distance given by the normal projection survives in the extension to the full state space.

Theorem 6 For arbitrary $\rho \in \overline{S}(A)$ we have $\inf_{\sigma\in\mathcal{E}} S_\rho(\sigma) = \min_{\sigma\in\mathcal{E}^{\mathrm{cmb}}} S_\rho(\sigma)$. The minimum on $\mathcal{E}^{\mathrm{cmb}}$ is unique at $N^{\mathrm{cmb}}(\rho)$. [Proof on page 233]

Corollary 7.17 One has $\mathcal{E}^{\mathrm{cmb}} = \mathrm{cl}_{rI}(\mathcal{E})$. [Proof on page 234]

We consider the two examples of *rI*-closures depicted in Figure 7.2. It is instructive to find the components of the *rI*-closures on the frustum barrel of Figure 6.2 and Figure 6.3 on page 132.

Example 7.18 (a) The swallow family $\mathcal{E} := \exp_1\left(\mathrm{Lin}\{\sigma_1 \oplus 1, \sigma_2 \oplus 1\}\right)$ has the projector lattice calculated in Example 5.34

$$\begin{aligned} \mathcal{P}_U &= \left\{0,\ 0_2 \oplus 1,\ p_+(\sigma_1) \oplus 1,\ p_+(\sigma_2) \oplus 1,\ \mathbb{1}\right\} \\ &\cup \left\{p_+(\sigma_1 \cos(\alpha) + \sigma_2 \sin(\alpha)) \oplus 0 :\ \alpha \in (\tfrac{\pi}{2}, 2\pi)\right\}. \end{aligned}$$

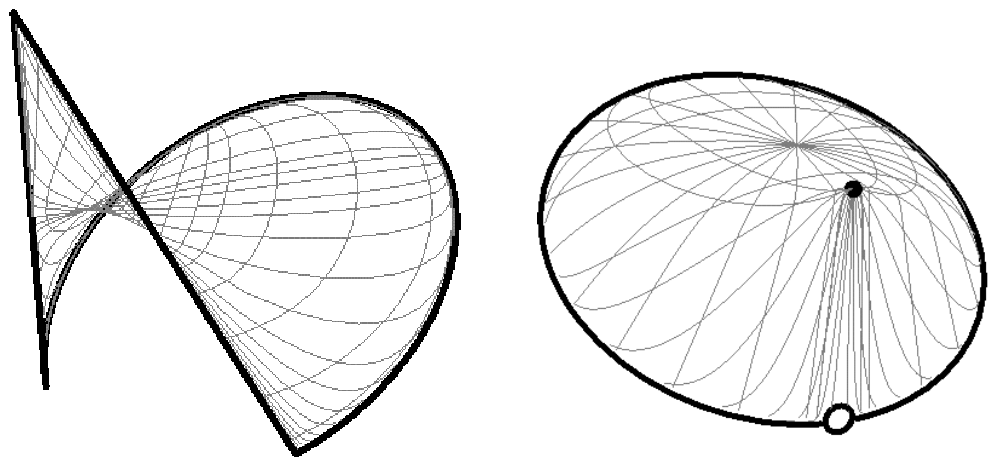


Figure 7.2: The *rI*-closures of the swallow family (left) and the Staffelberg family (right) are depicted. Points on an *rI*-closure but outside a family are drawn in thick black. The absence of a point is indicated by a small circle.

The canonical parameter space is $\Theta = \mathrm{Lin}_{\mathbb{R}}\{(\sigma_1 \oplus 1) - \frac{\mathbb{1}}{3},\ (\sigma_2 \oplus 1) - \frac{\mathbb{1}}{3}\}$ and this is also the canonical tangent space U of $\mathcal{E}$. For a projector $p \in \mathcal{P}_U$ of rank one $A^p = \mathbb{C}$ holds and then the compression of $\mathcal{E}$ by p is the single pure state $\mathcal{E}_p = \overline{S}(A^p) = \{1\}$. This gives for rank one projectors $p \in \mathcal{P}_U$

$$\kappa^p(\mathcal{E}_p) = \{p\}.$$

There are two projectors $p := p_+(\sigma_1) \oplus 1$ and $q := p_+(\sigma_2) \oplus 1$ of rank two in $\mathcal{P}_U$. Example 5.34 provides the solutions $\pi_{\mathrm{lin}(\mathbb{F}(p))}(\Theta) = \mathbb{R}[p_+(\sigma_1) \oplus (-1)]$ and $\pi_{\mathrm{lin}(\mathbb{F}(q))}(\Theta) = \mathbb{R}[p_+(\sigma_2) \oplus (-1)]$. Since $A^p = A^q = \mathbb{C}^2$, one has $\varsigma^p(\Theta) = \varsigma^q(\Theta) = (A^p)^0_{\mathrm{sa}}$ and the exponential family $\mathcal{E}_p = \mathcal{E}_q = S(A^p)$ is a relative open segment. One has

$$\kappa^p(\mathcal{E}_p) =]p_+(\sigma_1) \oplus 0,\ 0_2 \oplus 1[\qquad \text{and} \qquad \kappa^q(\mathcal{E}_q) =]p_+(\sigma_2) \oplus 0,\ 0_2 \oplus 1[\,.$$

Hence the *rI*-closure $\mathrm{cl}_{rI}(\mathcal{E})$ is the union of $\mathcal{E}$, the pure state $0_2 \oplus 1$, the closed arc

$$\left\{p_+(\sigma_1 \cos(\alpha) + \sigma_2 \sin(\alpha)) \oplus 0 :\ \alpha \in [\frac{\pi}{2}, 2\pi]\right\}$$

and the two relative open segments $\kappa^p(\mathcal{E}_p)$ and $\kappa^q(\mathcal{E}_q)$ linking the endpoints of the arc to $0_2 \oplus 1$. We see that $\mathrm{cl}_{rI}(\mathcal{E}) \setminus \mathcal{E}$ is homeomorphic to a circle and it should be possible to prove that $\mathrm{cl}_{rI}(\mathcal{E})$ is closed in norm topology.

(b) The Staffelberg family $\mathcal{E} := \exp_1(\mathrm{Lin}\{\sigma_1 \oplus 0, \sigma_2 \oplus 1\})$ has the projector lattice calculated in Example 5.34

$$\mathcal{P}_{V,\perp} \ = \ \{0,\ p,\ \mathbb{1}\} \ \cup \ \left\{p_+(\sigma_2 \cos(\alpha - \frac{\pi}{4}) - \sigma_1 \sin(\alpha - \frac{\pi}{4})) \oplus 0 \ :\ \alpha \in [0, 2\pi) \setminus \{\frac{\pi}{4}\}\right\}.$$

with $p := p_+(\sigma_2) \oplus 1$. The canonical parameter space and tangent space of $\mathcal{E}$ is

$$\Theta = U = \mathrm{Lin}_{\mathbb{R}}\{\sigma_1 \oplus 0,\ (\sigma_2 \oplus 1) - \frac{\mathbb{1}}{3}\}.$$

The only projector of rank two in $\mathcal{P}_U$ is p and Example 5.34 provides the solution $\varsigma^p(\Theta) = \{0\}$ for the canonical parameter space of $\mathcal{E}_p$. This proves $\kappa^p(\mathcal{E}_p) = \left\{\frac{1}{2}p\right\}$. The rI-closure is the union

$$\mathrm{cl}_{rI}(\mathcal{E}) \;=\; \mathcal{E} \;\cup\; \left\{\frac{1}{2}p\right\} \;\cup\; \{p_+(\sigma_2\cos(\beta) + \sigma_1\sin(\beta)) \oplus 0 :\ \beta \in (0, 2\pi)\}\}\,. \tag{7.23}$$

Since $p_+(\sigma_2) \oplus 0 - \frac{1}{2}p \perp U$ the equation (7.21) $N^{\mathrm{cmb}} = M^{\mathrm{cmb}} \circ \pi_U|_{\overline{S}(A)+U^\perp}$ proves that

$$N^{\mathrm{cmb}}(p_+(\sigma_2) \oplus 0) = \frac{1}{2}p. \tag{7.24}$$

For metric properties of the reverse information closure consider with $\beta \in \mathbb{R}$ the distance

$$\|\frac{1}{2}p - p_+(\sigma_2\cos(\beta) + \sigma_1\sin(\beta)) \oplus 0\|_2 = \sqrt{1 - \frac{1}{2}\cos(\beta)}$$

which is minimized for $\beta = 0$ with $\|\frac{1}{2}p - p_+(\sigma_2) \oplus 0\|_2 = \frac{1}{\sqrt{2}}$. We see that $\mathrm{cl}_{rI}(\mathcal{E}) \setminus \mathcal{E}$ is not connected. Since $p_+(\sigma_2) \oplus 0$ is not in $\mathrm{cl}_{rI}(\mathcal{E})$ we see that $\mathrm{cl}_{rI}(\mathcal{E})$ is not closed in norm topology. We notice a discontinuity of the combinatorial normal projection N^{cmb} at $p_+(\sigma_2 \oplus 0)$ and thus at each point in $p_+(\sigma_2 \oplus 0) + U^\perp$. In particular, the combinatorial mean value parametrization M^{cmb} has a discontinuity at $\pi_U(p_+(\sigma_2 \oplus 0))$. Topological properties of the mean value parametrization are analyzed more in depth in Section 7.2.

Remark 7.19 (Statistics and convex support) The parts $\kappa^p(\mathcal{E}_p)$ of the combinatorial extension $\mathcal{E}^{\mathrm{cmb}}$ can be defined independently (7.8) of a choice of κ^p. Another independent definition can be the isomorphism to the rI-closure in Corollary 7.17. We show that the combinatorial mean value parametrization (via the convex support) has an independent description, too. We put as in Remark 6.32

$$\mathcal{E} := \exp_1\left(a_0 + \mathrm{Lin}\{a_1, \ldots, a_k\}\right)$$

for a finite number of self-adjoint matrices $a_0, \ldots, a_k \in A_{\mathrm{sa}}$. At the beginning we settle the projector lattice $\mathcal{P}_U$ for the canonical tangent space $U = \pi_{A^0_{\mathrm{sa}}}(\mathrm{Lin}\{a_1, \ldots, a_k\})$ of $\mathcal{E}$. This lattice can be calculated directly from the statistic $a_1, \ldots, a_k$ using Remark 5.33.

The projector lattice $\mathcal{P}_U$ being established, we obtain by Proposition 7.11 for an arbitrary density matrix $\rho \in \overline{S}(A)$ the unique projector

$$p := \bigwedge\{q \in \mathcal{P}_U : q \geq s(\rho)\}$$

in $\mathcal{P}_U$ such that $\rho \in \kappa^p(\mathrm{Dom}_p)$. The mean value mapping for $\mathcal{E}$ (6.30) $m : A_{\mathrm{sa}} \to \mathbb{R}^k$, $a \mapsto (\langle a_1, a\rangle, \dots, \langle a_k, a\rangle)$ induces a bijection from the combinatorial extension of $\mathcal{E}$

$$m|_{\mathcal{E}^{\mathrm{cmb}}} : \mathcal{E}^{\mathrm{cmb}} \to \overline{m(\mathcal{E})}$$

to the convex support $\overline{m(\mathcal{E})}$. This follows from the bijection $\pi_U|_{\mathcal{E}^{\mathrm{cmb}}} : \mathcal{E}^{\mathrm{cmb}} \to \mathrm{sr}_U$ (7.19) and the extension of the diagram (6.31). Let $(\xi_1, \dots, \xi_k) \in \overline{m(\mathcal{E})}$ and let $\rho \in \overline{S}(A)$. We study conditions for

$$\rho \in \mathcal{E}^{\mathrm{cmb}} \qquad \text{and} \qquad m(\rho) = (\xi_1, \dots, \xi_k). \tag{7.25}$$

If $p = \mathbb{1}$ then ρ satisfies (7.25) if and only if there are coefficients $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ such that $\rho = \exp_1(a_0 + \sum_{i=1}^k \lambda_i a_i)$ and for $j = 1, \dots, k$ we have (6.33)

$$\frac{\partial}{\partial \lambda_j} F(a_0 + \sum_{i=1}^k \lambda_i a_i) = \xi_j.$$

Otherwise if $p \neq \mathbb{1}$ then there is no solution to (6.33) because $\rho \notin \mathrm{Dom}$ in this case. Let $p \in \mathcal{P}_U$ be non-zero projector. We use the statistic $\varsigma^p(a_1), \dots, \varsigma^p(a_k)$ for the compression

$$\mathcal{E}_p = \exp_1\left(\varsigma^p(a_0) + \mathrm{Lin}\{\varsigma^p(a_1), \dots, \varsigma^p(a_k)\}\right)$$

and we put $\rho_p := (\kappa^p)^{-1}(\rho)$. Then we can use (7.4) and Proposition 7.11 and we get

$$\rho \in \mathcal{E}^{\mathrm{cmb}} \qquad \Longleftrightarrow \qquad \rho \in \kappa^p(\mathcal{E}_p) \qquad \Longleftrightarrow \qquad \rho_p \in \mathcal{E}_p.$$

The mean value mapping for $\mathcal{E}_p$

$$m_p : (A^p)_{\mathrm{sa}} \to \mathbb{R}^k, \qquad a \mapsto (\langle \varsigma^p(a_1), a\rangle, \dots, \langle \varsigma^p(a_k), a\rangle)$$

induces a diffeomorphism $m_p|_{\mathcal{E}_p} : \mathcal{E}_p \to m_p(\mathcal{E}_p) \subset \mathbb{R}^k$ in analogy to the mean value mapping (6.30) for $\mathcal{E}$. We have for $j = 1, \dots, k$

$$\langle \varsigma^p(a_j), \rho_p\rangle = \langle a_j, \rho\rangle - \frac{\mathrm{tr}(p a_j)}{\mathrm{tr}(p)}$$

and for a tuple $\lambda_1, \dots, \lambda_k \in \mathbb{R}^k$ we have

$$\begin{aligned} & \tfrac{\partial}{\partial \lambda_j} F\left(\varsigma^p(a_0) + \textstyle\sum_{i=1}^k \lambda^i \varsigma^p(a_i)\right) \\ = \quad & -\tfrac{\mathrm{tr}(p a_j)}{\mathrm{tr}(p)} + \tfrac{\partial}{\partial \lambda_j} \ln\left(\mathrm{tr}(\mathbb{1} - p) + \mathrm{tr}(\exp(p a_0 p + \textstyle\sum_{i=1}^k \lambda_i p a_i p))\right). \end{aligned}$$

Hence we obtain for $\lambda_1, \dots, \lambda_k \in \mathbb{R}^k$ with (7.7) that $\rho \in \mathcal{E}^{\mathrm{cmb}}$ and $m(\rho) = (\xi_1, \dots, \xi_k)$ if and only if $\rho = \frac{p e^{p a_0 p + \sum_{i=1}^k \lambda_i p a_i p}}{\mathrm{tr}\left(p e^{p a_0 p + \sum_{i=1}^k \lambda_i p a_i p}\right)}$ and for $j = 1, \dots, k$ we have

$$\frac{\partial}{\partial \lambda_j} \ln\left(\mathrm{tr}(\mathbb{1} - p) + \mathrm{tr}(\exp(p a_0 p + \sum_{i=1}^k \lambda_i p a_i p))\right) = \xi_j. \tag{7.26}$$

A solution to (7.26) is only possible with $p \in \mathcal{P}_U$ if $p = \bigwedge\{q \in \mathcal{P}_U : q \geq s(\rho)\}$. For the special case $p = \mathbb{1}$ the equation simplifies to the derivative of free energy F as above.

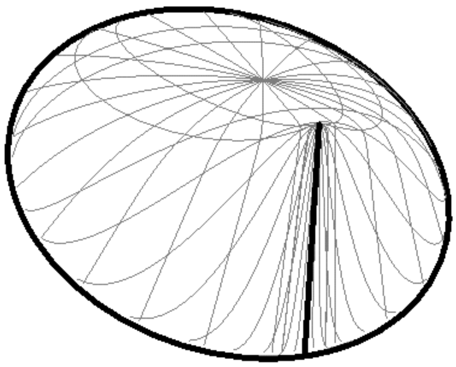


Figure 7.3: The variation closure $\overline{\mathcal{E}}$ of the Staffelberg family is depicted. Points on the variation closure but outside the family are drawn in thick black. The circle about the family is the closure of the pure states in $\mathcal{P}_U$. The vertical segment is proved in Example 7.21 to belong to $\overline{\mathcal{E}}$. Only its top vertex belongs to the rI-closure of $\mathcal{E}$.

7.2 Topology of exponential families

The closure in norm topology of an exponential family $\mathcal{E}$ can be strictly larger than the rI-closure $\mathrm{cl}_{rI}(\mathcal{E})$. A discrepancy $\mathrm{cl}_{rI}(\mathcal{E}) \subsetneq \overline{\mathcal{E}}$ exists for the Staffelberg family in Example 7.21. This is a feature, which does not appear for the classical (finite-dimensional) case and more generally under a regularity discussed with examples in Remark 7.27.

It turns out that the topology of $\mathcal{E}$ is related to the topology of the projector lattice $\mathcal{P}_U$ for U the canonical tangent space of $\mathcal{E}$. Question 5 from Section 5.4 about the inclusion $\mathcal{P}_U \subset \overline{\mathcal{P}_{U,\perp}}$ has a meaning in the context of exponential families. This is explained in Remark 7.25.

For completeness [Cs05] we also mention the I-closure in analogy with the classical case. By custom in probability theory we use the trace norm $\|\cdot\|_1$. There is no difference to the two-norm because all norms on a finite-dimensional vector space are equivalent.

Definition 7.20 The ***I*-closure** of $\mathcal{E}$ is $\mathrm{cl}_I(\mathcal{E}) := \{\rho \in \overline{S}(A) : \inf_{\sigma\in\mathcal{E}} S(\sigma,\rho) = 0\}$. The **variation closure** or **closure** of $\mathcal{E}$ is the norm closure $\overline{\mathcal{E}} = \{\rho \in \overline{S}(A) : \inf_{\sigma\in\mathcal{E}} \|\rho-\sigma\|_1 = 0\}$.

Example 7.21 We consider the Staffelberg family $\mathcal{E} = \exp_1(\mathrm{Lin}\{\sigma_1 \oplus 0,\ \sigma_2 \oplus 1\})$ with the aim to detect a difference between the *rI*-closure and the variation closure of $\mathcal{E}$. The *rI*-closure is the union (7.23)

$$\mathrm{cl}_{rI}(\mathcal{E}) \;=\; \mathcal{E} \;\cup\; \left\{\frac{p}{2}\right\} \;\cup\; \{p_+(\sigma_2\cos(\beta) + \sigma_1\sin(\beta)) \oplus 0 :\ \beta \in (0, 2\pi)\}\}$$

with $p := p_+(\sigma_2) \oplus 1$. Using $q := p_+(\sigma_2) \oplus 0$, the segment $g := [q, \frac{p}{2}]$ projects under π_U to a single point. We prove on page 235 that g belongs to the variation closure $\overline{\mathcal{E}}$. Then

$$\overline{\mathcal{E}} \setminus \mathrm{cl}_{rI}(\mathcal{E}) \supset g \setminus \{\frac{p}{2}\}.$$

In addition, the half-open segment $g \setminus \{q\}$ consists of density matrices with support p. The matrices in $\overline{\mathcal{E}}$ with support p are not confined to $\kappa^p(\mathcal{E}_p) = \{\frac{p}{2}\}$.

The following lemma was created to prove the equality $\overline{\mathcal{E}} = \mathrm{cl}_{rI}(\mathcal{E})$ for classical exponential families in Section 8.2. In the final proof, the result is not necessary.

Lemma 7.22 *If there are two density matrices $\rho \neq \sigma$ in $\overline{\mathcal{E}}$ such that $\pi_U(\rho) = \pi_U(\sigma)$ then there is an uncountable set of density matrices $\{\rho_\alpha\}_{\alpha \in I}$ in $\overline{\mathcal{E}}$ such that $\pi_U(\rho_\alpha) = \pi_U(\rho)$ for $\alpha \in I$.* [Proof on page 236]

We compare the distinct closures of a general exponential family in a matrix algebra.

Remark 7.23 **(a)** The quantum mechanical generalization of the Pinsker-Csiszár inequality [Pe08] allows to compare relative entropy distance to trace norm distance, for density matrices $\rho, \sigma \in \overline{S}(A)$

$$S(\rho, \sigma) \geq \frac{1}{2}\|\rho - \sigma\|_1^2$$

holds. This implies the inclusions

$$\mathrm{cl}_I(\mathcal{E}) \subset \overline{\mathcal{E}} \qquad \text{and} \qquad \mathrm{cl}_{rI}(\mathcal{E}) \subset \overline{\mathcal{E}}. \tag{7.27}$$

(b) The *I*-closure is easy to describe, it is $\mathrm{cl}_I(\mathcal{E}) = \mathcal{E}$. A density matrix $\rho \in \overline{\mathcal{E}} \setminus \mathcal{E}$ is singular by Corollary 6.26. Hence $S(\sigma, \rho) = \infty$ for all $\sigma \in \mathcal{E}$ and this gives $\rho \notin \mathrm{cl}_I(\mathcal{E})$. With (7.27) we find $\mathrm{cl}_I(\mathcal{E}) \subset \mathcal{E}$.

(c) The *rI*-closure was the subject of study in Section 7.1. It is defined by (7.2)

$$\mathrm{cl}_{rI}(\mathcal{E}) = \{\rho \in \overline{S}(A) : \inf_{\sigma\in\mathcal{E}} S(\rho,\sigma) = 0\}.$$

Corollary 7.17 proves the equality $\mathrm{cl}_{rI}(\mathcal{E}) = \mathcal{E}^{\mathrm{cmb}}$ to the combinatorial extension (7.4)

$$\mathcal{E}^{\mathrm{cmb}} = \bigcup_{p\in\mathcal{P}_U\setminus\{0\}} \kappa^p\left(\mathcal{E}_p\right).$$

When studying the variation closure $\overline{\mathcal{E}}$ it is fruitful to study the pair $\mathrm{cl}_{rI}(\mathcal{E}) \subset \overline{\mathcal{E}}$ (7.27). By Theorem 4 on the mean value chart there is the continuous surjection

$$\pi_U|_{\overline{\mathcal{E}}} : \overline{\mathcal{E}} \to \mathrm{sr}_U$$

to the state reflection. The combinatorial mean value chart is $\pi^{\mathrm{cmb}} = \pi_U|_{\mathrm{cl}_{rI}(\mathcal{E})}$ (7.19). This is a continuous bijection $\pi^{\mathrm{cmb}} : \mathrm{cl}_{rI}(\mathcal{E}) \to \mathrm{sr}_U$. The inverse is the combinatorial mean value parametrization, which is a bijection defined on the compact state reflection (7.15)

$$M^{\mathrm{cmb}} : \mathrm{sr}_U \to \mathrm{cl}_{rI}(\mathcal{E}).$$

One has

$$\begin{aligned} \mathrm{cl}_{rI}(\mathcal{E}) = \overline{\mathcal{E}} \quad &\Longleftrightarrow \quad \pi_U|_{\overline{\mathcal{E}}} \text{ is injective} \\ &\Longleftrightarrow \quad M^{\mathrm{cmb}} : \mathrm{sr}_U \to \mathrm{cl}_{rI}(\mathcal{E}) \text{ is continuous} \\ &\Longleftrightarrow \quad \mathrm{cl}_{rI}(\mathcal{E}) \text{ is closed.} \end{aligned} \tag{7.28}$$

In case $\mathrm{cl}_{rI}(\mathcal{E}) \subsetneq \overline{\mathcal{E}}$ we can suspect the exact position of a discontinuity in (7.28) by looking at Example 7.18 (b). The Staffelberg family $\mathcal{E}$ has a discontinuity of the combinatorial mean value parametrization located at $\pi_U(p_+(\sigma_2 \oplus 0))$. On the other hand, every symmetrization map for the state space $\overline{S}(A)$ at U has a discontinuity at $p_+(\sigma_2 \oplus 0)$ by Remark 5.54 (c).

Question 8 The combinatorial mean value parametrization $M^{\mathrm{cmb}} : \mathrm{sr}_U \to \mathcal{E}^{\mathrm{cmb}}$ has a discontinuity at $x \in \mathrm{sr}_U$ if and only if every symmetrization map of the state space $\overline{S}(A)$ at U has a discontinuity on $x + U^{\perp}$.

Let us have a closer look at the difference $\overline{\mathcal{E}} \setminus \mathrm{cl}_{rI}(\mathcal{E})$.

Lemma 7.24 If (p_i) is a sequence of non-zero orthogonal projectors with limit $p \in \mathcal{P}$ then $\kappa^p(\mathcal{E}_p) \subset \overline{\bigcup_{i\in\mathbb{N}} \kappa^{p_i}(\mathcal{E}_{p_i})}$. If $p \in \overline{\mathcal{P}_U} \setminus \{0\}$ then $\kappa^p(\mathcal{E}_p) \subset \overline{\mathcal{E}}$. [Proof on page 236]

Remark 7.25 (Geodesic and Grassmannian approximation) We saw in the previous section in Theorem 6 that a density matrix ρ in the rI-closure of an exponential family $\mathcal{E}$ is approximated by a curve in $\mathrm{cl}_{rI}(\mathcal{E})$ that is piecewise an e-geodesic running in compressions $\kappa^p(\mathcal{E}_p)$ with p in an access sequence for sr_U. The projectors in an access sequence have a strictly decreasing rank. Thus they are separated by spectral norm one (Lemma 2.38).

In contrast the approximation in Lemma 7.24 takes place in the projector lattice $\mathcal{P}$ (which is isomorphic to the Grassmannian). By Theorem 5 a density matrix ρ in the compression $\kappa^p(\mathcal{E}_p)$ for a non-zero projector $p \in \mathcal{P}_{U,\perp}$ belongs to the closure of an e-geodesic in $\mathcal{E}$. If $p \neq \mathbb{1}$ then ρ is a relative boundary point of the state space and is the limit point of an e-geodesic. Thus the affirmative to Question 5 from Section 5.4,

$$\mathcal{P}_U \subset \overline{\mathcal{P}_{U,\perp}},$$

implies that $\mathrm{cl}_{rI}(\mathcal{E}) \setminus \mathcal{E}$ is approximated by limit points of e-geodesics in $\mathcal{E}$.

Let us consolidate more topological links between the projector lattice $\mathcal{P}_U$ and the exponential family $\mathcal{E}$.

Proposition 7.26 If $\overline{\mathcal{E}} = \mathrm{cl}_{rI}(\mathcal{E})$ then $\mathcal{P}_U$ is closed. [Proof on page 237]

Remark 7.27 (Geodesic closure) **(a)** A special instance of the equality $\mathrm{cl}_{rI}(\mathcal{E}) = \overline{\mathcal{E}}$ is

$$\overline{\mathcal{E}} = \bigcup_{p\in\mathcal{P}_{U,\perp}\setminus\{0\}} \kappa^p(\mathcal{E}_p). \tag{7.29}$$

By Theorem 5 this equality is equivalent to the variation closure $\overline{\mathcal{E}}$ being the union of closures of e-geodesics in $\mathcal{E}$. As discussed in Remark 7.23 (c) the union on the right-hand side is contained in the rI-closure $\mathrm{cl}_{rI}(\mathcal{E})$ and $\mathrm{cl}_{rI}(\mathcal{E}) \subset \overline{\mathcal{E}}$. This gives $\mathrm{cl}_{rI}(\mathcal{E}) = \overline{\mathcal{E}}$ and

$$\mathcal{P}_{U,\perp} = \mathcal{P}_U. \tag{7.30}$$

We deduce that the state reflection sr_U has only exposed faces. In addition from Proposition 7.26 it follows that the lattice $\mathcal{P}_U$ is closed.

(b) Some exponential families important for applications have the regularity (7.29), among them the exponential family $S(A) = \exp_1(A^0_{\mathrm{sa}})$ of invertible density matrices. If $\rho \in \overline{S}(A)$ is an arbitrary density matrix then we can choose a vector $u \in A^0_{\mathrm{sa}}$ with maximal projector $p_+(u) = s(\rho)$. For example we can use $u = s(\rho)\,\mathrm{tr}(\mathbb{1} - s(\rho)) - (\mathbb{1} - s(\rho))\,\mathrm{tr}(s(\rho))$. The function

$$\widehat{\ln} \;:\; [0,\infty) \to \mathbb{R},\;\; x \mapsto \begin{cases} 0 & \text{if } x = 0, \\ \ln(x) & \text{otherwise} \end{cases}$$

is a continuous function when restricted to a finite subset $F \subset [0,\infty)$. We can define by functional calculus

$$\theta := \widehat{\ln}(\rho) - \frac{\mathrm{tr}(\widehat{\ln}(\rho))}{\mathrm{tr}(\mathbb{1})}\mathbb{1} \in A^0_{\mathrm{sa}}$$

and obtain $\lim_{\lambda\to\infty} \exp_1(\theta + \lambda u) = \frac{pe^{p\theta p}}{\mathrm{tr}(pe^{p\theta p})} = \rho$ by Lemma 7.6. We will generalize this example to factorizable families, where the regularity (7.29) is also true. Other families with the regularity are m-convex families and classical families, which are discussed in Chapter 8.

We can ask whether the converse to Proposition 7.26 holds.

Question 9 Does closedness of $\mathcal{P}_U$ imply $\overline{\mathcal{E}} = \mathrm{cl}_{rI}(\mathcal{E})$?

To approach Question 9 we guess how the variation closure is formed consulting the Staffelberg family $\mathcal{E}$ in Example 7.21. Let $p := p_+(\sigma_2) \oplus 1$ and $q := p_+(\sigma_2) \oplus 0$. Then the segment

$$g := \left[q,\; \frac{1}{2}p\right]$$

is included in $\overline{\mathcal{E}}$. The vertex q of g belongs to the face $\kappa^p(\mathrm{Dom}_p) = \kappa^p(\overline{S}(A^p)) = \mathbb{F}(p)$. The face $\mathbb{F}(p)$ belongs to the lifted face lattice $\mathcal{L}_U$ (Remark 5.13), it is the segment

$$\kappa^p(\mathrm{Dom}_p) = [q,\; 0_2 \oplus 1] = g \cup \left[\frac{1}{2}p,\; 0_2 \oplus 1\right].$$

The segment g is depicted in Figure 7.3 whereas the lifted face $\mathbb{F}(p)$ is a segment in the barrel of the frustum $\overline{S}(M_2 \oplus \mathbb{C}) \cap \mathrm{aff}(\mathcal{E})$ about $\mathcal{E}$ depicted in Figure 6.3 on page 132. The main observation is that $g \subset \overline{\mathcal{E}}$ is the convex hull of two compressed exponential families $\kappa^p(\mathcal{E}_p) = \left\{\frac{1}{2}p\right\}$ and $\kappa^q(\mathcal{E}_q) = \{q\}$ with the projectors p and q included in $\overline{\mathcal{P}_U}$:

Question 10 For $p \in \mathcal{P}_U$ we set $C(p) := \big\{q \in \overline{\mathcal{P}_U} : p = \bigwedge\{r \in \mathcal{P}_U : r \geq q\}\big\}$. Is the variation closure of $\overline{\mathcal{E}}$ equal to the union of sets of convex combinations

$$\left\{\sum_{i=1}^{k} \lambda_i \rho_i \ : \ \lambda_i \geq 0, \ \sum_{i=1}^{k} \lambda_i = 1, \ \rho_i \in \kappa^{p_i}(\mathcal{E}_{p_i}), \ |\{p_1, \dots, p_k\}| = k, \ p_i \in C(p), \ k \in \mathbb{N}\right\}$$

for non-zero projectors $p \in \mathcal{P}_U$?

The conjectured shape of the variation closure in Question 10 implies the affirmative to Question 9. This follows because closedness of $\mathcal{P}_U$ implies $C(p) = \{p\}$ for $p \in \mathcal{P}_U$.

7.3 Discontinuity and local maximizers

We can prove some additional results about entropy distance.

Theorem 7 The entropy distance from $\mathcal{E}$ is continuous if and only if $\overline{\mathcal{E}} = \mathrm{cl}_{rI}(\mathcal{E})$.
[Proof on page 237]

Example 7.28 (Discontinuous entropy distance) The Staffelberg family $\mathcal{E}$ discussed in Example 7.18 (b) has a discontinuous entropy distance. The punctured circle of rank one projectors $\{p_+(\sigma_2 \cos(\beta) + \sigma_1 \sin(\beta)) \oplus 0 : \ \beta \in (0, 2\pi)\}$ belongs to the rI-closure $\mathrm{cl}_{rI}(\mathcal{E})$ (7.23), hence each of these pure states has zero entropy distance from $\mathcal{E}$. The missing pure state $p_+(\sigma_2) \oplus 0$ has the combinatorial normal projection $N^{\mathrm{cmb}}(p_+(\sigma_2) \oplus 0) = \frac{1}{2}(p_+(\sigma_2) \oplus 1)$ (7.24). With Theorem 6

$$S_{\mathcal{E}}(p_+(\sigma_2) \oplus 0) = S(p_+(\sigma_2) \oplus 0, \ \frac{1}{2}(p_+(\sigma_2) \oplus 1)) = \ln(2)$$

holds. The Staffelberg family is depicted in Figure 7.2.

In the previous example it is instructive to compare the location of the discontinuity to the mean value parametrization. This leads us to extend Question 8.

Question 11 If $\mathcal{E}$ is an exponential family in a matrix algebra A, are the following assertions equivalent? For $\rho \in \overline{S}(A)$

(a) the entropy distance has a discontinuity at ρ,
(b) the combinatorial mean value parametrization has a discontinuity at $\pi_U(\rho)$,
(c) every symmetrization map of $\overline{S}(A)$ at U has a discontinuity on $\rho + U^{\perp}$.

The main idea to the following estimate is to confine the entropy distance to a subset where the function is strictly convex. This basic idea was used in [Ay02]. We can generalize it in two directions. We use the state space of a matrix algebra instead of a probability simplex and we consider local maximizers on the entire state space $\overline{S}(A)$ instead of a restriction to the domain $\mathrm{Dom}(\mathcal{E})$.

Proposition 7.29 If $\rho \in \overline{S}(A)$ is a local maximizer of entropy distance from $\mathcal{E}$ then $\dim(F(\rho)) \leq \dim(\mathcal{E})$. [Proof on page 238]

Remark 7.30 The rank of a density matrix $\rho \in \overline{S}(A)$ is frequently the more accessible quantity compared to the face dimension. If the matrix algebra is $A = M_{n_1} \oplus \cdots \oplus M_{n_N}$ for $N \in \mathbb{N}$ and a multi-index $n \in \mathbb{N}_0^N$ then the dimension of the face of a state $\rho = \rho_1 \oplus \cdots \oplus \rho_N \in \overline{S}(A)$ is $\dim(F(\rho)) = \sum_{i=1}^N \mathrm{rk}(\rho_i)^2 - 1$ (4.46). We retrieve the classical bound [Ay02] for a local maximizer $\rho \in \overline{S}(A)$ of entropy distance

$$\mathrm{rk}(\rho) = \sum_{i=1}^{N} \mathrm{rk}(\rho_i) \leq \sum_{i=1}^{N} \mathrm{rk}(\rho_i)^2 = \dim(F(\rho)) + 1 \leq \dim(\mathcal{E}) + 1. \tag{7.31}$$

We abandon the direct sum case and consider a full matrix algebra $A = M_n$ for $n \in \mathbb{N}$. The face of a state $\rho \in \overline{S}(A)$ has dimension $\dim(F(\rho)) = \mathrm{rk}(\rho)^2 - 1$. The rank bound for a local maximizer $\rho \in \overline{S}(A)$ of entropy distance is in the full matrix algebra case

$$\mathrm{rk}(\rho) \leq \sqrt{\dim(\mathcal{E}) + 1}. \tag{7.32}$$

Lemma 7.31 Let F be a non-empty face of the state space $\overline{S}(A)$ with $\dim(F) = \dim(\overline{S}(A)) - 1$. Let $\mathcal{E}$ be an exponential family in A with canonical tangent space $\mathrm{lin}(F)$. Then the centroid $\frac{s(F)}{\mathrm{tr}(s(F))}$ is a local maximizer of entropy distance from $\mathcal{E}$.

[Proof on page 238]

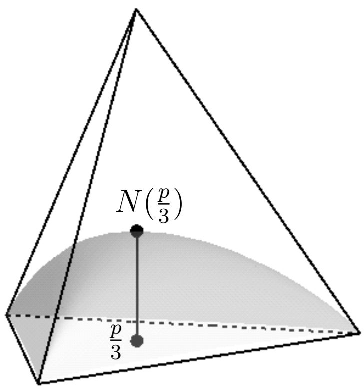


Figure 7.4: For $p := (1, 1, 1, 0)$ the exponential family $\mathcal{E} := \exp_1(\mathrm{lin}(\mathbb{F}(p)) - \ln(\frac{3}{2})p)$ is drawn within the state space of the algebra $\mathbb{C}^4$. The local maximizer $\frac{p}{3}$ of entropy distance from $\mathcal{E}$ has the normal projection $N(\frac{p}{3}) = \frac{1}{9}(2, 2, 2, 3)$ and entropy distance $S_{\mathcal{E}}(\frac{p}{3}) = S(\frac{p}{3}, N(\frac{p}{3})) = \ln(3) - \ln(2)$.

Example 7.32 **(a)** The face dimension bound for a local maximizer of multi-information in Proposition 7.29 is sharp in some cases. The exponential family $S(A)$ of invertible density matrices is the simplest one and Lemma 7.31 presents more examples (one is depicted in Figure 7.4). However, these are almost classical examples because they require a proper face of the state space with codimension one (in the affine hull of the state space).

(b) One might ask whether the bound in Proposition 7.29 can be improved in the quantum case for an exponential family distinct from $S(A)$. A largest proper face F of the state space can be too small to isolate a local maximizer on F from other boundary points outside of F.

For instance, in the algebra M_2 of 2×2-matrices we consider for $\lambda \in \mathbb{R}$ the one-point exponential family $\mathcal{E}$ consisting of the density matrix

$$\exp_1(-\lambda p_+(\sigma_1)) = \exp_1(\lambda p_-(\sigma_1)) = \frac{p_+(\sigma_1) + e^{\lambda} p_-(\sigma_1)}{1 + e^{\lambda}}$$

with Pauli matrices and notation from Example 4.6. We consider the pure states on the equator of the Bloch sphere for $\varphi \in \mathbb{R}$

$$\rho(\varphi) := p_+(\sigma_1 \cos(\varphi) + \sigma_2 \sin(\varphi)).$$

For $\varphi \in \mathbb{R}$ we have $S_{\mathcal{E}}(\rho(\varphi)) = S(\rho(\varphi), \exp_1(\lambda p_-(\sigma_1))) = \ln(1 + e^{\lambda}) - \frac{\lambda}{2}(1 - \cos(\varphi))$. If $\lambda = 0$ then $\mathcal{E} = \{\frac{\mathbb{1}}{2}\}$ is the trace state and every pure state of the Bloch sphere is a maximizer of entropy distance $\ln(2)$. If $\lambda > 0$ then for $\varphi \in \mathbb{R}$ the entropy distance for equator points is bounded by

$$\ln(1 + e^{\lambda}) - \lambda \leq S_{\mathcal{E}}(\rho(\varphi)) \leq \ln(1 + e^{\lambda})$$

and $p_+(\sigma_1)$ at $\varphi = 0$ is the unique global maximizer with $S_{\mathcal{E}}(p_+(\sigma_1)) = \ln(1+e^{\lambda}) > \ln(2)$. If $\lambda < 0$ then $p_+(\sigma_1)$ is not a maximizer of entropy distance. Indeed, one has

$$\ln(1 + e^{\lambda}) \leq S_{\mathcal{E}}(\rho(\varphi)) \leq \ln(1 + e^{\lambda}) - \lambda,$$

the pure state $p_+(\sigma_1)$ is the unique global minimizer on the Bloch sphere and the pure state $p_-(\sigma_1)$ is the unique global maximizer on the Bloch ball.

8 Some examples

The choice of examples is organized historically and by our interest in multi-information for quantum systems. Except for the case of Gibbs ensembles, these examples are very regularly structured having the simple closure discussed in Remark 7.27. In particular, the state reflection (mean value set) has only exposed faces and entropy distance is continuous.

A Gibbs ensemble is the mathematical object that describes a very successful idea in Statistical Physics dating back to Boltzmann in 1866. Information about the state of a realistic physical system, e.g. $2.7 \cdot 10^{25}$ molecules of an ideal gas in one cubic meter, is often available only through the macroscopic quantity of its energy E. One would agree to model the system by a probability distribution with mean energy E. The choice of a particular probability distribution with mean energy E is not unique and the idea is to choose the distribution that has most disorder. A modern mathematical solution to this problem is the notion of a Gibbs ensemble [Ru]. In this context we close a small gap about finite-level quantum systems in Section 8.1. Before, the solutions were unknown for singular states. We show that the rI-closure of the Gibbs ensembles is a suitable solution.

In 1978 Barndorff-Nielsen [Bar] describes the closure of an exponential family $\mathcal{E}$ of probability distributions with finite support. We repeat his proof in Section 8.2 showing that the rI-closure of $\mathcal{E}$ is the variation closure $\mathcal{E}$.

Factorizable families have multi-information as their entropy distance. This is a measure to quantify stochastic dependencies in a composite system, see Section 1.1. In Section 8.3 we analyze factorizable families for finite-level quantum systems and their multi-information in-depth. Factorizable families are covered by convex exponential families. We write an algebraic classification of convex exponential families (up to a subtle commutator condition) in Section 8.4.

8.1 Gibbs ensembles

The *rI*-closure of Gibbs ensembles consists of the set of maximizers, subject to a number of mean value constraints, of von Neumann entropy. The Gibbs ensembles are included as the invertible solutions. For these, a theorem of Klein is often used as a proof, see [Kn07, Ru]. Below we completely solve the optimization problem on the full state space.

We begin and explain the folkloric idea and solution. Let A be a matrix algebra and $a_1 \ldots, a_k \in A_{\rm sa}$ be a finite number of self-adjoint matrices. Then, given an invertible density matrix $\rho \in S(A)$ with mean values $\xi_i := \langle a_i, \rho\rangle$ for $i = 1, \ldots, k$ one can consider the problem of maximizing the von Neumann entropy $S(\sigma) = -\operatorname{tr}(\sigma \ln(\sigma))$ among all density matrices $\sigma \in \overline{S}(A)$ that have mean values

$$\langle a_i, \sigma\rangle = \xi_i \qquad \text{for} \qquad i = 1, \ldots, k \tag{8.1}$$

equal to ρ. In case of existence, the solution is the well-known **Gibbs ensemble** [Ru, In, Co]. The entity of all solutions for arbitrary invertible $\rho \in S(A)$ is the exponential family

$$\mathbb{G}(a_1, \ldots, a_k) := \left\{ \frac{e^{-\sum_{i=1}^k \beta_i a_i}}{\operatorname{tr}\left(e^{-\sum_{i=1}^k \beta_i a_i}\right)} \ : \ \beta_i \in \mathbb{R},\ i = 1, \ldots, k \right\} \tag{8.2}$$

with **inverse temperatures** $\beta_1, \ldots, \beta_k$. We can prove in Theorem 8 that the solutions $\sigma \in \overline{S}(A)$ to the problem of maximizing von Neumann entropy subject to (8.1) given arbitrary $\rho \in \overline{S}(A)$ not necessarily invertible, is the union

$$\mathbb{G}(a_1, \ldots, a_k)^{\rm cmb} = \bigcup_{p \in \mathcal{P}_U \setminus \{0\}} \left\{ \frac{p e^{-\sum_{i=1}^k \beta_i p a_i p}}{\operatorname{tr}\left(p e^{-\sum_{i=1}^k \beta_i p a_i p}\right)} \ : \ \beta_i \in \mathbb{R},\ i = 1, \ldots, k \right\}$$

for $U := \pi_{A^0_{\rm sa}}(\operatorname{Lin}\{a_1, \ldots, a_k\})$. In Remark 7.19 we explained how to calculate the projector lattice $\mathcal{P}_U$ and why the union on the right-hand side is the combinatorial extension (7.4) of the Gibbs ensembles $\mathbb{G}(a_1, \ldots, a_k)$.

Moreover it is explained for the mapping $m : \overline{S} \to \mathbb{R}^k$, $\rho \mapsto (\langle a_1, \rho\rangle, \ldots, \langle a_k, \rho\rangle)$ associating mean values, that $m|_{\mathbb{G}(a_1,\ldots,a_k)}$ is a diffeomorphism and $m|_{\mathbb{G}(a_1,\ldots,a_k)^{\rm cmb}}$ is a bijection with range the convex support of the statistic $a_1, \ldots, a_k$. Finally, given an arbitrary tuple of mean values in the convex support $(\xi_1, \ldots, \xi_k) \in m\left(\mathbb{G}(a_1, \ldots, a_k)^{\rm cmb}\right)$ the unique solution $\sigma \in \mathbb{G}(a_1, \ldots, a_k)^{\rm cmb}$ with $\langle a_i, \sigma\rangle = \xi_i$ for $i = 1, \ldots, k$ is created by

$$\sigma = \frac{p e^{-\sum_{i=1}^k \beta_i p a_i p}}{\operatorname{tr}\left(p e^{-\sum_{i=1}^k \beta_i p a_i p}\right)} \tag{8.3}$$

such that there exist a non-zero $p \in \mathcal{P}_U$ and $\beta_1, \ldots, \beta_k \in \mathbb{R}$ which solve for $j = 1, \ldots, k$ the equations (7.26)

$$-\frac{\partial}{\partial \beta_j} \ln \left(\operatorname{tr}(\mathbb{1} - p) + \operatorname{tr}(\exp(-\sum_{i=1}^{k} \beta_i p a_i p)) \right) = \xi_j. \tag{8.4}$$

To prove these statements we can use Remark 6.32 and Remark 7.19 together with Theorem 8. We changed signs of the scalars $\{\lambda_i\}_{i=1}^k$ appearing in the two remarks and we use inverse temperatures $\{\beta_i\}_{i=1}^k$ by custom in thermodynamics. When the Gibbs ensembles $\mathbb{G}(a_1, \ldots, a_k)$ are considered then $p = \mathbb{1}$ and the conditions (8.4) simplify to the well-known equations with free energy $-\frac{\partial}{\partial \beta_j} F(-\sum_{i=1}^k \beta_i a_i) = \xi_j$, see page 125 in [In]. The von Neumann entropy of the solution (8.3) is

$$\ln(\operatorname{tr}(p e^{-\sum_{i=1}^k \beta_i p a_i p})) + \sum_{i=1}^{k} \beta_i \xi_i.$$

The von Neumann entropy simplifies to $F\left(-\sum_{i=1}^k \beta_i a_i\right) + \sum_{i=1}^k \beta_i \xi_i$ for Gibbs ensembles.

Remark 8.1 Below we consider a vector space $U \subset A_{\mathrm{sa}}^0$ and the exponential family $\mathcal{E} := \exp_1(U)$. Then U is the canonical parameter space and the canonical tangent space of $\mathcal{E}$. The combinatorial mean value parametrization (7.15) is the bijection

$$M^{\mathrm{cmb}} : \mathrm{sr}_U \to \mathcal{E}^{\mathrm{cmb}}$$

between the state reflection sr_U and the combinatorial extension $\mathcal{E}^{\mathrm{cmb}}$ of $\mathcal{E}$ (7.4). The inverse to M^{cmb} is the combinatorial mean value chart $\pi^{\mathrm{cmb}} = \pi_U|_{\mathcal{E}^{\mathrm{cmb}}}$.

Lemma 8.2 Let $U \subset A_{\mathrm{sa}}^0$ be a vector space and put $\mathcal{E} := \exp_1(U)$. For each $x \in \mathrm{ri}(\mathrm{sr}_U)$ there exists a unique density matrix $\sigma \in \overline{S}(A)$ that maximizes von Neumann entropy on the set $\overline{S}(A) \cap (x + U^{\perp})$ and the density matrix is given by the mean value parametrization $\sigma = M(x)$ for $\mathcal{E}$. [Proof on page 240]

Remark 8.3 If the solutions (8.2) were unknown to us, they can be found with the method of Lagrange multipliers that gives a necessary condition for invertible solutions. However, the (strictly concave) von Neumann entropy may have *a priori* a maximum on the relative boundary of the constraint set but none in the relative interior. The method is not an alternative proof for Lemma 8.2. The inverse temperatures $\beta_1, \ldots, \beta_k$ appear as Lagrange multipliers. [Proof on page 241]

Theorem 8 Let $U \subset A_{\mathrm{sa}}^0$ be a vector space and put $\mathcal{E} := \exp_1(U)$. For each $x \in \mathrm{sr}_U$ there exists a unique density matrix $\sigma \in \overline{S}(A)$ that maximizes von Neumann entropy on the set $\overline{S}(A) \cap (x + U^\perp)$ and the density matrix is given by the combinatorial mean value parametrization $\sigma = M^{\mathrm{cmb}}(x) \in \mathcal{E}^{\mathrm{cmb}}$ for $\mathcal{E}$. [Proof on page 242]

8.2 The abelian case

We write with full proof the equality of rI-closure and variation closure for an exponential family in an abelian matrix algebra. The equality of the two closures is equivalent to the description of the closure of an exponential family of probability distributions in Barndorff-Nielsen [Bar]. A consequence of the result is the continuity of entropy distance.

Definition 8.4 If $\mathcal{E}$ is an exponential family in an abelian matrix algebra then $\mathcal{E}$ is called a **classical family**.

Remark 8.5 **(a)** Closures of an exponential family of probability distributions, dominated by a Borel measure on $\mathbb{R}^d$, are well understood, see Csiszár and Matúš [Cs05]. The much simpler case of finite support is known for longer, see Barndorff-Nielsen [Bar]. The case of finite support corresponds to our notion of a classical family. The link is the following.

(b) We consider the matrix algebra $A = \mathbb{C}^N$ and put $X := \{1, \dots, N\}$. The support of a vector $x \in \mathbb{C}^N$ is $\mathrm{supp}(x) = \{i \in X : x_i \neq 0\}$. The support induces a lattice isomorphism between the projector lattice $\mathcal{P}$ of A and the power set 2^X of X

$$\mathcal{P} \to 2^X, \qquad p \mapsto \mathrm{supp}(p),$$

see Remark 2.36 (c). This isomorphism implies two connections from the quantum to the classical world:

Firstly, the atoms of the projector lattice $\mathcal{P}$, the rank one projectors, correspond to the elements of X. We can write any projector as a sum of these elements (the atoms are only in the abelian case a complete set of projectors, i.e. mutually orthogonal and summing to $\mathbb{1}$). The state space $\overline{S}(A)$ is identified with the probability simplex (1.25)

$$\overline{S}(A) = \{\rho \in \mathbb{R}^N : \sum_{i=1}^{N} \rho_i = 1, \rho_i \geq 0 \text{ for } i = 1, \dots, N\}$$

by measurement probabilities on X. The probability simplex is a simplex in the geometric sense, see Example 4.9.

Secondly, the inverse $E := (\mathrm{supp}\,|_{\mathcal{P}})^{-1}$ of the support, restricted to the projector lattice,

$$E : 2^X \to \mathcal{P}$$

is a von Neumann measurement (Definition 1.13). Given a subset $Z \subset X$ and a classical system $\rho \in \overline{S}(A)$ with positive measurement probability

$$P_\rho(Z) := \sum_{z\in Z} \mathrm{tr}(\rho E(z)) = \sum_{z\in Z} \rho_z > 0,$$

the state reduction

$$\frac{\rho E(Z)}{\mathrm{tr}(\rho E(Z))}$$

is the probability density of the conditional probability distribution $P_\rho(\,\cdot\,|Z)$, see (1.22). In the proof of Proposition 8.6 the σ_i's are such that

$$P_{\sigma_i} = P_{\rho_i}(\,\cdot\,|\,\mathrm{supp}(p))$$

for a non-zero projector $p \in \mathcal{P}$. Conditional probability distributions are the natural way to construct the closure of an exponential family, see Barndorff-Nielsen [Bar] or see Section 1.3.

(c) A **polytope** in $\mathbb{R}^N$ is the convex hull of a finite point set. Equivalently, a polytope is the orthogonal projection of a simplex (in some higher dimensional space) to a linear space (Theorem 2.15 in [Zi]). The state reflection sr_U is defined as the orthogonal projection of the state space to the canonical tangent space U of an exponential family. Since the state space of $A = \mathbb{C}^N$ is a simplex, the state reflection is a polytope. It is proved in Theorem 75 in [Brø], that every face of a polytope is an exposed face. Thus we obtain equality $\mathcal{P}_U = \mathcal{P}_{U,\perp}$ of projector lattice (5.50) and exposed projector lattice (5.11).

Proposition 8.6 If $\mathcal{E}$ is a classical family then $\mathrm{cl}_{rI}(\mathcal{E}) = \overline{\mathcal{E}}$. [Proof on page 242]

Corollary 8.7 If $\mathcal{E}$ is a classical family then the entropy distance from $\mathcal{E}$ is continuous. [Proof on page 244]

Remark 8.8 The Staffelberg family in Example 7.21—an exponential family $\mathcal{E}$ in the non-abelian algebra $M_2 \oplus \mathbb{C}$—has $\mathrm{cl}_{rI}(\mathcal{E}) \subsetneq \overline{\mathcal{E}}$. For this family the entropy distance is discontinuous, see Example 7.28.

8.3 Multi-information

Multi-information is a measure of stochastic dependence between stochastic units. It has applications in Neuroscience and Statistical Mechanics, see the end of Section 1.1. It is well-known in the classical case that multi-information is the entropy distance from a family of factorizable probability distributions. We generalize the distance description to the case of a matrix algebra. We describe the structure of the variation closure of a factorizable family, which is equal to the rI-closure. We show continuity of multi-information.

Another subject is the question about maximizers of multi-information. As an evidence of infomax principles in Neuroscience, we prove an upper bound on the rank of a local maximizer of multi-information. The bound grows like the square-root of the number of units. This is a quadratic improvement over the classical case.

A composite quantum system is modeled on the tensor product (also called Kronecker product) of the Hilbert spaces of the individual particles (or units) [Pe08]. A source for definition of the Kronecker product can be Chapter I.4 in [Bh].

Definition 8.9 For $N \geq 1$ we consider the set $[N] = \{1, \ldots, N\}$ whose elements are called **units**. To each unit $i \in [N]$ we associate a matrix algebra A_i. We denote by $\mathbb{1}_i$ the identity and by 0_i the zero element in A_i for each unit $i \in [N]$. Let $a_i \in A_i$ for $i \in [N]$. For a non-empty subset $I \subset [N]$ we abbreviate the **Kronecker product**

$$a_I := \bigotimes_{i \in I} a_i$$

where the product is understood in lexicographical order of the indices. The **joint algebra** of the units I is the matrix algebra of dimension $\prod_{i \in I} \dim(A_i)$

$$A_I := \bigotimes_{i \in I} A_i.$$

This consists of all linear combinations of Kronecker products a_I with $a_i \in A_i$ for $i \in I$. (Direct sum and tensor product are connected by a distributive law [La].) We put $A_\emptyset := \mathbb{C}$ and $a_\emptyset := 1 \in \mathbb{C}$. We denote by $\mathbb{1}_I$ respectively by 0_I the identity respectively the zero element in the algebra A_I. The **Kronecker sum** is defined for non-empty I by

$$a \cdot I := \sum_{i \in I} \mathbb{1}_{I \setminus \{i\}} \otimes a_i$$

and we use the convention $a \cdot \emptyset = 0 \in \mathbb{C}$.

Remark 8.10 For disjoint subsets $I, J \subset [N]$ and $a_i, b_i \in A_i$ for $i \in I \cup J$ one has the equality

$$\operatorname{tr}(a_I) = \prod_{i \in I} \operatorname{tr}(a_i). \tag{8.5}$$

Hence for the Hilbert-Schmidt inner product

$$\langle a_I, b_I \rangle = \prod_{i \in I} \langle a_i, b_i \rangle \qquad \text{and} \qquad \|a_I\|_2 = \prod_{i \in I} \|a_i\|_2$$

holds. For $t \in \mathbb{C}$ and $b_i := a_i t$ for $i \in I$, when $(at) \cdot I$ is understood as $b \cdot I$ then

$$(ta) \cdot I = t(a \cdot I) \qquad \text{and} \qquad (a+b) \cdot I = a \cdot I + b \cdot I. \tag{8.6}$$

We use the notation

$$a_I \otimes a_J := a_{I \cup J} \quad \text{and} \quad A_I \otimes A_J := A_{I \cup J}.$$

Then

$$a \cdot (I \cup J) = (a \cdot I) \otimes \mathbb{1}_J + (a \cdot J) \otimes \mathbb{1}_I. \tag{8.7}$$

Lemma 8.11 For $I \subset [N]$ and $a_i \in A_i$, $i \in I$, we have $\exp(a \cdot I) = \bigotimes_{i \in I} \exp(a_i)$ and $\exp_1(a \cdot I) = \bigotimes_{i \in I} \exp_1(a_i)$. [Proof on page 244]

Definition 8.12 For $I \subset [N]$ a matrix in A_I is **factorizable** if it is of the form a_I for $a_i \in A_i$ with $i \in I$. We consider the **factorizable family** in A_I

$$\mathcal{F}_A(I) := \{\rho_I \ : \ \rho_i \in S(A_i),\ i \in I\} \subset S(A_I) \tag{8.8}$$

consisting of the factorizable and invertible density matrices in A_I. We consider also the vector space

$$\chi_A(I) := \left\{a \cdot I \ : \ a_i \in (A_i)^0_{\mathrm{sa}},\ i \in I\right\} \subset (A_I)^0_{\mathrm{sa}}. \tag{8.9}$$

We drop the references, $\mathcal{F}(I) = \mathcal{F}_A(I)$ and $\chi(I) = \chi_A(I)$, to the algebra if they are not necessary.

Remark 8.13 **(a)** The simplest non-trivial example of a factorizable family is depicted in Figure 6.1 on page 129 where $I = \{1, 2\}$ and $A_1 = A_2 = \mathbb{C}^2$.

(b) For $I \subset [N]$ one has by Lemma 8.11 the equation $\mathcal{F}(I) = \exp_1(\chi(I))$. In particular, $\chi(I)$ is the canonical parameter space and the canonical tangent space of $\mathcal{F}(I)$. If $\mathcal{F}(I) \neq \emptyset$ then

$$\dim(\mathcal{F}(I)) = \dim(\chi(I)) = \sum_{i \in I} \dim_{\mathbb{R}}\left((A_i)^0_{\mathrm{sa}}\right) = \sum_{i \in I} \dim_{\mathbb{C}}(A_i) - |I|. \tag{8.10}$$

Since the factorizable family is an exponential family we can study it using the extensions developed in previous chapters.

Lemma 8.14 For $I \subset [N]$ we have $\overline{\mathcal{F}(I)} = \{\rho_I \;:\; \rho_i \in \overline{S}(A_i),\; i \in I\}$. [Proof on page 244]

Proposition 8.15 For $I \subset [N]$ we have $\overline{\mathcal{F}(I)} = \bigcup_{\substack{p \in \mathcal{P}_{\chi(I),\perp} \\ p \neq 0}} \kappa^p(\mathcal{F}(I)_p)$. [Proof on page 244]

To appreciate Proposition 8.15 we calculate the projector lattice $\mathcal{P}_{\chi(I)}$ and for non-zero $p \in \mathcal{P}_{\chi(I)}$ the exponential families $\kappa^p(\mathcal{F}(I)_p)$. Traceless compressions ς^p are used to make their structure more transparent (5.60). We denote the projector lattice of the algebra A_i by $\mathcal{P}(A_i)$ and we use for $p_i \in \mathcal{P}(A_i)$ the compression $B_i := A_i^{p_i}$, $i \in I$.

Lemma 8.16 If $I \subset [N]$ and $a_i \in (A_i)_{\text{sa}}$ for $i \in I$ then $s(a_I) = \bigotimes_{i \in I} s(a_i)$ and $p_+(a \cdot I) = \bigotimes_{i \in I} p_+(a_i)$. [Proof on page 245]

Lemma 8.17 For $I \subset [N]$ we have $\mathcal{P}_{\chi(I),\perp} = \mathcal{P}_{\chi(I)} = \{p_I : p_i \in \mathcal{P}(A_i), i \in I\}$. [Proof on page 245]

Proposition 8.18 If $I \subset [N]$ and a projector $p_I \in \mathcal{P}_{\chi(I)}$ is defined by non-zero projectors $p_i \in \mathcal{P}(A_i)$ for $i \in I$ then $\varsigma^{p_I}(\chi_A(I)) = \chi_B(I)$ and $\mathcal{F}_A(I)_{p_I} = \exp_1(\chi_B(I))$ is a factorizable family in B_I. Moreover

$$\kappa^{p_I}(\mathcal{F}_A(I)_{p_I}) = \left\{ \frac{p_I e^{\theta \cdot I}}{\operatorname{tr}(p_I e^{\theta \cdot I})} \;:\; \theta_i \in \operatorname{lin}(\mathbb{F}(p_i)),\; i \in I \right\}.$$

[Proof on page 246]

Remark 8.19 Consider disjoint subsets $I, J \subset [N]$. For a factorizable matrix $a_{I \cup J} \in A_{I \cup J}$ with $a_i \in A_i$, $i \in I \cup J$, the **partial trace** of a to I is defined by

$$\operatorname{tr}_I(a_{I \cup J}) := a_I \operatorname{tr}(a_J).$$

The partial trace extends consistently to a linear mapping

$$\mathrm{tr}_I : A_{I\cup J} \to A_I.$$

For arbitrary $a \in A_I$ and $b \in A_{I\cup J}$ we have

$$\mathrm{tr}\left((a \otimes \mathbb{1}_J)b\right) = \mathrm{tr}(a\,\mathrm{tr}_I(b)). \tag{8.11}$$

If b is self-adjoint then $\mathrm{tr}_I(b)$ is self-adjoint. This follows from the fact that $\mathrm{tr}(p\,\mathrm{tr}_I(b)) = \mathrm{tr}\left((p \otimes \mathbb{1}_J)b\right)$ is real for an arbitrary projector $p \in \mathcal{P}(A_I)$. Moreover, for $b \geq 0$ one has $\mathrm{tr}(p\,\mathrm{tr}_I(b)) = \mathrm{tr}\left((p \otimes \mathbb{1}_J)b\right) \geq 0$, hence $\mathrm{tr}_I(b) \geq 0$ if $b \geq 0$. Altogether and with $\mathrm{tr}(\mathrm{tr}_I(b)) = \mathrm{tr}(b)$ the partial trace preserves the state space,

$$\mathrm{tr}_I|_{\overline{S}(A_{I\cup J})} \;:\; \overline{S}(A_{I\cup J}) \to \overline{S}(A_I)$$

is a surjective mapping.

Lemma 8.20 Let $I \subset [N]$, let $a \in A_I$ and for $i \in I$ let $u_i \in (A_i)^0_{\mathrm{sa}}$ and put $b_i := \mathrm{tr}_{\{i\}}\left(\pi_{(A_I)^1_{\mathrm{sa}}}(a)\right)$. Then $\langle u \cdot I, a\rangle = \langle u \cdot I, b_I\rangle$. [Proof on page 246]

Lemma 8.21 Let $I \subset [N]$ and $\rho \in \overline{S}(A_I)$. The combinatorial normal projection for the factorizable family $\mathcal{F}(I)$ evaluated in ρ is $N^{\mathrm{cmb}}(\rho) = \bigotimes_{i\in I} \mathrm{tr}_{\{i\}}(\rho)$. [Proof on page 247]

Theorem 9 Let $I \subset [N]$ and $\rho \in \overline{S}(A_I)$. The entropy distance from the factorizable family is a difference of von Neumann entropies $S_{\mathcal{F}(I)}(\rho) = \sum_{i\in I} S(\mathrm{tr}_{\{i\}}(\rho)) - S(\rho)$.
[Proof on page 247]

Definition 8.22 Let $I \subset [N]$. The difference of von Neumann entropies

$$\overline{S}(A_I) \to \mathbb{R}, \qquad \rho \mapsto \sum_{i\in I} S(\mathrm{tr}_{\{i\}}(\rho)) - S(\rho) \tag{8.12}$$

is called **multi-information**.

Remark 8.23 Let $I \subset [N]$. **(a)** Multi-information is continuous on $\overline{S}(A_I)$ for the following reasons. Multi-information is entropy distance from the factorizable family $\mathcal{F}(I)$ (Theorem 9). Since the rI- and variation closures of $\mathcal{F}(I)$ coincide by Proposition 8.15 and Remark 7.27 (a) one obtains continuity of entropy distance from Theorem 7.

(b) If $\rho \in \overline{S}(A_I)$ is a local maximizer of multi-information then the estimate

$$\dim(F(\rho)) \leq \dim(\mathcal{F}(I)) = \sum_{i \in I} \dim_{\mathbb{C}}(A_i) - |I| \tag{8.13}$$

holds. The reason is the following. A maximizer $\rho \in \overline{S}(A)$ of entropy distance from the factorizable family $\mathcal{F}(I)$ has the face dimension $F(\rho)$ bounded by the dimension of $\mathcal{F}(I)$, see Proposition 7.29. The dimension of the exponential family is calculated in (8.10).

(c) The rank of a density matrix $\rho \in \overline{S}(A_I)$ is a measure of disorder. We compare the rank of a local maximizer for composite classical and quantum systems of the same level sizes $n_i \in \mathbb{N}$ and $i \in I$. In the classical case the algebras are $A_i := \mathbb{C}^{n_i}$ for $i \in [N]$ and the factorizable family has dimension $\dim(\mathcal{F}(I)) = \sum_{i \in I} n_i - |I|$. This implies for a local maximizer $\rho \in A_I$ of multi-information the rank bound, see (7.31) and (8.13),

$$\mathrm{rk}(\rho) \leq \sum_{i \in I} n_i - |I| + 1. \tag{8.14}$$

When the matrix algebra is chosen maximal in the quantum case, that is $A_i = M_{n_i}$ for $i \in I$ then $\dim(\mathcal{F}(I)) = \sum_{i \in I} n_i^2 - |I|$. Now the face dimension of a local maximizer ρ is quadratic in the rank so (7.32)

$$\mathrm{rk}(\rho) \leq \sqrt{\sum_{i \in I} n_i^2 - |I| + 1}. \tag{8.15}$$

We have found a sub-linear growth of the rank of a local maximizer of multi-information in the number $|I|$ of units compared to the exponential growth of the maximal possible rank

$$\mathrm{rk}(\mathbb{1}) = \prod_{i \in I} n_i.$$

In dependence of the number of units this is a quadratic improvement of the quantum case over the classical case.

(d) The discussion in (c) is a worst-case scenario. The classical estimate for equal units $A_i = \mathbb{C}^n$, $i \in [N]$ becomes redundant when a global maximizer $\rho \in A_I$ is considered. Then

$$\mathrm{rk}(\rho) = n \tag{8.16}$$

independent of the number $|I|$ of units [AK]. Indeed, ρ has a completely deterministic dependence among the units while one fixed unit is uniformly distributed. A similar deterministic behavior of a local maximizer is true for a much larger class of abelian algebras.

8.4 Convex families

Factorizable families have convex sub-families allowing for sharper results on maximizers of multi-information[1]. Analysis of the convex sub-families implies curvature estimates[2]. We describe the closure of a convex exponential family. Modulo a more subtle commutator relation we can give an algebraic characterization of convex exponential families. This characterization is done for the classical case in [MA].

Example 8.24 It is observed in Remark 6.2 on page 129 that the factorizable family of density matrices in the algebra $\mathbb{C}^2 \otimes \mathbb{C}^2$ is a piece of a hyperbolic paraboloid. In particular, it is covered by two transverse families of straight line segments, which are exponential families by themselves. More generally, for $N \in \mathbb{N}$, a factorizable family (studied in Section 8.3) in the N-fold tensor product $A_1 \otimes \cdots \otimes A_N$ with matrix algebras A_i, $i = 1, \ldots, N$ is covered by N transverse families of convex exponential sub-families.

Definition 8.25 An exponential family $\mathcal{E}$ in a matrix algebra A is a **convex family** if $\mathcal{E}$ is convex. This means that the segment $[\rho, \sigma]$ belongs to $\mathcal{E}$ when $\rho, \sigma \in \mathcal{E}$. As before we write Θ for the canonical parameter space and U for the canonical tangent space of $\mathcal{E}$.

We describe the closure of a convex family and find that the state reflection has only exposed faces.

Lemma 8.26 If $\mathcal{E}$ is a convex family then the restricted projection $\pi_U|_{\mathrm{aff}(\mathcal{E})} : \mathrm{aff}(\mathcal{E}) \to U$ is an affine isomorphism and $\overline{\mathcal{E}} = \mathrm{aff}(\mathcal{E}) \cap \overline{S}(A)$. [Proof on page 248]

Proposition 8.27 If $\mathcal{E}$ is a convex family then $\overline{\mathcal{E}} = \bigcup_{p \in \mathcal{P}_{U,\perp} \setminus \{0\}} \kappa^p(\mathcal{E}_p)$.
[Proof on page 248]

In addition to the easy structured variation closure of a convex family established in

[1] Private communication by N. Ay.
[2] Private communication by A. Knauf.

Proposition 8.27 and discussed in Remark 7.27 (a), a strong algebraic characterization is available. To this end let us introduce some basic ideas.

Remark 8.28 **(a)** We consider for an orthogonal projector $p \in \mathcal{P}$ the complementary projector $p' = \mathbb{1} - p$. Then for a vector space $V \subset A^0_{\rm sa}$ of traceless self-adjoint matrices we have

$$p \in V + \mathbb{R}\mathbb{1} \qquad \Longleftrightarrow \qquad p\,\mathrm{tr}(p') - p'\,\mathrm{tr}(p) \in V. \tag{8.17}$$

(b) We use for $a, b \in A$ the **commutator** notation $[a, b] = ab - ba$ and we say a and b **commute** or a **commutes with** b if $[a, b] = 0$. Recall (Satz 5.27 in [Kn01]) if two normal matrices $a, b \in A$ commute then there exists an orthogonal basis of the Hilbert space of A consisting of vectors that are eigenvectors both of a and of b. In particular, the spectral projectors of a and b commute and for arbitrary functions $f : \mathrm{spec}(a) \to \mathbb{C}$ and $g : \mathrm{spec}(b) \to \mathbb{C}$ we have $[f(a), g(b)] = 0$ by functional calculus.

Lemma 8.29 If there is $\theta \in \Theta$ that commutes with every matrix in U and if $U + \mathbb{R}\mathbb{1}$ contains the spectral projectors of every matrix in U then $\mathcal{E}$ is a convex family.[3]

[Proof on page 248]

The one-dimensional case of Lemma 8.29 is the following.

Remark 8.30 If $\theta, u \in A^0_{\rm sa}$ commute and u has at most two different eigenvalues then the e-geodesic $\mathbb{R} \to S(A)$, $\lambda \mapsto \exp_1(\theta + \lambda u)$ is convex.

Modulo commutator relations, the converse to Lemma 8.29 is the following.

Proposition 8.31 If $\mathcal{E}$ is a convex family and $u \in U$ commutes with some $\theta \in \Theta$ then $U + \mathbb{R}\mathbb{1}$ contains the spectral projectors of u. [Proof on page 249]

Remark 8.32 One can think that a convex family $\mathcal{E}$ has the canonical tangent space U

[3] The lemma and proof are a private communication by A. Knauf.

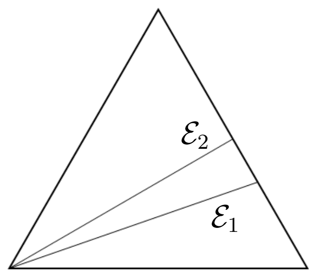


Figure 8.1: The drawing shows for $\theta := \ln(2)(-3, 2, 1)$ and $u := (-2, 1, 1)$ the convex families $\mathcal{E}_1 := \exp_1(\theta + \mathbb{R}u)$ and $\mathcal{E}_2 := \exp_1(\mathbb{R}u)$ within the state space of $\mathbb{C}^3$. Both families have the same canonical tangent space $\mathbb{R}u$.

equal to the traceless self-adjoint part of a sub-algebra of A. This is not true in general. An example is $A = M_2$ with $\Theta = U = \mathrm{Lin}\{\sigma_1, \sigma_2\}$ and Pauli matrices σ_1, σ_2. Using $b(\varphi) := (\cos(\varphi), \sin(\varphi), 0)$ we can parametrize $\mathcal{E}$ by $\mathbb{R}^+ \times [0, 2\pi) \to \mathcal{E} \setminus \{\frac{\mathbb{1}}{2}\}$, where (4.8)

$$(t, \varphi) \mapsto \exp_1(tb(\varphi)\widehat{\sigma}) = \frac{1}{2}(\mathbb{1} + \tanh(t)b(\varphi)\widehat{\sigma}).$$

Then $\mathcal{E}$ consists of the relative interior of the disk which is the convex hull of the equator $\{b(\varphi)\widehat{\sigma} : \varphi \in [0, 2\pi)\}$ of the Bloch sphere. This proves that $\mathcal{E}$ is convex. On the other hand, the product $\sigma_1\sigma_2 = i\sigma_3$ is not an element of $U + iU + \mathbb{C}\mathbb{1}$. Hence $U + \mathbb{R}\mathbb{1}$ is not the self-adjoint part of an algebra.

Question 12 What is the meaning of the commutator condition in Lemma 8.29 and in Proposition 8.31 ?

An approach to this question can be the following.

Question 13 If $\mathcal{E}$ is a convex family, is then the exponential family $\exp_1(U)$ convex?

The affirmative to Question 13 implies that the commutator condition can be dropped in Proposition 8.31. By example, the commutator condition can not be dropped from Lemma 8.29. The exponential family $\exp_1(\sigma_1 + \mathbb{R}\sigma_2)$ with Pauli matrices σ_1, σ_2 is not convex while $\exp_1(\mathbb{R}\sigma_2)$ is. Both families have the same antipodal boundary points on the Bloch sphere by Lemma 7.6 but only the second family contains the trace state $\frac{\mathbb{1}_2}{2}$ which is the center of the Bloch sphere.

The affirmative to Question 13 does not imply (by Minkowski Theorem) that $\overline{\exp_1(U)} = \mathcal{E}$. The extreme points of $\overline{\mathcal{E}}$ may not be pure. Then $\overline{\exp_1(U)}$ can have different extreme points as $\overline{\mathcal{E}}$. An example is depicted in Figure 8.1.

9 Application: Stationary Markov transitions

We extend the range of multi-information to a dynamical situation by considering a Markovian kernel on a composite system. Interaction measures for these systems are investigated in experimental and theoretical Neuroscience, see e.g. Linsker [Lin] or Tononi et al. [To]. It is believed that complex structures in biology can emerge through self-organization governed by very simple laws like maximization of interaction measures.

Here we treat a temporal interaction measure defined by Ay [Ay01] and we can prove a high degree of determinism for local maximizers of the measure. With increasing effort we prove three bounds of reduced disorder. Methods include a maximization of multi-information in the context of a joint probability simplex (Remark 9.7) and a cyclic decomposition of Markov transitions (Remark 9.16 and Remark 9.18). Cyclic decompositions of Markov transitions are known from Kalpazidou [KaS]. The geometric counterpart is proved here in detail for the purpose of dimension estimates based on convex geometry.

Definition 9.1 The **configuration space** Ω is an arbitrary finite set. We write $P(\Omega)$ for the set of **probability distributions** on the configuration space Ω:

$$P(\Omega) := \big\{ p \in \mathbb{R}^{\Omega} : p(\omega) \geq 0 \text{ for all } \omega \in \Omega \text{ and } \sum_{\omega \in \Omega} p(\omega) = 1 \big\}.$$

In this section we write the coefficients of a vector or matrix in brackets behind the symbol. We call $P(\Omega)$ also the **probability simplex on** Ω and $P(\Omega \times \Omega)$ is called the **joint probability simplex on** Ω.

Remark 9.2 The probability simplex on Ω is the state space $\overline{S}(A)$ of the abelian matrix algebra $A = \mathbb{C}^{\Omega}$ treated in Section 8.2. The geometric studies take place in the Euclidean space of self-adjoint elements $A_{\mathrm{sa}} = \mathbb{R}^{\Omega}$. The matrix algebra $A = \mathbb{C}^{\Omega}$ is isometrically identified with the Hilbert space $H = \mathbb{C}^{\Omega}$. The trace of $x \in A = H$ is $\mathrm{tr}(x) = \sum_{\omega \in \Omega} x(\omega)$. Mainly we will address the algebra $A = \mathbb{C}^{\Omega \times \Omega} = \mathbb{C}^{\Omega} \otimes \mathbb{C}^{\Omega}$ and Kronecker products of such algebras. The Kronecker product $\otimes$ is introduced in Section 8.3.

Definition 9.3 A **Markov transition kernel** on Ω is defined as a function

$$k : \Omega \times \Omega \longrightarrow [0,1], \quad (\omega, \omega') \longmapsto k(\omega|\omega') \tag{9.1}$$

such that $k(\cdot|\omega') \in P(\Omega)$ for all $\omega' \in \Omega$. We write $K(\Omega)$ for the set of all Markov transition kernels on Ω. A pair $(p, k) \in P(\Omega) \times K(\Omega)$ is called a **Markov transition** on Ω. The **joint distribution** of a Markov transition $(p, k) \in P(\Omega) \times K(\Omega)$ is given by the vector $J(p, K) \in \mathbb{R}^{\Omega \times \Omega}$ with coefficients

$$J(p, k)(\omega, \omega') := p(\omega')k(\omega|\omega') \tag{9.2}$$

for $\omega, \omega' \in \Omega$. The **first marginal** of a vector $x \in \mathbb{C}^{\Omega\times\Omega}$ is defined as the vector $x_1 \in \mathbb{C}^{\Omega}$ with coefficients for $\omega \in \Omega$

$$x_1(\omega) := \sum_{\omega' \in \Omega} x(\omega, \omega'). \tag{9.3}$$

The **second marginal** of x is the vector $x_2 \in \mathbb{C}^{\Omega}$ with coefficients for $\omega' \in \Omega$

$$x_2(\omega') := \sum_{\omega \in \Omega} x(\omega, \omega'). \tag{9.4}$$

A Markov transition $(p, k) \in P(\Omega) \times K(\Omega)$ is **stationary** if $J(p, k)_1 = p$. We denote the set of stationary Markov transitions on Ω by

$$T(\Omega) := \{(p, k) \in P(\Omega) \times K(\Omega) \ : \ J(p, k)_1 = p\}. \tag{9.5}$$

We define the **cycle space** of Ω as the vector space

$$C(\Omega) := \{x \in \mathbb{C}^{\Omega\times\Omega} \ : \ x_1 = x_2\}. \tag{9.6}$$

The elements of $C(\Omega)$ are called **cycles**. The **Kirchhoff polytope on** Ω is

$$\mathrm{Kirch}(\Omega) := P(\Omega \times \Omega) \cap C(\Omega). \tag{9.7}$$

Remark 9.4 **(a)** The first marginal x_1 of a vector $x \in \mathbb{C}^{\Omega\times\Omega} = \mathbb{C}^{\Omega} \otimes \mathbb{C}^{\Omega}$ is a special case of the partial trace. Indeed, association of a first marginal is a linear mapping, thus we may argue with a factorizable vector $x = y \otimes z$ for $y, z \in \mathbb{C}^{\Omega}$. One has for $\omega \in \Omega$

$$x_1(\omega) = \sum_{\omega' \in \Omega} y(\omega) \otimes z(\omega') = y(\omega)\,\mathrm{tr}(z) = [\mathrm{tr}_{\{1\}}(x)](\omega),$$

that is $x_1 = \mathrm{tr}_{\{1\}}(x)$ and similarly for the second marginal one has $x_2 = \mathrm{tr}_{\{2\}}(x)$. In particular, if $p \in P(\Omega \times \Omega)$ is a probability distribution then both the first marginal p_1 and the second marginal p_2 are probability distributions on Ω.

(b) For a Markov transition $(p, k) \in P(\Omega) \times K(\Omega)$ the joint distribution $J(p, k)$ is a probability distribution on $\Omega \times \Omega$ and the second marginal is $J(p, k)_2 = p$. If $\Omega = \{1, \dots, n\}$ for $n \in \mathbb{N}$ it is customary to write the Markov transition kernel k in matrix notation

$$k = \begin{pmatrix} k(1|1) & \cdots & k(n|1) \\ \vdots & & \vdots \\ k(1|n) & \cdots & k(n|n) \end{pmatrix}.$$

Then the rows of k are probability distributions. This is a transposed matrix notation, and we denote the joint distribution of (p, k) in the same way

$$J(p, k) = \begin{pmatrix} p(1)k(1|1) & \cdots & p(1)k(n|1) \\ \vdots & & \vdots \\ p(n)k(1|n) & \cdots & p(n)k(n|n) \end{pmatrix}$$

to preserve orientation between matrix entries and indices where the first index $\omega \in \Omega$ labels the columns and the second index $\omega' \in \Omega$ the rows of $k(\omega|\omega')$ or of $J(p, k)(\omega, \omega')$. With the usual matrix product one has for a row vector $p = (p_1, \dots, p_n)$ the first marginal of the joint distribution

$$J(p, k)_1 = pk. \tag{9.8}$$

If a system is in the state p then the first marginal pk is the state of the system after one application of the Markov transition kernel k. It is instructive to view k as a dynamics acting on probability distributions. A stationary Markov transition (p, k) is characterized with (9.8) by the equation

$$p = pk.$$

The probability distribution p is a fixed point of the dynamics of k likewise p is a (left) eigenvector of k.

(c) The disorder of a Markov transition $(p, k) \in P(\Omega) \times K(\Omega)$ can be quantified by the **conditional entropy** [Ay03] with von Neumann entropy S

$$H(p, k) := \sum_{\omega \in \Omega} p(\omega) S(k(\cdot|\omega)).$$

Lemma 9.5 The map $J : P(\Omega) \times K(\Omega) \to P(\Omega \times \Omega)$ is surjective. Two Markov transitions (p, k) and (p', k') have the same image $J(p, k) = J(p', k')$ if and only if $p = p'$ and $k(\cdot|\omega') =$

$k'(\cdot|\omega')$ for all $\omega' \in \mathrm{supp}(p)$. Stationary Markov transitions correspond (in a non-unique way) to points in the Kirchhoff polytope, $T(\Omega) = J^{-1}(\mathrm{Kirch}(\Omega))$. [Proof on page 250]

Definition 9.6 (Temporal interaction) For $N \geq 1$ and the set $[N] = \{1, \ldots, N\}$ of units a (local) configuration space is a finite set Ω_i for $i \in [N]$ and the configuration space is $\Omega := \times_{i \in [N]} \Omega_i$. The **local joint algebra** is $A_i := \mathbb{C}^{\Omega_i} \otimes \mathbb{C}^{\Omega_i}$ for a unit $i \in [N]$ and the **temporal marginal algebra** is $\mathbb{C}^{\Omega} = \bigotimes_{i \in [N]} \mathbb{C}^{\Omega_i}$. Let us consider the factorizable family in $A_{[N]}$

$$\mathcal{F}([N]) := \{Q_{[N]} \;:\; Q_i \in S(A_i),\; i \in [N]\} \subset S(A_{[N]})$$

with canonical tangent space $\chi([N])$ (Definition 8.12). The factorizable family has dimension $\dim(\mathcal{F}([N])) = \sum_{i \in [N]} |\Omega_i|^2 - N$ (8.10). Given a Markov transition $(p, k) \in P(\mathbb{C}^{\Omega}) \times K(\mathbb{C}^{\Omega})$ the entropy distance of the joint distribution $J(p, k)$ from $\mathcal{F}([N])$

$$\inf_{Q \in \mathcal{F}([N])} S(J(p, k), Q) \tag{9.9}$$

is a measure of **temporal interaction** for the Markov transition (p, k). This is discussed in Example 2.1 and in Example 3.1 (3) in [Ay01].

Remark 9.7 (Local maxima on $P(\Omega) \times K(\Omega)$.) Temporal interaction (9.9) of a Markov transition (p, k) is multi-information of the joint distribution $J(p, k)$ with respect to the factorizable family $\mathcal{F}([N])$ in the Kronecker product $A_{[N]}$ of local joint algebras $A_i = \mathbb{C}^{\Omega_i} \otimes \mathbb{C}^{\Omega_i}$, $i \in [N]$. A local maximizer of multi-information on $P(\Omega) \times K(\Omega)$ satisfies the support bound (8.14)

$$\mathrm{supp}(J(p, k)) \leq \sum_{i \in [N]} |\Omega_i|^2 - N + 1.$$

The geometric reason behind the bound is a bounded face dimension on $J(p, k)$. This inequality implies the support bound for $\omega \in \mathrm{supp}(p)$

$$\mathrm{supp}(k(\cdot|\omega)) \leq \sum_{i \in [N]} |\Omega_i|^2 - N + 1 \tag{9.10}$$

and the bound for conditional entropy

$$H(p, k) \leq \ln\Big(\sum_{i \in [N]} |\Omega_i|^2 - N + 1\Big). \tag{9.11}$$

To characterize maximizers of temporal interaction among stationary Markov transitions we study the Kirchhoff polytope. Its structure as a polytope is determined by the vertices which are cycles. Cyclic decompositions of more general (than stationary) Markov transitions are studied in [KaS]. We use concepts from graph theory.

Definition 9.8 A **one-dimensional complex** [Bam] consists of **nodes** and **branches**. Here, nodes are considered elements of Ω and branches elements in $\Omega \times \Omega$. We denote a complex by its set of branches $B \subset \Omega \times \Omega$. For any subset $B' \subset B$, the complex B' is called a **subcomplex** of B. A branch $(\omega, \omega') \in \Omega \times \Omega$ has **initial node** ω' and **terminal node** ω. The **node set** of the complex B is

$$N(B) := \{\omega \in \Omega : \omega \text{ is an initial or a terminal node of some branch } b \in B\}.$$

The complex B is **connected** if for every pair of distinct nodes $\omega, \omega' \in N(B)$ there exists $k \in \mathbb{N}$ and a sequence $\omega_0, \ldots, \omega_k$ of nodes in Ω with $\omega_0 = \omega'$ and $\omega_k = \omega$ such that one of the branches (ω_{i+1}, ω_i) or (ω_i, ω_{i+1}) belongs to B for $i = 0, \ldots, k-1$. A **connected component** of B is a maximal connected subcomplex of B (partially ordered by inclusion). The set of non-empty connected components of B is denoted $\mathrm{conn}(B)$. The **cyclomatic number** of B is

$$\mu(B) := |B| - |N(B)| + |\,\mathrm{conn}(B)|.$$

The **cycle space** of a complex $B \subset \Omega \times \Omega$ is

$$C(B) := \{x \in C(\Omega) : \mathrm{supp}(x) \subset B\}.$$

An element of $C(B)$ is called a **cycle dominated by** B. We represent B graphically. For each node in $N(B)$ we draw a dot in the plane. For each branch $b \in B$ with initial node ω' and terminal node ω we draw an arrow from the dot of ω' to the dot of ω. The complex $\mathrm{supp}(x) \subset \Omega \times \Omega$ for a point $x \in \mathrm{Kirch}(\Omega)$ is a **Kirchhoff complex** on Ω.

Remark 9.9 **(a)** The cycle space $C(\Omega)$ appears in the context of Kirchhoff's current law for electrical networks ([Bam], p. 420). A positive cycle (in the sense of a matrix algebra) is one with non-negative real coefficients.

(b) The dimension of the cycle space of a complex $B \subset (\Omega \times \Omega)$ is the cyclomatic number (see page 425 in [Bam])

$$\dim(C(B)) = \mu(B). \tag{9.12}$$

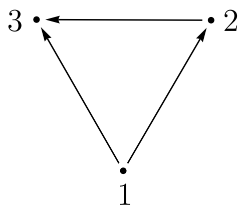


Figure 9.1: Graphical representation of the complex $\{(2,1),(3,2),(3,1)\}$.

The intersection of the cycle space with the positive cone can have a smaller dimension than the cycle space. For example, the complex $\{(2,1),(3,2),(3,1)\}$ (Figure 9.1) has a one-dimensional cycles space generated by the vector $x \in \mathbb{C}^{\{1,2,3\}\times\{1,2,3\}}$ which is zero except for $x(2,1) = x(3,2) = 1$ and $x(3,1) = -1$. We will see for a Kirchhoff complex there is no dimension difference.

Definition 9.10 Given a non-empty subset $U \subset \Omega$ and a cyclic permutation $\pi : U \to U$ we define an **elementary cycle** on Ω for $\omega, \omega' \in \Omega$ by

$$c(U,\pi)(\omega,\omega') := \begin{cases} 1 & \text{if } \omega' \in U \text{ and } \omega = \pi(\omega'), \\ 0 & \text{otherwise,} \end{cases}$$

and an **elementary probability cycle** on Ω by $\frac{c(U,\pi)}{|U|}$. The set of elementary probability cycles on Ω is denoted by $Z(\Omega)$.

Lemma 9.11 The Kirchhoff polytope $\mathrm{Kirch}(\Omega)$ is a polytope and $Z(\Omega) \subset \mathrm{Kirch}(\Omega)$. For $x, y \in \mathrm{Kirch}(\Omega)$ we have $\mathrm{supp}(x) \subset \mathrm{supp}(y)$ if and only if $x \in F(\mathrm{Kirch}(\Omega), y)$.
[Proof on page 250]

Lemma 9.12 Let $x \in C(\Omega)$ be a cycle. If there is an elementary cycle c on Ω such that $\mathrm{supp}(x) \subset \mathrm{supp}(c)$ then $x = \lambda c$ for some $\lambda \in \mathbb{C}$. If $x \geq 0$ and $x \neq 0$ then there exists an elementary cycle c on Ω such that $\mathrm{supp}(c) \subset \mathrm{supp}(x)$. [Proof on page 251]

Lemma 9.13 If $x \in \mathrm{Kirch}(\Omega)$ then the extreme points of the face $F(\mathrm{Kirch}(\Omega), x)$ are the elementary probability cycles $\{\zeta \in Z(\Omega) : \mathrm{supp}(\zeta) \subset \mathrm{supp}(x)\}$. The cycle space of the Kirchhoff complex $\mathrm{supp}(x)$ is $C(\mathrm{supp}(x)) = \mathrm{Lin}_{\mathbb{C}}(\{\zeta \in Z(\Omega) : \mathrm{supp}(\zeta) \subset \mathrm{supp}(x)\})$.
[Proof on page 251]

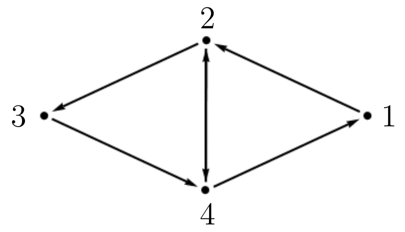


Figure 9.2: The Kirchhoff complex $\{(2,1),(3,2),(4,3),(1,4),(4,2),(2,4)\}$ in graphical representation.

Proposition 9.14 The face of $x \in \mathrm{Kirch}(\Omega)$ has dimension $\dim\left(F(\mathrm{Kirch}(\Omega), x)\right) = \mu(\mathrm{supp}(x)) - 1$ with cyclomatic number μ of the Kirchhoff complex $\mathrm{supp}(x)$. The face is

$$\begin{aligned} F(\mathrm{Kirch}(\Omega), x) &= \mathrm{conv}\left(\{\zeta \in Z(\Omega) : \mathrm{supp}(\zeta) \subset \mathrm{supp}(x)\}\right) \\ &= P(\Omega \times \Omega) \cap C(\mathrm{supp}(x)). \end{aligned}$$

In particular, if $\Omega \neq \emptyset$ then $\dim(\mathrm{Kirch}(\Omega)) = |\Omega|^2 - |\Omega|$. [Proof on page 252]

Example 9.15 For $\Omega := \{1,2,3,4\}$ we consider the Kirchhoff complex $\mathrm{supp}(p)$ for the joint probability distribution

$$p := \frac{1}{6}\begin{pmatrix} 0&0&0&1 \\ 1&0&0&1 \\ 0&1&0&0 \\ 0&1&1&0 \end{pmatrix}.$$

The complex has 6 branches, 4 nodes and one connected component (Figure 9.2). The face $F := F(\mathrm{Kirch}(\Omega), p)$ is depicted in Figure 9.3. By Proposition 9.14 the dimension of F is $\mu(\mathrm{supp}(p)) - 1 = 6 - 4 + 1 - 1 = 2$ and $\mathrm{Kirch}(\Omega)$ has dimension $4^2 - 4 = 12$. By Lemma 9.13 the four extreme points of F are

$$\frac{1}{4}\begin{pmatrix} 0&0&0&1 \\ 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \end{pmatrix}, \quad \frac{1}{3}\begin{pmatrix} 0&0&0&1 \\ 1&0&0&0 \\ 0&0&0&0 \\ 0&1&0&0 \end{pmatrix}, \quad \frac{1}{3}\begin{pmatrix} 0&0&0&0 \\ 0&0&0&1 \\ 0&1&0&0 \\ 0&0&1&0 \end{pmatrix} \quad \text{and} \quad \frac{1}{2}\begin{pmatrix} 0&0&0&0 \\ 0&0&0&1 \\ 0&0&0&0 \\ 0&1&0&0 \end{pmatrix}.$$

Through the extreme points of $\mathrm{Kirch}(\Omega)$ we can characterize local maximizers of temporal interaction on the Kirchhoff polytope.

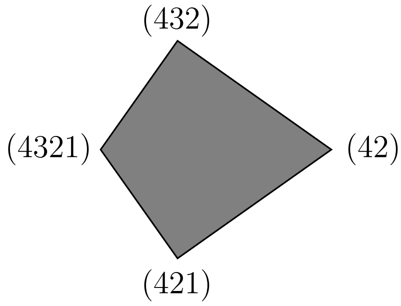


Figure 9.3: The kite in the figure is a homothetic drawing of the face F of the Kirchhoff polytope Kirch($\{1,2,3,4\}$) defined in Example 9.15. The extreme points of F are labeled by the permutation (in cycle notation) of the corresponding combinatorial cycles.

Remark 9.16 (Local maxima on $T(\Omega)$.) **(a)** The **theorem of Carathéodory** [Ro] says if $x \in \mathbb{R}^d$ belongs to the convex hull of a set $C \subset \mathbb{R}^d$ then there exists a subset $C' \subset C$ consisting of $d+1$ or fewer points such that x lies in the convex hull of C'.

(b) Optimization of temporal interaction among the stationary Markov transitions allows the same characterization of local maximizers as the unrestricted optimization in Remark 9.7. If (p,k) is a local maximizer of temporal interaction on the set $T(\Omega)$ of stationary Markov transitions then $Q := J(p,k)$ is a local maximizer of entropy distance from $\mathcal{F}([N])$ on the Kirchhoff polytope Kirch(Ω) by transition to joint probability distributions (Lemma 9.5 and Definition 9.6). The proof of optimization of entropy distance in Proposition 7.29 can be adapted. If $Q' \in Q + \chi([N])^\perp$ then by Theorem 6 the entropy distance from $\mathcal{F}([N])$ is

$$S_{\mathcal{F}([N])}(Q') = S(Q', N^{\mathrm{cmb}}(Q)) = -S(Q') - \mathrm{tr}(Q' \ln(N^{\mathrm{cmb}}(Q))).$$

The distribution Q is a local maximizer of multi-information on $\mathrm{Kirch}(\Omega) \cap (Q + \chi([N])^\perp)$ and von Neumann entropy S is strictly concave on this set. Thus

$$F(\mathrm{Kirch}(\Omega), Q) \cap (Q + \chi([N])^\perp) = \{Q\}$$

and this implies

$$\dim(F(\mathrm{Kirch}(\Omega), Q)) \leq \dim(\chi([N])) = \sum_{i \in [N]} |\Omega_i|^2 - N. \tag{9.13}$$

See Definition 9.6 for the dimension of $\chi([N])$. By Minkowski theorem, the Kirchhoff polytope is the convex hull of its extreme points and by Carathéodory theorem we obtain that Q is a convex combination of at most $\sum_{i\in[N]} |\Omega_i|^2 - N + 1$ extreme points. This gives for $\omega \in \mathrm{supp}(p) \subset \Omega$ a support bound $\mathrm{supp}(k(\cdot|\omega)) \leq \sum_{i\in[N]} |\Omega_i|^2 - N + 1$ and a conditional entropy bound $H(p,k) \leq \ln(\sum_{i\in[N]} |\Omega_i|^2 - N + 1)$ follows.

We can improve the bound (9.13) slightly having a closer look at the intersection of the factorizable family $\mathcal{F}([N])$ with the Kirchhoff polytope $\mathrm{Kirch}(\Omega)$.

Proposition 9.17 For $N \geq 0$ and $i \in [N]$ let Ω_i be finite sets and put $\Omega := \times_{i\in[N]} \Omega_i$ and $A_i := \mathbb{C}^{\Omega_i} \otimes \mathbb{C}^{\Omega_i}$ for $i \in [N]$. Then

$$\mathcal{F}([N]) \cap \mathrm{Kirch}(\Omega) = \{Q_{[N]} \ : \ Q_i \in \mathrm{ri}(\mathrm{Kirch}(\Omega_i)), \ i \in [N]\},$$

$$\overline{\mathcal{F}([N])} \cap \mathrm{Kirch}(\Omega) = \{Q_{[N]} \ : \ Q_i \in \mathrm{Kirch}(\Omega_i), \ i \in [N]\}.$$

The combinatorial normal projection $N^{\mathrm{cmb}}(Q) \in \overline{\mathcal{F}([N])}$ (7.14) of a distribution $Q \in \mathrm{Kirch}(\Omega)$ belongs to $\overline{\mathcal{F}([N])} \cap \mathrm{Kirch}(\Omega)$. [Proof on page 253]

Remark 9.18 (Local maxima on $T(\Omega)$—improved.) We can improve the bound (9.13). Let (p, k) be a local maximizer of temporal interaction on the set $T(\Omega)$ of stationary Markov transitions. Then $Q := J(p, k)$ is a local maximizer of entropy distance from the factorizable family $\mathcal{F}([N])$ on the Kirchhoff polytope $\mathrm{Kirch}(\Omega)$. The combinatorial normal projection N^{cmb} maps $\mathrm{Kirch}(\Omega)$ into itself by Proposition 9.17. The combinatorial normal projection preserves fibers of the projection to $\chi([N])$ by definition (7.21). This proves that

$$L := \pi_{\chi([N])}(\mathrm{Kirch}(\Omega)) = \pi_{\chi([N])}(\mathrm{Kirch}(\Omega) \cap \overline{\mathcal{F}([N])}).$$

Since the mean value chart (6.28) is a diffeomorphism $\pi = \pi_{\chi([N])}|_{\mathcal{F}([N])} : \mathcal{F}([N]) \to \mathrm{sr}_{\chi([N])}$, the polytope L has the same dimension as $\mathrm{Kirch}(\Omega) \cap \overline{\mathcal{F}([N])}$, which is by Proposition 9.14

$$\dim(L) = \sum_{i\in[N]} (|\Omega_i|^2 - |\Omega_i|). \tag{9.14}$$

In the same way as in Remark 9.16 (b), the strict concavity of von Neumann entropy implies that

$$F(\mathrm{Kirch}(\Omega), Q) \cap (Q + \chi([N])^{\perp}) = \{Q\}.$$

This proves $\mathrm{lin}(F(\mathrm{Kirch}(\Omega), Q)) \cap (\chi([N])^{\perp} \cap \mathrm{lin}(\mathrm{Kirch}(\Omega))) = \{0\}$ and a dimension estimate follows,

$$\begin{aligned}
&\dim(F(\mathrm{Kirch}(\Omega), Q)) + \dim(\chi([N])^{\perp} \cap \mathrm{lin}(\mathrm{Kirch}(\Omega))) \\
\leq \ & \dim(\mathrm{Kirch}(\Omega)) \\
= \ & \dim(L) + \dim(\chi([N])^{\perp} \cap \mathrm{lin}(\mathrm{Kirch}(\Omega))).
\end{aligned}$$

With the equality (9.14) this implies the support estimate for $\omega \in \mathrm{supp}(p)$

$$\mathrm{supp}(k(\cdot|\omega)) \leq \sum_{i\in[N]} (|\Omega_i|^2 - |\Omega_i|) + 1 \tag{9.15}$$

and the bound for conditional entropy

$$H(p,k) \leq \ln\left(\sum_{i\in[N]} (|\Omega_i|^2 - |\Omega_i|) + 1\right). \tag{9.16}$$

Remark 9.19 If $x \in \mathrm{Kirch}(\Omega)$ is a local maximizer of multi-information then we can control in Remark 9.18 the face dimension d of x. With Carathéodory's Theorem we deduce a minimum number of elementary probability cycles necessary for a representation of x. Here we want to point out that d has no a compelling influence on the support of x or on the number of extreme points of its face.

(a) There is no bound on the support size $|\operatorname{supp}(x)|$ in dependence of d. For an extreme point, the dimension is constant $d = 0$. The extreme points of the Kirchhoff polytope are the elementary probability cycles (Lemma 9.13) that can have support size $|\Omega|$.

(b) There is no polynomial bound in d for the number of extreme points of the face $F(\mathrm{Kirch}(\Omega), x)$. We give an example where the number of extreme points grows exponentially with d.

For $n \in \mathbb{N}$ we consider the complex with nodes $\{0, 1, \ldots, n+1\}$ and with branches $\{(0,1), \ldots, (0,n)\} \cup \{(1, n+1), \ldots, (n, n+1)\}$. For $m \in \mathbb{N}$ we glue m copies $C_1, \ldots, C_m$ of the described complex and create a complex C in the following way. We identify the vertex $n+1$ of the complex C_i with the vertex 0 of the complex C_{i+1} for $i = 1, \ldots, m-1$ and we identify the vertex $n+1$ of the complex C_m with the vertex 0 of the complex C_1.

The connected complex C has $|N(C)| = m(n+1)$ nodes and $|C| = 2mn$ branches. The cyclomatic numbers is

$$\mu(C) = |C| - |N(C)| + |\operatorname{conn}(C)| = 2mn - m(n+1) + 1 = m(n-1) + 1.$$

We can model the complex C as a Kirchhoff complex for a configuration space Ω of cardinality at least $|N(C)| = m(n+1)$. Using identical distribution on $C \subset \Omega \times \Omega$, for $\omega, \omega' \in \Omega$

$$x(\omega, \omega') := \frac{1}{2mn} \begin{cases} 1 & \text{for } (\omega, \omega') \in C, \\ 0 & \text{otherwise,} \end{cases}$$

we have $C = \operatorname{supp}(x)$. By Proposition 9.14 the dimension of the face $F(\mathrm{Kirch}(\Omega), x)$ is

$$d = \mu(\operatorname{supp}(x)) - 1 = m(n-1).$$

The number of extreme points of this face is the number of elementary cycles dominated by $\operatorname{supp}(x)$,

$$|\{\zeta \in Z(\Omega) : \operatorname{supp}(\zeta) \subset \operatorname{supp}(x)\}| = n^m.$$

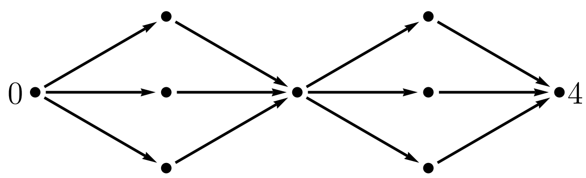


Figure 9.4: The depicted complex with node 0 and node 4 identified is the complex C in Remark 9.19 (b) for $n = 3$ and $m = 2$. For fixed $n = 3$ a cyclic concatenation C of m copies of a diamond shaped subcomplex has the linear cyclomatic number $\mu(C) = 2m+1$ and the exponential number 3^m of elementary cycles dominated by C.

For fixed n and variable $m = \frac{d}{n-1}$ this number is exponential in the dimension d

$$n^m = (\sqrt[n-1]{n})^d.$$

For $n = 2$ we get the largest base $\sqrt[1]{2} = 2$ for exponentiation, the face dimension is $d = m = \mu(C) - 1$ and the number of extreme points is $2^d = 2^{\mu(C)-1}$. We may compare to the largest possible number $2^{\mu(C)} - 1$ of cycles in an undirected graph [Vo]. The complex C for $m = 2$ and $n = 3$ is represented graphically in Figure 9.4.

10 Evidence

Proof of Remark 2.13. (b) We can write $\mathrm{tr}(ab) = \mathrm{tr}((\sqrt{a}\sqrt{b})^*(\sqrt{a}\sqrt{b}))$ using the square root in Remark 2.11 (e) and cyclic invariance under the trace. So $a \perp b$ implies $\|\sqrt{a}\sqrt{b}\|_2 = 0$ for the HS norm. By (a) the HS norm is a norm in the sense of Definition 2.1 (b) hence $\sqrt{a}\sqrt{b} = 0$ and this implies $ab = 0$. The converse is trivial. `qed`

Proof of Remark 2.16. (b) Observe $\mathbb{1} - p - q \geq 0$ so that Remark 2.11 (c) shows $-pqp = p(\mathbb{1} - p - q)p \geq 0$. Since $pqp \geq 0$ we get $pqp = 0$ by Remark 2.11 (d). In other words, the reason is that $\leq$ is a partial ordering on A_{sa}, see Remark 2.24. Then we obtain that p and q are perpendicular, $\mathrm{tr}(pq) = \mathrm{tr}(pqp) = 0$. Since orthogonal projectors are positive by (a) we can use Remark 2.13 (b) and obtain $pq = 0$ from orthogonality.

(d) This follows from the definition of functional calculus in Remark 2.9 (d) through interpolation of a function $f : \mathrm{spec}(a) \to \mathbb{C}$ by Newton polynomials. `qed`

Proof of Remark 2.25. (b) The equivalence between (iv) and (v) follows from Remark 2.13 (b), the non-trivial part is the positive definiteness of HS inner product. On assumption of (v) we find $q - p = q - qpq = q(\mathbb{1} - p)q \geq 0$, which is (iii). Assuming (iii), that is $p - q \leq 0$ we find

$$\mathrm{tr}(p(\mathbb{1} - q)) = \mathrm{tr}(p(p - q)p) = \mathrm{tr}(p(\mathbb{1} - q)p) = 0,$$

which is (iv). We have to link the first two conditions to the other conditions. Assuming (v) we find (i), $\mathrm{Im}(p) = \mathrm{Im}(qp) = q(\mathrm{Im}(p)) \subset \mathrm{Im}(q)$. The assumption that $\mathrm{Im}(p) \subset \mathrm{Im}(q)$ trivially implies $\mathrm{Im}(p) \perp \mathrm{Im}(q)^\perp$. The converse direction comes from (2.37): $\mathrm{Im}(p) \subset (\mathrm{Im}(q)^\perp)^\perp = \mathrm{Im}(q)$. Finally let us prove that $\mathrm{Im}(p) \subset \mathrm{Im}(q)$ implies $p \leq q$. For a vector $x \in \mathcal{H}$ we use the decomposition $x = y + z$ where $y \in \mathrm{Im}(p) \subset \mathrm{Im}(q)$ and $z \in \ker(p)$. Then

$$\langle x, (q - p)(x) \rangle = \langle z, q(z) \rangle \geq 0$$

gives $q - p \geq 0$, see Remark 2.11 (d). `qed`

Proof of Remark 2.35 (e). For $x, y \in \mathcal{L}$ the relation $x \leq y$ is equivalent to $x \wedge y = x$ and to $x \vee y = y$ by (2.42). We assume (2.43). If $x \leq y$ then $x = x \wedge y$ then $x' = (x \wedge y)' = x' \vee y'$ then $x' \geq y'$.

Secondly, the equivalence (2.44) expresses that the involution $x \mapsto x'$ is an antitone bijection with antitone inverse. Now Lemma 2 on page 24 in [Bi] shows that $x \mapsto x'$ is a lattice isomorphism (changing the ordering from $\leq$ to $\geq$ and interchanging $\wedge$ and $\vee$). qed

Proof of Remark 2.36 (d). Notice that $s(a)a = a$ by the spectral theorem (2.17) and the definition of the support projector (2.20). On the other hand, if an orthogonal projector $p \in \mathcal{P}(A)$ satisfies $pa = a$, then $\mathrm{Im}(p) \supset \mathrm{Im}(a)$. This is $\mathrm{Im}(p) \supset \mathrm{Im}(s(a))$ because a and the support projector $s(a)$ have the same image (2.24). Application of the inverse of the lattice isomorphism $p \mapsto \mathrm{Im}(p)$ gives $p \geq s(a)$. qed

Proof of Lemma 2.38. At first we calculate the diameter for the projector lattice. For two projectors $p, q \in \mathcal{P}(A)$ and a unit vector u in the Hilbert space H

$$\|(p-q)(u)\|_2^2 = \|u\|_2^2 - \|(\mathbb{1} - p - q)(u)\|_2^2 \leq 1$$

holds because $(p-q)^2 + (\mathbb{1}-p-q)^2 = \mathbb{1}$, see [Av]. Thus $\|p-q\| = \sup_{\substack{u \in H \\ \|u\|_2 = 1}} \|(p-q)(u)\|_2 \leq 1$.

For a non-zero projector $p \in \mathcal{P}(A)$ we have $\|p\| = 1$. If $A \neq \{0\}$ then

$$1 \geq \sup_{q_1, q_2 \in \mathcal{P}(A)} \|q_1 - q_2\| \geq \|p\| = 1.$$

In particular, the projector lattice is bounded. For compactness we show that $\mathcal{P}(A)$ is closed. First of all, the space of self-adjoint matrices is closed. Secondly for matrices $a, p \in A$ with $p^2 = p$ we have $a^2 - p = (a-p)^2 + (a-p)p + p(a-p)$, so

$$\|a^2 - p\| \leq \|a - p\|(\|a - p\| + 2\|p\|).$$

This proves that the space of matrices $\{p \in A : p^2 = p\}$ is closed.

For the remaining questions, we first consider the case $A = M_n$ of the full matrix algebra of complex $n \times n$ matrices for $n \in \mathbb{N}$. If $0 \leq k \leq n$ then compactness of $\mathcal{P}_k(M_n)$ follows from closedness because the projector lattice $\mathcal{P}(M_n)$ is compact. The closedness of $\mathcal{P}_k(M_n)$ follows from the conditional for projectors $p, q \in \mathcal{P}(M_n)$

$$\|p - q\| < 1 \implies \mathrm{rk}(p) = \mathrm{rk}(q) \tag{10.1}$$

which is proved in [Av]. The conditional (10.1) is equivalent to the maximal separation of two conjugation manifolds for $k \neq l$ with $0 \leq k \leq n$ and $0 \leq l \leq n$,

$$\inf_{p \in \mathcal{P}_k(M_n), q \in \mathcal{P}_l(M_n)} \|p - q\| = 1.$$

The differentiable structure of $\mathcal{P}(M_n)$ is deduced from the Grassmannian $\mathcal{G}(\mathbb{C}^n)$ through the isomorphism (2.38)

$$\mathcal{P}(M_n) \to \mathcal{G}(M_n), \quad p \mapsto \operatorname{Im}(p).$$

Indeed, the projectors $p \in \mathcal{P}(M_n)$ of constant rank k correspond to the k-dimensional subspaces of $\mathbb{C}^n$. These form a complex differentiable manifold of dimension $k(n-k)$ [Hi]. Likewise this is a real differentiable manifold of dimension $2k(n-k)$. The charts are described by Plücker coordinates and chart changes are analytic (they are rational functions) [Wey, Har].

Let $A = M_{n_1} \oplus \cdots \oplus M_{n_N}$ have the direct sum representation (2.6) for a multi-index $n = (n_1, \ldots, n_N) \in \mathbb{N}^N$. Let $p, q \in \mathcal{P}(A)$ be orthogonal projectors. The projector q is conjugate to p if and only if q_i is conjugate to p_i for $i = 1, \ldots, N$. Thus the conjugation classes of $\mathcal{P}(A)$ are the conjugation manifolds

$$\mathcal{P}_k(A) = \mathcal{P}_{k_1}(M_{n_1}) \oplus \cdots \oplus \mathcal{P}_{k_N}(M_{n_N})$$

for multi-indices $0 \leq k \leq n$. The topological and differentiable properties translate from the above full matrix algebra case. The dimension is $2\sum_{i=1}^N k_i(n_i - k_i)$. The C*-norm distance is $\|p - q\| = \max_{i \in \{1,\ldots,N\}} \|p_i - q_i\|$ for $p, q \in \mathcal{P}(A)$. Thus for multi-indices $k \neq l$

$$\inf_{p \in \mathcal{P}_k(A), q \in \mathcal{P}_l(A)} \|p - q\| = 1. \qquad \texttt{qed}$$

Proof of (2.50). By convergence of $a_n := \mathbb{1} - p - p_n$ to $\mathbb{1} - 2p$ we can use perturbation theory [KaT]. One has $\operatorname{spec}(\mathbb{1} - 2p) \subset \{-1, 1\}$ and for $n \in \mathbb{N}$ the total projectors of a_n are

$$q_n(\mu) := \frac{1}{-2\pi i} \int_{\gamma_\mu} (a_n - \zeta)^{-1} \, \mathrm{d}\, \zeta$$

where $\mu \in \{-1, 1\}$ and γ_μ is a positively oriented curve in $\mathbb{C} \setminus \{-1, 1\}$ enclosing μ but not $-\mu$. By analyticity of eigenvalues we have for $n \in \mathbb{N}$ large enough $\operatorname{sgn}(a_n) = q_n(1) - q_n(-1) \overset{n\to\infty}{\longrightarrow} (\mathbb{1} - p) - p = \mathbb{1} - 2p$. qed

Proof of (3.18). Notice that for $c_1 \in \operatorname{ri}(C_1)$ we have $\operatorname{aff}(C_1) = \{c_1 + \lambda(\widetilde{c_1} - c_1) : \widetilde{c_1} \in C_1, \lambda \in \mathbb{R}\}$. By (3.17) we can choose c_1 in the relative interior of each of the three convex sets C_1, C_2 and $C_1 \cap C_2$ and obtain the relation. qed

Proof of Lemma 3.5. We use Theorem 7.1 in [Ro]. This states that an extended real-valued function defined on $\mathbb{R}^m$ is lower semi-continuous throughout $\mathbb{R}^m$ if and only if the level sets of the function are closed for all real values ($\pm\infty$ excluded).

Let us consider the extension of f

$$f^{\mathrm{ex}}(x) := \begin{cases} f(x) & \text{if } x \in C, \\ +\infty & \text{otherwise} \end{cases}$$

defined on $\mathbb{R}^m$. Since C is closed and f^{ex} is constant on the complement $\mathbb{R}^m \setminus C$, the functions f and f^{ex} are either both lower semi-continuous or they are both not lower semi-continuous. On the other hand, the level sets of f and f^{ex} are all equal except for the value $+\infty$. `qed`

Proof of Remark 3.9. (a) The inclusion $F \subset \mathrm{aff}(F) \cap C$ is clear. Conversely, if $x \in \mathrm{aff}(F)$ then there are $y, z \in F$ and $\lambda \in \mathbb{R}$ such that $x = y + \lambda(z - y)$. If $\lambda \in [0, 1]$ then $x \in [y, z] \subset F$. Otherwise, if $\lambda < 0$ then $y \in]x, z[$ and this implies $x \in F$ by the face property of F. Likewise $\lambda > 1$ gives $z \in]y, x[$ with the same result.

(d) The previous property (c) says that being a face of C is a closure property in the sense of Definition 2.30. Thus by Lemma 2.31 the lattice is complete with intersection as infimum.

(e) is proved in Theorem 18.1 in [Ro].

(f) is proved in Theorem 18.2 in [Ro]. `qed`

Proof of Remark 3.11. (a) follows from Remark 3.9 (e).

(b) Since $F(C_1, x) \cap C_2$ is a face of $C_1 \cap C_2$ we must show that x belongs to the relative interior $\mathrm{ri}(F(C_1, x) \cap C_2)$. Since $x \in \mathrm{ri}(F(C_1, x))$ and C_2 is affine, this follows from (3.17).

(c) One has

$$F(G, x) = F(C \cap \mathrm{aff}(G), x) = F(C, x) \cap \mathrm{aff}(G) = F(C, x)$$

where the first equality is Remark 3.9 (a), then (b) and (a) of the current remark are used.

(d) is proved in Corollary 18.5.1 in [Ro]. `qed`

Proof of Lemma 3.16. The desired details were not found in the literature, hence a short proof at this point. Let $F := \bigcap_{u \in U} F_\perp(C, u)$ and $G := \bigcap_{u \in \mathrm{conv}(U) \setminus \{0\}} F_\perp(C, u)$.

First we show $F = G$. The non-trivial part is to prove $F \subset G$. A vector $v \in \mathrm{conv}(U) \setminus \{0\}$ is a convex combination $v = \sum_i \lambda_i u_i$ for $u_i \in U$ and non-negative real scalars λ_i. If $x \in F$ then $x \in F_\perp(C, u_i)$ for all i and then

$$\langle v, x\rangle = \sum_i \lambda_i \langle u_i, x\rangle = \sum_i \lambda_i \max_{s \in C} \langle u_i, s\rangle \geq \max_{s \in C} \sum_i \lambda_i \langle u_i, s\rangle = \max_{s \in C} \langle v, s\rangle,$$

so $x \in F_\perp(C, v)$. The vector v was arbitrary. So $x \in G$ and we have $F = G$ indeed.

We assume that $G \neq \emptyset$ and prove $G = F_\perp(C, v)$ for $v \in \mathrm{ri}(\mathrm{conv}(U)) \setminus \{0\}$. Observe that $\mathrm{ri}(\mathrm{conv}(U)) \setminus \{0\} \neq \emptyset$, otherwise $U = \{0\}$ or $U = \emptyset$ which excluded in the assumptions. To prove the non-trivial inclusion $F_\perp(C, v) \subset G$ assume by contradiction that there is a point $y \in F_\perp(C, v) \setminus G$. Then for some vector $u_0 \in \mathrm{conv}(U) \setminus \{0\}$ we have

$$y \in F_\perp(C, v) \setminus F_\perp(C, u_0).$$

Since v lies in the relative interior of $\mathrm{conv}(U)$ and u_0 lies in $\mathrm{conv}(U)$ there exists $\lambda \in (0, 1)$ and $u_1 \in \mathrm{conv}(U)$ such that $v = \lambda u_0 + (1-\lambda) u_1$ (see Theorem 6.4 in [Ro]). We can assume that $u_1 \neq 0$ by performing a small perturbation of this point along the direction $v - u_0$ if necessary. Let $x \in G$. Then $x \in F_\perp(C, u_0) \cap F_\perp(C, u_1)$. The estimate

$$\begin{aligned} \langle v, y\rangle &= \lambda \langle u_0, y\rangle + (1-\lambda)\langle u_1, y\rangle < \lambda \max_{z \in C} \langle u_0, z\rangle + (1-\lambda)\langle u_1, y\rangle \\ &\leq \lambda \langle u_0, x\rangle + (1-\lambda)\langle u_1, x\rangle = \langle v, x\rangle \end{aligned}$$

gives the contradiction $y \notin F_\perp(C, v)$. `qed`

Proof of Remark 3.19. The equation (3.41) is proved for closed and bounded convex sets in Theorem 2.2.1 in [Sch]. Since the normal cone of a convex set is a local property, the equation is true for closed convex sets. The equation is true if C_1, C_2 are only convex: one has $\mathrm{N}(C_i, x) = \mathrm{N}(\overline{C_i}, x)$ $(i = 1, 2)$ and $\mathrm{N}(C_1 \cap C_2, x) = \mathrm{N}(\overline{C_1 \cap C_2}, x)$. Since C_1, C_2 share a relative interior point, we have $\overline{C_1 \cap C_2} = \overline{C_1} \cap \overline{C_2}$ by Theorem 6.5 in [Ro]. `qed`

Proof of Lemma 3.20. The first inclusion $\mathrm{lin}(C)^\perp \subset \mathrm{N}(C, x)$ for $x \in C$ is easy to see: if $u \in \mathrm{lin}(C)^\perp$ then $\langle u, y - x\rangle = 0$ for all $y \in C$ so $u \in \mathrm{N}(C, x)$. Now for some $x \in C$ let us assume that $\mathrm{N}(C, x)$ is a vector space. Then for $u \in \mathrm{N}(C, x)$ we have $\pm u \in \mathrm{N}(C, x)$ and for $y \in C$ we get

$$h(C, u) = \langle u, y - x\rangle = 0 = -\langle -u, y - x\rangle = -h(C, -u).$$

Thus, for the vectors $u \in \mathbb{R}^n$ with $h(C,u) \neq -h(C,u)$ follows $u \notin \mathrm{N}(C,x)$. This means by the maximization characterization of normal cones (3.38) that $\langle u, x\rangle < h(C,x)$. These are exactly the assumption of Theorem 13.1 in [Ro] to prove that $x \in \mathrm{ri}(C)$. Clearly, if $x \in \mathrm{ri}(C)$ then $\mathrm{N}(C,x) = \mathrm{lin}(C)^{\perp}$.

The inclusion $\mathrm{N}(C,x) \subset B(C)$ for each $x \in C$ was remarked upon in (3.37). Let us assume that $\mathrm{N}(C,x) = B(C)$ and define $D := \overline{C - x}$. Translations and taking closures of a convex set do not change normal cones nor barrier cones. So

$$\mathrm{N}(D,0) = \mathrm{N}(C,x) = B(C) = B(D).$$

For a point $y \in D$ we find $\langle u, y\rangle \leq h(D,u)$ for all $u \in \mathbb{R}^m$ by definition of the support function. Since $B(D) = \mathrm{N}(D,0)$ we find for all $u \in B(D)$ the equality $h(D,u) = \langle u, 0\rangle = 0$ by the duality (3.39). Hence for all $y \in D$ for all $u \in B(D)$ and for all $\lambda \geq 0$

$$\langle u, \lambda y\rangle = \lambda\langle u, y\rangle \leq \lambda h(D,u) = 0 = h(D,u)$$

holds. A closed convex set is the intersection of half spaces that contain it, see Theorem 13.1 in [Ro], so $\lambda y \in D$. This shows that D is a convex cone. `qed`

Proof of (3.45). Both assertions are trivial if $F = \emptyset$ or if $F_{\perp}(C,u) = \emptyset$. Let $u \in B(C) \setminus \{0\}$. For a point $x \in \mathrm{ri}(F_{\perp}(C,u))$ we have by duality (3.39) and by Remark 3.19 (g) (iii)

$$u \in \mathrm{N}(C,x) = \mathrm{N}(F_{\perp}(C,u)).$$

From this equation the antitone assignment of normal cones (3.44) gives that $F \subset F_{\perp}(C,u)$ implies $u \in \mathrm{N}(C, F_{\perp}(C,u)) \subset \mathrm{N}(C,F)$. Conversely, if $u \in \mathrm{N}(C,F)$ then for $x \in \mathrm{ri}(F)$ we have $u \in \mathrm{N}(C,x)$. Thus $x \in F_{\perp}(C,u)$ and Remark 3.11 (a) gives $F \subset F_{\perp}(C,u)$. `qed`

Proof of Proposition 3.23. The proposition is trivial if $G = \emptyset$. Then $\mathrm{N}(C,G) = \mathbb{R}^m$ and $F \vee G = F$. Let $F := F(C,x)$ and $G := F(C,y)$ for points $x, y \in C$ throughout the proof.

Observe that for any $z \in]x,y[$ we have $F(C,z) = F \vee G$. The points x and y belong to $F(C,z)$ hence the whole faces F, G belong to $F(C,z)$, see Remark 3.11 (a). Thus $F \vee G \subset F(C,z)$. On the other hand z belongs to $[x,y] \subset F \vee G$ so we get $F(C,z) \subset F \vee G$.

Now we can prove the inclusions $\mathrm{N}(C,F) \cap \mathrm{N}(C,G) \subset \mathrm{N}(C, F \vee G)$. For $u \in \mathrm{N}(C,x) \cap \mathrm{N}(C,y)$ and arbitrary $\xi \in C$ we find

$$\langle u, \xi - \frac{1}{2}(x+y)\rangle = \frac{1}{2}\langle u, \xi - x\rangle + \frac{1}{2}\langle u, \xi - y\rangle \leq 0.$$

This proves $u \in \mathrm{N}(C, F \vee G)$. The inclusions $\mathrm{N}(C, F), \mathrm{N}(C, G) \supset \mathrm{N}(C, F \vee G)$ follow from the antitone assignment (3.44) of normal cones.

Let us prove that $\mathrm{N}(C, F) \cap \mathrm{N}(C, G)$ is a face of $\mathrm{N}(C, F)$. We must show for $u, v, w \in \mathrm{N}(C, F)$ and $v \in \mathrm{N}(C, F \vee G) \cap]u, w[$ that $u, w \in \mathrm{N}(C, G)$ holds. If $u = 0$ then $w = \lambda v$ for some real $\lambda > 0$. Then $u, w \in \mathrm{N}(C, G)$ because $\mathrm{N}(C, G)$ is a closed cone including v. If $u, w \neq 0$ and $v = 0$ then $u, w \in \mathrm{lin}(C)^{\perp}$. By Lemma 3.20 the vector space $\mathrm{lin}(C)^{\perp}$ belongs to the normal cone of every point of C so $u, w \in \mathrm{N}(C, G)$.

Finally, assume that $u, v, w \neq 0$. Since $v \in \mathrm{N}(C, G)$ we have $G \subset F_{\perp}(C, v)$ by (3.45). Now $F_{\perp}(C, v) = F_{\perp}(C, u) \cap F_{\perp}(C, w)$ holds by Lemma 3.16 so

$$G \subset F_{\perp}(C, v) = F_{\perp}(C, u) \cap F_{\perp}(C, w) \subset F_{\perp}(C, u)$$

gives $\mathrm{N}(C, F_{\perp}(C, u)) \subset \mathrm{N}(C, G)$ and (3.45) completes the proof with $u \in \mathrm{N}(C, F_{\perp}(C, u))$. The proof that $v \in \mathrm{N}(C, G)$ is a complete analogue. qed

Proof of Corollary 3.24. The normal cone lattice is complete by Remark 2.34 because it has finite length. If for two faces $F, G \in \mathcal{F}(C) \setminus \{\emptyset\}$ the normal cones are properly included in each other,

$$\mathrm{N}(C, F) \subsetneq \mathrm{N}(C, G),$$

then by Proposition 3.23 the cone $\mathrm{N}(C, F) = \mathrm{N}(C, F) \cap \mathrm{N}(C, G)$ is a proper face of $\mathrm{N}(C, G)$. The dimension of $\mathrm{N}(C, F)$ is strictly smaller than the dimension of the cone $\mathrm{N}(C, G)$ by the stratification property, see Remark 3.9 (e). The length of $\mathcal{N}(C)$ is bounded by $m + 1$.

For two cones $K, L \in \mathcal{N}(C)$ the intersection $K \cap L$ belongs to $\mathcal{N}(C)$ by Proposition 3.23 and the intersection is also the largest common subset of K and L, so $K \wedge L = K \cap L$. The smallest element of $\mathcal{N}(C)$ is calculated in Lemma 3.20. qed

Proof of Lemma 3.26. The intersection expression follows from the duality (3.45). Since $F \subset \overset{\perp}{F}$, the inclusion $\mathrm{N}(C, \overset{\perp}{F}) \subset \mathrm{N}(C, F)$ follows from antitone assignment of normal cones (3.44). The zero vector belongs to every normal cone. For a non-zero vector $u \in \mathrm{N}(C, F)$ we have by duality (3.45) $F \subset F_{\perp}(C, u)$ hence $\overset{\perp}{F} \subset F_{\perp}(C, u)$. Duality applied again gives $u \in \mathrm{N}(C, \overset{\perp}{F})$. qed

Proof of Lemma 3.27. Before proving the assertion let us show for $x \in \mathbb{R}^m$ and $\{x\} \subsetneq C$ the equality

$$\mathrm{ri}(\mathrm{conv}(C \setminus \{x\})) = \mathrm{ri}(C). \tag{10.2}$$

The set C will be substituted later for a normal cone. If $C \setminus \{x\}$ is not convex then $\operatorname{conv}(C \setminus \{x\}) = C$ and the equality follows. If $C \setminus \{x\}$ is convex then x is an extreme point of C, hence $\operatorname{ri}(C) \subset C \setminus \{0\} \subset C$, unless $C = \{x\}$. The set $C \setminus \{x\}$ is sandwiched between the relative interior and the closure of C, thus the relative interiors of the convex sets $C \setminus \{x\}$ and C are equal, see Corollary 6.3.1 in [Ro].

By assumption, the normal cone $\mathrm{N}(C, F)$ has more than one point. Then (3.47) shows $F = \bigcap_{u \in \mathrm{N}(C,F)\setminus\{0\}} F_\perp(C, u)$ and by Lemma 3.16 this intersection of exposed faces is equal to $F_\perp(C, v)$ for any vector $v \in \operatorname{ri}(\operatorname{conv}(\mathrm{N}(C, F)\setminus\{0\}))\setminus\{0\}$. The latter set is $\operatorname{ri}(\mathrm{N}(C, F))\setminus\{0\}$ by (10.2). qed

Proof of Proposition 3.28. (a) If $F \subset G$ then $\mathrm{N}(G) \subset \mathrm{N}(F)$ since the assignment of normal cones is antitone (3.44).

(b) By (3.47), the inclusion $\mathrm{N}(G) \subset \mathrm{N}(F)$ implies $\overset{\perp}{F} \subset \overset{\perp}{G}$. If $G = \overset{\perp}{G}$ then $F \subset \overset{\perp}{F} \subset \overset{\perp}{G} = G$. Conversely, if G is not exposed then $G \subsetneq \overset{\perp}{G}$ but the two faces G and $\overset{\perp}{G}$ have the same normal cones by (3.48).

(c) The inclusion $\mathrm{N}(G) \subset \mathrm{N}(F)$ follows from (a). If $F = \overline{F}$ and $\mathrm{N}(F) \subset \mathrm{N}(G)$ then $G \subset F$ follows by (b). Conversely, if F is not exposed then $F \subsetneq \overset{\perp}{F}$ and $\mathrm{N}(F) = \mathrm{N}(\overset{\perp}{F})$.

(d) The inclusion follows from (b). If $F = G$ then $\mathrm{N}(G) = \mathrm{N}(F)$. qed

Proof of Proposition 3.29. The two lattices $\mathcal{F}_\perp(C)$ and $\mathcal{N}(C)$ are partially ordered by set inclusion. They are linked by the antitone mapping

$$\mathrm{N}(C)|_{\mathcal{F}_\perp(C)} : \mathcal{F}_\perp(C) \to \mathcal{N}(C), \quad F \mapsto \mathrm{N}(C, F),$$

see (3.44). This mapping is surjective because a face F of C has the same normal cone as the smallest exposed face that contains F, see (3.48).

We can show that $\mathrm{N}(C)|_{\mathcal{F}_\perp(C)}$ has an antitone inverse. Then Lemma 2.32 implies that the mapping is an (antitone) lattice isomorphism. Let us prove that $\mathrm{N}(C)|_{\mathcal{F}_\perp(C)}$ is injective. Notice that $\mathrm{N}(C)|_{\mathcal{F}_\perp(C)\setminus\{\emptyset, C\}}$ is injective by Lemma 3.27. (To apply the lemma recall from Lemma 3.20 that the normal cone of a proper face has a non-zero vector.) Moreover, by the Lemma 3.20, only the improper face C has the smallest normal cone $\operatorname{lin}(C)^\perp$ and it remains to show that $\mathrm{N}(C, F) = \mathbb{R}^m$ implies $F = \emptyset$ for a face F of C. If $\mathrm{N}(C, F) = \mathbb{R}^m$ holds for a non-empty face F then Lemma 3.20 shows that $F = C$ and $\operatorname{lin}(C) = (\mathbb{R}^m)^\perp = \{0\}$. Thus, C has exactly one point. This case was excluded in the assumptions.

The antitone character of the inverse $\mathrm{N}(C)|_{\mathcal{F}_\perp(C)\setminus\{\emptyset,C\}}$ follows from Proposition 3.28 (b). With the relations $\mathrm{N}(C,\emptyset)=\mathbb{R}^m$ and $\mathrm{N}(C,C)=\mathrm{lin}(C)^\perp$ we conclude that $\mathrm{N}(C)|_{\mathcal{F}_\perp(C)}$ is an antitone bijection. qed

Proof of Lemma 3.31. The inclusion $(X+Y)\cap Z\subset X+(Y\cap Z)$ is proved by taking vectors $x\in X$ and $y\in Y$ such that $x+y\in Z$. Then $y=(x+y)-x\in Z$. For the converse $X+(Y\cap Z)\subset(X+Y)\cap Z$ we choose vectors $x\in X$ and $t\in Y\cap Z$. Then $t+x\in Z$. qed

Proof of Lemma 3.36. We prove that $L_V^C(F)$ is a face of C for F a face of $\pi_V(C)$. Let $y\in L_V^C(F)$. Then $\pi_V(y)\in F$. Let $x,z\in C$ with $y\in]x,z[$. A linear map commutes with reduction to the relative interior (3.15) so $\pi_V(y)\in]\pi_V(x),\pi_V(z)[$. Since F is a face, we get $\pi_V(x),\pi_V(z)\in F$. Then

$$x\in L_V^C(x)=L_V^C\circ\pi_V(x)=(\pi_V(x)+V^\perp)\cap C\subset(F+V^\perp)\cap C=L_V^C(F).$$

Also $z\in L_V^C(F)$, so $L_V^C(F)$ is a face of C.

The support functions of C and $\pi_V(C)$ are equal on V because for all $x\in\mathbb{R}^m$ and $v\in V$ we have $\langle v,x\rangle=\langle v,\pi_V(x)\rangle$. If $v\in V\setminus B(C)$ then $F_\perp(C,v)=F_\perp(\pi_V(C),v)=\emptyset$ and accordingly $L_V^C(\emptyset)=\emptyset$. If v is a non-zero vector in $B(C)$ then the supporting hyperplanes $H(C,v)$ and $H(\pi_V(C),v)$ are equal, see (3.31). By the modular law for affine spaces applied to $V^\perp\subset\mathrm{lin}(H(\pi_V(C),v))$ we get

$$\begin{aligned}F_\perp(\pi_V(C),v)+V^\perp &= (H(\pi_V(C),v)\cap\pi_V(C))+V^\perp=H(\pi_V(C),v)\cap(\pi_V(C)+V^\perp)\\ &= H(C,v)\cap(C+V^\perp).\end{aligned}$$

This gives

$$\begin{aligned}L_V^C(F_\perp(\pi_V(C),v)) &= (F_\perp(\pi_V(C),v)+V^\perp)\cap C=H(C,v)\cap(C+V^\perp)\cap C\\ &= H(C,v)\cap C=F_\perp(C,v)\end{aligned}$$

finally. qed

Proof of Proposition 3.38. The mapping L_V^C restricted to $\mathcal{F}(\pi_V(C))$ resp. to $\mathcal{F}_\perp(\pi_V(C))$ is a bijection to $\mathcal{F}_V^C$ resp. to $\mathcal{F}_{V,\perp}^C$ with inverse mapping π_V by Lemma 3.34. The ranges are included in the face lattice of C respectively in the exposed face lattice of C by Lemma 3.36.

The mappings L_V^C and π_V are inverse to each other and they are isotone on the considered domains with respect to set inclusion by Lemma 3.34. Hence the lift is a lattice isomorphism in each case by Lemma 2.32. The formula for lift and projection of exposed faces follows from Lemma 3.36.

Finally, by direct sum structure of $\mathbb{R}^m = V + V^\perp$ we have for faces F, G of $\pi_V(C)$

$$L_V^C(F \cap G) = ((F \cap G) + V^\perp) \cap C = (F + V^\perp) \cap (G + V^\perp) \cap C = L_V^C(F) \cap L_V^C(G),$$

the infimum in the lifted face lattices is the intersection. `qed`

Proof of Proposition 3.39. If the face $F \in \mathcal{F}(C)$ belongs to the lifted face lattice $\mathcal{F}_V^C$ then there is a face $G \in \mathcal{F}(\pi_V(C))$ such that $F = L_V^C(G)$. With Lemma 3.34 we find

$$L_V^C(F) = L_V^C \circ L_V^C(G) = L_V^C(G) = F.$$

Conversely we assume that a face $F \in \mathcal{F}(C)$ satisfies $L_V^C(F) = F$. Let us first prove the inclusion

$$L_V^C(F(\pi_V(C), y)) \subset F,$$

where $y := \pi_V(x)$ for an arbitrary point $x \in F$. This inclusion depends on the assumption $L_V^C(F) = F$. One has

$$L_V^C(y) = L_V^C \circ \pi_V(x) \subset L_V^C \circ \pi_V(F) = L_V^C(F) = F$$

because L_V^C is isotone for subsets of $\pi_V(C)$ and since $L_V^C = L_V^C \circ \pi_V$, see Lemma 3.34. The map π_V is inverse to L_V^C for subsets of $\pi_V(C)$ so

$$y \in \mathrm{ri}(F(\pi_V(C), y)) = \mathrm{ri}(\pi_V \circ L_V^C(F(\pi_V(C), y))) = \pi_V(\mathrm{ri}(L_V^C(F(\pi_V(C), y))))$$

where the second equation is true because reduction to the relative interior for a convex set commutes with a linear map (3.15). We get

$$\emptyset \neq \mathrm{ri}(L_V^C(F(\pi_V(C), y))) \cap L_V^C(y).$$

Since $L_V^C(y) \subset F$, the relative interior of the face on the left hand side meets F and we get by Remark 3.9 (e) on page 54

$$L_V^C(F(\pi_V(C), y)) \subset F.$$

Provided that we choose $x \in \mathrm{ri}(F)$ then the converse inclusion

$$F \subset L_V^C(F(\pi_V(C), y))$$

holds independent of the assumption $L_V^C(F) = F$. Indeed, let $x \in \mathrm{ri}(F)$. Then $y = \pi_V(x) \in \mathrm{ri}(\pi_V(F))$, hence $\pi_V(F) \subset F(\pi_V(C), y)$ by Remark 3.9 (e). We obtain

$$F \subset L_V^C(F) = L_V^C \circ \pi_V(F) \subset L_V^C(F(\pi_V(C), y)).$$

Notice that the assumed equality $L_V^C(F) = F$ surpasses the generally valid and trivial inclusion $F \subset L_V^C(F)$ for a subset F of C. qed

Proof of Lemma 3.40. (a) $\Longrightarrow$ (b): let us assume a convex subset $F \subset C$ projects to $\mathrm{rb}(\pi_V(C))$. Then $\mathrm{ri}(\pi_V(F))$ meets a proper exposed face G of $\pi_V(C)$ by Remark 3.15 (f) because the relative boundary of $\pi_V(C)$ is covered by proper exposed faces. Since $\pi_V(F)$ is a convex set, Remark 3.9 (e) gives $\pi_V(F) \subset G$. Since $F \subset C$ the inclusion $F \subset L_V^C(F)$ is obvious. Then by Lemma 3.34 we have that

$$F \subset L_V^C(F) = L_V^C \circ \pi_V(F) \subset L_V^C(G).$$

In addition, by Proposition 3.38, the set $L_V^C(G)$ is a proper exposed face of C.

(b) $\Longrightarrow$ (c) is trivial. (c) $\Longrightarrow$ (a): a proper face G in the lifted face lattice $\mathcal{F}_V^C$ projects to a proper face $\pi_V(G)$ of $\pi_V(C)$ by Proposition 3.38. By the stratification property of a convex set we have $\pi_V(G) \subset \mathrm{rb}(\pi_V(C))$. If $F \subset G$, then F projects to the relative boundary *a fortiori.* qed

Proof of Lemma 3.41. Let $a \in C + V^\perp$. We use the duality (3.39) and the support function identity in Remark 3.35 (a) to prove the first identity. Let $u = v + w \in \mathbb{R}^n$ for $v \in V$ and $w \in V^\perp$. If $u \in \mathrm{N}(\pi_V(C), \pi_V(a))$ then

$$h(C + V^\perp, v) = h(\pi_V(C), v) = h(\pi_V(C), u) = \langle u, \pi_V(a) \rangle = \langle v, \pi_V(a) \rangle = \langle v, a \rangle,$$

so $v \in \mathrm{N}(C + V^\perp, a)$ and $u \in \mathrm{N}(C + V^\perp, a) + V^\perp$. Conversely, we assume $v \in \mathrm{N}(C + V^\perp, a)$. Then

$$\langle u, \pi_V(a) \rangle = \langle v, \pi_V(a) \rangle = \langle v, a \rangle = h(C + V^\perp, v) = h(\pi_V(C), v) = h(\pi_V(C), u),$$

so $u \in \mathrm{N}(\pi_V(C), \pi_V(a))$. The second equation comes from the intersection formula for normal cones (3.40). qed

Proof of Lemma 3.43. By direct sum decomposition of $\mathbb{R}^m = \mathrm{lin}(C) + \mathrm{lin}(C)^\perp$, the normal cone of C in the restricted ambient space $\mathrm{lin}(C)$ is $\mathrm{N}(C, a) \cap \mathrm{lin}(C)$ for any point $a \in C$. Then the distributive formula for intersection and relative interior (3.17) proves that an acute normal vector v for C is also acute normal for the restricted ambient space.

Assume that $v \in V$ is a non-zero vector. We can assume that $F_\perp(C, v) \neq \emptyset$. Otherwise v is trivially acute normal for C. Let $x \in \mathrm{ri}(F_\perp(\pi_V(C), v))$. By Proposition 3.38 we have

$$F_\perp(\pi_V(C), v) = \pi_V(F_\perp(C, v))$$

and we can choose a point $a \in \mathrm{ri}(F_\perp(C, v))$ such that $x = \pi_V(a)$, see (3.15). If v is acute normal for C then $v \in \mathrm{ri}(\mathrm{N}(C, a))$. The projection formula for normal cones in Lemma 3.41 gives

$$\mathrm{N}(\pi_V(C), x) = (\mathrm{N}(C, a) \cap V) + V^\perp.$$

Since v is in the relative interior of $\mathrm{N}(C, a)$ and in V the equations (3.16) and (3.17) give $v \in \mathrm{ri}(\mathrm{N}(\pi_V(C), x))$, that is, v is acute normal for $\pi_V(C)$. qed

Proof of Theorem 1. On the one hand coatoms of the face lattice are exposed faces and intersections of exposed faces are exposed by Lemma 3.16. For the converse it is sufficient to prove that a proper exposed face of C is either a coatom or that it is the intersection of two strictly larger exposed faces.

Justified by restriction of the ambient space $\mathbb{R}^m$ to $\mathrm{aff}(C)$, which is allowed by Lemma 3.43, let us assume that $\mathrm{int}(C) \neq \emptyset$. Let F be a proper exposed face of C. Notice that $\mathrm{N}(C, F)$ does not contain a line. Otherwise $\mathrm{int}(C) = \emptyset$ by Remark 3.19 (g) (iv). The normal cone $\mathrm{N}(C, F)$ is non-zero by Lemma 3.20 since $F \neq C$. So the normal cone of F is not an affine space. There are two cases to distinguish.

If the cone $\mathrm{N}(C, F)$ is a closed half of an affine space then it is a ray (no lines are contained in the cone). Then, if there is a face $G \in \mathcal{F}(C)$ containing F properly, we get

$$\mathrm{N}(C, G) = \mathrm{N}(C, \overset{\perp}{G}) \subsetneq \mathrm{N}(C, F)$$

by (3.48) and Proposition 3.29. Then $\mathrm{N}(C, G) = \{0\}$ because $\mathrm{N}(C, G)$ is a face of $\mathrm{N}(C, F)$ by Proposition 3.23. This implies $G = C$ so F is a coatom.

If the cone $\mathrm{N}(C, F)$ is not a closed half of an affine space then we apply Theorem 18.4 in [Ro] and this provides for a non-zero relative interior point $u \in \mathrm{ri}(\mathrm{N}(C, F)) \setminus \{0\}$ two relative boundary points v, w of $\mathrm{N}(C, F)$ such that u lies on the line segment joining v and w. Of course, u belongs to the relative interior of $[v, w]$ by the intersection formula of relative interiors (3.17). Since $\mathrm{N}(C, F)$ is a convex cone one has $v \neq 0$, otherwise for some $\lambda > 1$ we have $w = \lambda u$ and w would belong to the relative interior of the cone. Similarly $w \neq 0$. Thus

$$F = F_\perp(C, u) = F_\perp(C, v) \cap F_\perp(C, w)$$

by Lemma 3.27 and Lemma 3.16. The arguments so far are completely general. If v is acute normal for C then

$$v \in \mathrm{ri}(\mathrm{N}(C, F_\perp(C, v))).$$

In case $F = F_\perp(C, v)$ we get the contradiction $v \in \mathrm{ri}(\mathrm{N}(C, F))$. Hence $F \subsetneq F_\perp(C, v)$. Similarly one has $F \subsetneq F_\perp(C, w)$. qed

Proof of Lemma 3.46. Observe that every normal cone is the normal cone of an exposed face F by Proposition 3.29. By assumption the face F is non-empty and has a non-empty normal cone. Hence Lemma 3.27 shows that for some $u \in \mathrm{ri}(\mathrm{N}(C, F)) \setminus \{0\}$ we have $F = F_\perp(C, u)$. Now $u \in \mathrm{ri}(\mathrm{N}(C, F)) = \mathrm{ri}(\mathrm{N}(C, F_\perp(C, u)))$ gives $\mathrm{T}(C, u) = \mathrm{N}(C, F)$.

(a) Normal cones of non-empty faces are closed convex cones included in the barrier cone by Remark 3.19 (a) and (b). A face of a closed convex cone is a closed convex cone.

(b) Let $u \in \mathrm{ri}(K) \setminus \{0\}$. One has $\mathrm{ri}(K) \setminus \{0\} = \mathrm{ri}(\mathrm{conv}(K \setminus \{0\})) \setminus \{0\}$ by (10.2) on page 196. Thus

$$F_\perp(C, u) = \bigcap_{v \in K \setminus \{0\}} F_\perp(C, v) \tag{10.3}$$

by Lemma 3.16. The right-hand side is independent of the choice of u.

(c) We can assume that $K = T(C, u)$ for some $u \in B(C) \setminus \{0\}$. By definition of a touching cone $u \in \mathrm{ri}(K)$ holds. If we choose $v \in \mathrm{ri}(K) \setminus \{0\}$ then by (b) we have $F_\perp(C, u) = F_\perp(C, v)$, hence $T(C, u) = T(C, v)$.

(d) Note that a cone with zero in the relative interior is a linear space. With a non-zero $u \in \mathrm{ri}(K)$ the opposite vector $-u$ belongs also to $\mathrm{ri}(K)$. Then from (b) follows

$$F_\perp(C, u) = F_\perp(C, -u)$$

so $C = F_\perp(C, u)$. The normal cone of C is $\mathrm{N}(C, C) = \mathrm{lin}(C)^\perp$ by Lemma 3.20 hence

$$K(C, u) = F(\mathrm{N}(C, F_\perp(C, u)), u) = F(\mathrm{N}(C, C), u) = F(\mathrm{lin}(C)^\perp, u) = \mathrm{lin}(C)^\perp. \qquad \texttt{qed}$$

Proof of Proposition 3.47. Relative interiors of distinct touching cones do not meet at non-zero vectors by Lemma 3.46 (c) and they do not meet at zero by (d) of the lemma. If C is closed then the face of C exposed by a non-zero vector $u \in B(C)$ in non-empty. Thus the touching cone $T(C, u)$ is defined and u belongs to the relative interior of this cone. qed

Proof of Proposition 3.48. Let K be a touching cone of C and let us assume that $u \in \mathrm{ri}(K) \setminus \{0\}$ is acute normal for C. Then for a point $x \in \mathrm{ri}(F_\perp(C,u))$ we have $u \in \mathrm{ri}(\mathrm{N}(C,x))$. By definition of the normal cone of a face we have $\mathrm{N}(C,x) = \mathrm{N}(C, F_\perp(C,u))$ hence $u \in \mathrm{ri}(\mathrm{N}(C, F_\perp(C,u)))$ and this gives us $T(C,u) = \mathrm{N}(C, F_\perp(C,u))$. Since $u \in \mathrm{ri}(K)$ we have by Lemma 3.46 (c) $K = T(C,u)$. Hence K is the normal cone of the non-empty face $F_\perp(C,u)$.

Conversely let us assume that the touching cone K is the normal cone of a non-empty face of C. Then by Proposition 3.29 we have $K = \mathrm{N}(C,F)$ for some non-empty exposed face F of C and Lemma 3.27 shows that $F = F_\perp(C,u)$ for any non-zero $u \in \mathrm{ri}(K)$. Then for a point $x \in \mathrm{ri}(F_\perp(C,u))$

$$\mathrm{N}(C,x) = \mathrm{N}(C, F_\perp(C,u)) = \mathrm{N}(C,F) = K$$

holds and this shows that $u \in \mathrm{ri}(K) = \mathrm{ri}(\mathrm{N}(C,x))$. So u is acute normal for C. qed

Proof of Corollary 3.49. A vector $u \in B(C) \setminus \{0\}$ is in $\mathrm{ri}(K)$ for a (unique) touching cone K of C by Proposition 3.47. By Proposition 3.48 the cone K is the normal cone of a non-empty face of C if and only if u is acute normal. qed

Proof of Lemma 3.50. It is sufficient to prove that the face $F(C \cap \mathbb{A}, x)$ is exposed for $x \in \mathrm{rb}(C \cap \mathbb{A})$. Observe that x belongs to the relative boundary of C (by the intersection formula (3.17) for relative interiors). If the face of x in C is exposed by a non-zero vector u then by the maximum property (3.34) of exposed faces we have

$$x \in F_\perp(C,u) \cap \mathbb{A} = F_\perp(C \cap \mathbb{A}, u).$$

The sets on the left-hand side share the relative interior point x so the intersection formula (3.17) for relative interiors can be used,

$$x \in \mathrm{ri}(F_\perp(C,u)) \cap \mathbb{A} = \mathrm{ri}(F_\perp(C,u) \cap \mathbb{A}) = \mathrm{ri}(F_\perp(C \cap \mathbb{A}, u)).$$

This completes the proof. qed

Proof of Lemma 3.51. Recall that for each $x \in \mathrm{ri}(F)$ we have $F = F(C,x)$ by the stratification property (3.23) of C into relative interiors of faces. Moreover the normal cone (3.42) of F is $\mathrm{N}(C,F) = \mathrm{N}(C,x)$.

If $x \in \mathrm{ri}(F)$ and F is exposed then Lemma 3.27 tells us that $F = F_\perp(C,u)$ for each vector $u \in \mathrm{ri}(\mathrm{N}(C,x)) \setminus \{0\}$. So $x \in \mathrm{ri}(F) = \mathrm{ri}(F_\perp(C,u))$ proves that all $x \in \mathrm{ri}(F)$ are acute exposed.

Choose $x \in \mathrm{ri}(F)$ and pick a non-zero vector $u \in \mathrm{ri}(\mathrm{N}(C,x)) = \mathrm{ri}(\mathrm{N}(C,F))$. If x is acute exposed for C then $x \in \mathrm{ri}(F_\perp(C,u))$, hence $F = F_\perp(C,u)$ is an exposed face. If $\mathrm{N}(C,x) = \{0\}$ then $F = C$ by Lemma 3.20, which is also an exposed face. qed

Proof of Remark 4.5. (a) For $\rho, \sigma \in \overline{S}(A)$ the first inequality in the term

$$\|\rho - \sigma\|_2^2 = \|\rho\|_2^2 + \|\sigma\|_2^2 - 2\,\mathrm{tr}(\sqrt{\rho}\sigma\sqrt{\rho}) \leq \|\rho\|_2^2 + \|\sigma\|_2^2 \leq \|\rho\| + \|\sigma\| \leq 2.$$

comes from positivity of the matrix $\sqrt{\rho}\sigma\sqrt{\rho}$. The second inequality comes from the inequality $|\,\mathrm{tr}(ab)| \leq \|a\|_1 \|b\|$ with the trace norm $\|\cdot\|_1$ (2.12). The third inequality is due to the spectral bound of density matrices.

(b) Since $\overline{S}(A)$ is a compact convex set, the Minkowski theorem (3.26) implies that the state space $\overline{S}(A)$ is the convex hull of its extreme points. We can assume that two density matrices $\rho, \sigma \in \overline{S}(A)$ are convex combinations of rank one projectors p_i and q_j, with non-zero coefficients λ_i and μ_j such that $\rho = \sum_i \lambda_i p_i$ and $\sigma = \sum_j \mu_j q_j$. Then

$$\|\rho - \sigma\| = \|\sum_{i,j} \lambda_i \mu_j (p_i - q_j)\| \leq \sum_{i,j} \lambda_i \mu_j \|p_i - q_j\| \leq \sum_{i,j} \lambda_i \mu_j = 1.$$

The inequality $\|p - q\| \leq 1$ for arbitrary orthogonal projectors is well-known [Av]. We prove that orthogonality of ρ and σ is necessary for the equality to hold. If the second of the above inequalities is sharp then $\|p_i - q_j\| = 1$ for all i, j. This shows that -1 or 1 is an eigenvalue of $p_i - q_j$. The corresponding spectral projectors are $(\mathbb{1} - p_i) \wedge q_j$ and $p_i \wedge (\mathbb{1} - q_j)$. This follows by decomposition in commuting and non-commuting parts of p_i and q_j, see [Bo]. If $(\mathbb{1} - p_i) \wedge q_j$ has rank one, then $q_j \leq (\mathbb{1} - p_i)$ and this gives $p_i q_j = 0$ by (2.39). Similarly we argue if $p_i \wedge (\mathbb{1} - q_j)$ has rank one then $p_i q_j = 0$. The equality $p_i q_j = 0$ holds for arbitrary i, j so $\rho\sigma = 0$. qed

Proof of the intersection for Remark 4.7. The projection shape in Figure 4.1 is represented in coordinates of the basis of V

$$v_1 := \frac{1}{2}(\sigma_1 - \sigma_2) \oplus 0 \qquad \text{and} \qquad v_2 := \frac{1}{2}(\sigma_1 + \sigma_2) \oplus 1 - \frac{\mathbb{1}}{3}.$$

We use this basis to calculate the intersection $\overline{S}(M_2 \oplus \mathbb{C}) \cap (\frac{\mathbb{1}}{3} + V)$. Let

$$x = \frac{\mathbb{1}}{3} + \lambda v_1 + \mu v_2.$$

We denote the identity in M_2 by $\mathbb{1}_2$. Since

$$x = \left(\mathbb{1}_2 \frac{1}{3}(1 - \mu) + \frac{1}{2}(\sigma_1(\lambda + \mu) + \sigma_2(\mu - \lambda))\right) \oplus \frac{1}{3}(1 + 2\mu)$$

we obtain from (4.4) the eigenvalues $a_\pm := \frac{1}{3}(1-\mu) \pm \frac{1}{\sqrt{2}}\sqrt{\lambda^2+\mu^2}$ of x in addition to $a_3 := \frac{1}{3}(1+2\mu)$. Positivity of a_3 is equivalent to

$$\mu \geq -\frac{1}{2}. \tag{10.4}$$

A short calculation gives that positivity of a_- alone is equivalent to

$$\frac{\lambda^2}{(\sqrt{\frac{2}{7}})^2} + \frac{(\mu+\frac{2}{7})^2}{(\frac{3\sqrt{2}}{7})^2} \leq 1. \tag{10.5}$$

Clearly the conditions 10.4 and 10.5 are sufficient for positivity of x. This gives the truncated ellipse in the figure. For a correct scaling notice that $\|v_1\|_2 = 1$ and $\|v_2\|_2 = \sqrt{\frac{5}{3}}$ hold for the orthogonal vectors v_1 and v_2. qed

Proof of Lemma 4.8. This follows by induction from the case $N = 2$. One has $\overline{S}(A) = [(A_1)^+ \oplus (A_2)^+] \cap A^1_{\mathrm{sa}}$. A state ρ has the form $\rho = a_1 \oplus a_2$ for $a_i \in (A_i)^+$, $i = 1, 2$. One has $1 = \mathrm{tr}(\rho) = \mathrm{tr}(a_1) + \mathrm{tr}(a_2)$. Then either $\rho = a_1 \oplus 0$, $\rho = 0 \oplus a_2$ or $\rho = \mathrm{tr}(a_1)\frac{a_1}{\mathrm{tr}(a_1)} \oplus (1-\mathrm{tr}(a_1))\frac{a_2}{\mathrm{tr}(a_2)}$. Conversely, $\lambda\rho_1 \oplus (1-\lambda)\rho_2 \in \overline{S}(A)$ if $\rho_i \in \overline{S}(A_i)$ for $i = 1, 2$ and $\lambda \in [0, 1]$. qed

Proof of Proposition 4.11. As observed in Remark 2.11 (d) on page 35, the positive cone is the intersection

$$A^+ = \bigcap_{u \in H\setminus\{0\}} \{a \in A_{\mathrm{sa}} : \langle u, a(u)\rangle \geq 0\}.$$

This makes clear that A^+ is a convex cone. If $P_u \in \mathcal{B}(H)$ for non-zero $u \in A_{\mathrm{sa}}$ denotes the orthogonal projector to the linear span of u, the intersection representation

$$A^+ = \bigcap_{u \in H\setminus\{0\}} \{a \in A_{\mathrm{sa}} : \mathrm{tr}(P_u a) \geq 0\}$$

by closed half spaces is available, hence A^+ is closed.

Since every self-adjoint matrix $a \in A_{\mathrm{sa}}$ is the difference $a = a^+ - a^-$ of two positive matrices $a^+, a^- \in A^+$, see Remark 2.11 (f), the affine hull of the positive cone is as large as A_{sa}. The space of self-adjoint matrices is a vector space, so the translation vector space of the positive cone is $\mathrm{lin}(A^+) = A_{\mathrm{sa}}$.

For a convex cone the support function is either 0 or ∞ depending on whether being evaluated for a point inside or a point outside the barrier cone. To calculate the barrier cone, let $a, b \in A^+$. By Remark 2.11 (c)

$$\langle -b, a\rangle = -\operatorname{tr}(\sqrt{b}a\sqrt{b}) \leq 0$$

holds, so $-A^+ \subset B(A^+)$. Conversely, if $b \in A_{\mathrm{sa}} \setminus (-A^+)$ then the largest eigenvalue of b is positive, $\mu_+(b) > 0$. Thus

$$\sup_{a \in A^+} \langle b, a\rangle \geq \sup_{\lambda > 0} \langle b, \lambda p_+(b)\rangle = +\infty$$

for the eigenprojector $p_+(b)$ of b corresponding to the eigenvalue $\mu_+(b)$. This proves $B(A^+) = -A^+$ and hence for $b \in A_{\mathrm{sa}}$

$$h(A^+, -b) = \begin{cases} 0 & \text{if } b \in A^+, \\ +\infty & \text{otherwise,} \end{cases}$$

holds. We can calculate the interior of the positive cone A^+ using Theorem 13.1 in [Ro]. The theorem says that a point $a \in A_{\mathrm{sa}}$ belongs to $\operatorname{int}(A^+)$ if and only if for all non-zero $b \in A_{\mathrm{sa}}$

$$\langle b, a\rangle < h(A^+, b)$$

holds. Trivially, if $b \notin B(A^+) = -A^+$ then $\langle b, a\rangle < h(A^+, b) = \infty$. Otherwise, if $b \in A^+$ then $h(A^+, -b) = 0$. Since $\langle a, b\rangle \geq 0$ for all $a \in A^+$ one has

$$\langle -b, a\rangle < 0 \qquad \iff \qquad s(a)s(b) \neq 0$$

because $a \perp b$ is equivalent to $s(a)s(b) = 0$ for positive matrices, see (2.26). The property $s(a)s(b) \neq 0$ is true for all $b \in A^+$ if and only if $s(a) = \mathbb{1}$, so the interior of the positive cone consists of positive matrices a with full support $s(a) = \mathbb{1}$.

By definition, given a non-zero vector $b \in A^+$, the face of A^+ exposed by $-b$ contains all positive matrices $a \in A^+$ such that $\langle -b, a\rangle = h(A^+, -b)$. The support function is constant zero on $B(A^+) = -A^+$ hence a matrix in $F_\perp(A^+, -b)$ satisfies $a \perp b$. By the argument in the previous paragraph this is $s(a)s(b) = 0$. The latter equation is equivalent to $s(a)k(b) = s(a)$ with the kernel projector $k(b)$ of b. Using order theoretic relations for projectors (2.39) this is $s(a) \leq k(b)$. We have proved $F_\perp(A^+, -b) = \{a \in A^+ : s(a) \leq k(b)\}$.

With Remark 2.22 (b) on page 40 we can use the compression $A^{k(b)}$ to describe exposed faces, their affine span and translation vector space. Relative interiors of faces can be computed using Remark 3.3 (a) on page 52 because $\kappa^{k(b)}$ is linear. qed

Proof of Corollary 4.13. The vector $p - \mathbb{1}$ for an orthogonal projector $p \in \mathcal{P} \setminus \{\mathbb{1}\}$ exposes the face

$$F_\perp(A^+, p - \mathbb{1}) = \kappa^p((A^p)^+)$$

by (4.15). So the mapping $p \mapsto \kappa^p((A^p)^+)$ assigns faces of the positive cone A^+ to projectors of A. The image of the identity is $\kappa^{\mathbb{1}}((A^{\mathbb{1}})^+) = A^+$. The relative interior of each of the faces is

$$\mathrm{ri}(\kappa^p((A^p)^+)) = \{a \in A^+ : s(a) = p\}$$

by (4.17) for the proper faces and by (4.14) for A^+. The relative interior $\mathrm{ri}(\kappa^p((A^p)^+))$ contains the unique projector p, so the mapping

$$\mathcal{P} \to \mathcal{F}(A^+), \quad p \mapsto \kappa^p((A^p)^+)$$

is injective. A convex set is the disjoint union of relative interiors of faces (3.23) and here the relative interiors of the considered faces already cover A^+, hence the mapping is a bijection $\mathcal{P} \to \mathcal{F}(A^+) \setminus \{\emptyset\}$. The mapping and the inverse are isotone. This follows immediately from

$$\kappa^p((A^p)^+) = \{a \in A^+ : s(a) \leq p\}.$$

Hence by Lemma 2.32 the mapping is a lattice isomorphism. Both lattices were proved to be complete lattices: the projector lattice of von Neumann algebra is considered in Lemma 2.29 on page 45, the face lattice of a convex set in (3.22) on page 54. qed

Proof of Corollary 4.14. The barrier cone of the positive cone is $B(A^+) = -A^+$, see (4.12). The normal cone of A^+ at $a \in A^+$ is included in the barrier cone, $\mathrm{N}(A^+, a) \subset -A^+$, see (3.37). The zero vector belongs to every normal cone, so consider a non-zero vector in the barrier cone $b \in -A^+$. Then

$$b \in \mathrm{N}(A^+, a) \iff a \in F_\perp(A^+, b)$$

by the duality (3.39) between normal cones and exposed faces. By the expression (4.15) for the exposed face

$$F_\perp(A^+, b) = \kappa^{k(b)}\left((A^{k(b)})^+\right)$$

this is equivalent to $s(a) \leq k(b)$ or likewise to $s(b) \leq k(a)$. So this is the same as $-b \in \kappa^{k(a)}\left((A^{k(a)})^+\right)$. The relative interior is computed with (4.17) for non-zero a or with (4.14) for $a = 0$ at the apex of the positive cone. qed

Proof of Proposition 4.18. Many results are derived from the positive cone A^+, see Proposition 4.11. The intersection $\overline{S}(A) = A^+ \cap A^1_{\mathrm{sa}}$ is closed and bounded. Explicit bounds in HS norm and C*-norm are calculated in Remark 4.5.

Since the trace state $\widehat{\mathbb{1}}$ of A belongs to both the relative interior of the positive cone and the affine space of self-adjoint trace one matrices, the intersection formulas (3.18) and (3.17) give

$$\operatorname{aff}(\overline{S}(A)) = \operatorname{aff}(A^+ \cap A^1_{\text{sa}}) = \operatorname{aff}(A^+) \cap A^1_{\text{sa}} = A^1_{\text{sa}}$$

and

$$\operatorname{ri}(\overline{S}(A)) = \operatorname{ri}(A^+ \cap A^1_{\text{sa}}) = \operatorname{ri}(A^+) \cap \operatorname{ri}(A^1_{\text{sa}}) = \operatorname{ri}(A^+) \cap A^1_{\text{sa}}.$$

The relative interior of the positive cone consists of the invertible positive matrices (4.14) so $\operatorname{ri}(\overline{S}(A))$ is the space of invertible density matrices $S(A) = \{\rho \in \overline{S}(A) : s(\rho) = \mathbb{1}\}$ (4.2). From the affine hull we deduce that $\operatorname{lin}(\overline{S}(A)) = A^0_{\text{sa}}$ so the dimension formula follows and it is complemented numerically by (2.15) on page 37.

The barrier cone is the whole space of self-adjoint matrices A_{sa} since $\overline{S}(A)$ is compact. Let us calculate the support function first for vectors in the relative boundary of the barrier cone

$$u \in -(A^+ \setminus \operatorname{ri}(A^+)) \subset B(A^+).$$

Since the state space is included in the positive cone,

$$h(u) \leq h(A^+, u).$$

On the other hand the normalized kernel projector $\frac{k(u)}{\operatorname{tr}(k(u))}$ lies in $H(A^+, u) \cap \overline{S}(A)$ by (4.15) and therefore

$$h(A^+, u) = \langle u, \frac{k(u)}{\operatorname{tr}(k(u))} \rangle \leq h(u).$$

We have $h(u) = 0$ for all $u \in -(A^+ \setminus \operatorname{ri}(A^+))$, that is $h(u) = \mu_+(u)$ is the largest eigenvalue of u. Observe the decomposition

$$A_{\text{sa}} = \mathbb{R} \cdot \mathbb{1} - (A^+ \setminus \operatorname{ri}(A^+)) = \{\lambda \mathbb{1} - a : \lambda \in \mathbb{R}, a \in A^+ \setminus \operatorname{ri}(A^+)\}. \tag{10.6}$$

This is true because a self-adjoint matrix $u \in A_{\text{sa}}$ can be written

$$u = \mu_+(u)\mathbb{1} - (\mu_+(u)\mathbb{1} - u).$$

Since a density matrix has trace one we find for arbitrary $u \in A_{\text{sa}}$ that

$$\begin{aligned} h(u) = \sup_{\rho \in \overline{S}(A)} \langle u, \rho \rangle &= \sup_{\rho \in \overline{S}(A)} \langle \mu_+(u)\mathbb{1} - (\mu_+(u)\mathbb{1} - u), \rho \rangle \\ &= \mu_+(u) + \sup_{\rho \in \overline{S}(A)} \langle u - \mu_+(u)\mathbb{1}, \rho \rangle = \mu_+(u). \end{aligned}$$

Let us prove that all faces of $\overline{S}(A)$ are exposed. By the intersection $\overline{S}(A) = A^+ \cap A^1_{\text{sa}}$, the faces of $\overline{S}(A)$ are given by intersection of faces of the positive cone A^+ with A^1_{sa}, see

Remark 3.11 (b). All faces of the positive cone A^+ are exposed by Corollary 4.13. Then Lemma 3.50 transfers this property to the state space. Let us calculate the exposed face $F_\perp(u)$ for non-zero vectors $u \in -(A^+ \setminus \mathrm{int}(A^+))$ first. One has

$$\begin{aligned} F_\perp(u) &= A^+ \cap A^1_{\mathrm{sa}} \cap H(A^+, u) = F_\perp(A^+, u) \cap A^1_{\mathrm{sa}} \\ &= \kappa^{k(u)}\left((A^{k(u)})^+\right) \cap A^1_{\mathrm{sa}}. \end{aligned}$$

By the properties stated in Remark 2.22 (c) for the *-monomorphism $\kappa^{k(u)}$ we find that

$$F_\perp(u) = \kappa^{k(u)}\left((A^{k(u)})^+ \cap (A^{k(u)})^1_{\mathrm{sa}}\right) = \kappa^{k(u)}\left(\overline{S}(A^{k(u)})\right).$$

Since the kernel projector of u is the maximal eigenprojector, $k(u) = p_+(u)$, this gives

$$F_\perp(u) = \kappa^{p_+(u)}\left(\overline{S}(A^{p_+(u)})\right).$$

By the decomposition (10.6) and invariance of the previous formula under the additive group $\mathbb{R}\cdot\mathbb{1}$, the formula is true for all non-zero vectors $u \in A_{\mathrm{sa}}$. The affine hull, translation vector space and relative interior are derived from the state space with Remark 3.3 (b) because $\kappa^{p_+(u)}$ is linear. qed

Proof of Corollary 4.20. The arguments are the same as in the proof for Corollary 4.13 with some little changes. For an orthogonal projector $p \in \mathcal{P} \setminus \{0\}$, the vector p exposes the face with support p

$$F_\perp(p) = \mathbb{F}(p)$$

by (4.29). The relative interior is

$$\mathrm{ri}\,(\mathbb{F}(p)) = \{\rho \in \overline{S}(A) : s(\rho) = p\}$$

by (4.32). The relative interiors $\mathrm{ri}\,(\mathbb{F}(p))$ for non-zero $p \in \mathcal{P}$ cover the state space, so by the stratification property (3.23) the mapping $\mathcal{P} \to \mathcal{F}$ is onto. Conversely, given $p \neq 0$, each of the faces $\mathbb{F}(p)$ contains the centroid $\frac{p}{\mathrm{tr}(p)}$ in the relative interior, thus the mapping is injective. The mapping and the inverse are both isotone, hence by Lemma 2.32 it is a lattice isomorphism. Both lattices are complete: the projector lattice of a von Neumann algebra is considered in Lemma 2.29 on page 45, the face lattice of a convex set in (3.22) on page 54. The remaining properties of the faces $\mathbb{F}(p)$ follow from Proposition 4.18. qed

Proof of Proposition 4.24. The first equation follows from the duality between exposed faces and normal cones (3.39) on page 60 together with the description of exposed faces (4.29). For $a \in A_{\mathrm{sa}} \setminus \{0\}$ and $\rho \in \overline{S}(A)$

$$a \in \mathrm{N}(\rho) \iff \rho \in F_\perp(a) \iff s(\rho) \leq p_+(a)$$

holds. As argued in (3.41) on page 61 the normal cone is a sum

$$\mathrm{N}(\rho) = \mathrm{N}(A^+ \cap A^1_{\mathrm{sa}}, \rho) = \mathrm{N}(A^+, \rho) + \mathrm{N}(A^1_{\mathrm{sa}}, \rho).$$

because the positive cone and the affine space of trace one self-adjoint matrices share a relative interior point, the trace state $\widehat{\mathbb{1}} = \frac{\mathbb{1}}{\mathrm{tr}(\mathbb{1})}$. By (3.16) the relative interior of a sum of convex sets is the sum of their relative interiors, so

$$\mathrm{ri}\,(\mathrm{N}(\rho)) = \mathrm{ri}\left(\mathrm{N}(A^+, \rho)\right) + \mathrm{ri}\left(\mathrm{N}(A^1_{\mathrm{sa}}, \rho)\right).$$

We have $\mathrm{ri}\,(\mathrm{N}(A^1_{\mathrm{sa}}, \rho)) = \mathrm{N}(A^1_{\mathrm{sa}}, \rho) = \mathbb{R} \cdot \mathbb{1}$ and from Corollary 4.14 we recall that

$$\begin{aligned} \mathrm{ri}(\mathrm{N}(A^+, \rho)) &= -\kappa^{k(\rho)}\left(\mathrm{ri}(A^{k(\rho)})^+\right) \\ &= \{a \in -A^+ : s(a) = k(\rho)\} \\ &= \{a \in -A^+ : k(a) = s(\rho)\} \\ &= \{a \in -A^+ : p_+(a) = s(\rho)\} \end{aligned}$$

where $p_+(a)$ is the maximal eigenprojector of a. Then we find

$$\mathrm{ri}\,(\mathrm{N}(\rho)) = \{a \in -A^+ : p_+(a) = s(\rho)\} + \mathbb{R}\mathbb{1} = \{a \in A_{\mathrm{sa}} : p_+(a) = s(\rho)\}. \qquad \texttt{qed}$$

Proof of Lemma 4.28. For each unitary $v \in A$ and a self-adjoint matrix $a \in A_{\mathrm{sa}}$ the support projector of a is equivariant under conjugation with v: $v(s(\rho))v^* = s(v\rho v^*)$ by the spectral theorem. On the other hand, for two orthogonal projectors $p, q \in \mathcal{P}$ the conditions $v^*qv \leq p$ and $q \leq vpv^*$ are equivalent by (2.39) on page 43. Using these relations it is straight forward to show that

$$v\mathbb{F}(p)v^* = \mathbb{F}(vpv^*)$$

holds for arbitrary $p \in \mathcal{P}$ and unitaries $v \in A$. Using the isomorphism between projector lattice and face lattice in Corollary 4.20 we may assume that two faces F, G are of the form $F = \mathbb{F}(p)$ and $G = \mathbb{F}(q)$ for two projectors $p, q \in \mathcal{P}$. Then

$$G = vFv^* \iff \mathbb{F}(q) = v\mathbb{F}(p)v^* \iff \mathbb{F}(q) = \mathbb{F}(vpv^*)$$

where the right-hand side is equivalent to $q = vpv^*$ for $p = s(F)$ and $q = s(G)$.

Finally, by Lemma 2.38 on page 49 two projectors $p, q \in \mathcal{P}$ are conjugate if and only if they belong to the same conjugation manifold $\mathcal{P}_k$ for a multi-index $k \in \mathbb{N}_0^N$ with $k \leq n$. Hence two faces are conjugate if and only if they belong to the same face manifold. `qed`

Proof of Proposition 4.32. Let us prove continuity of the mapping $\mathcal{P} \to \mathcal{F}$, $p \mapsto \mathbb{F}(p)$. Let $(p_i) \subset \mathcal{P}$ be a sequence of projectors with limit $p = \lim_{i\to\infty} p_i$. We show $\mathbb{F}(p) = \lim_{i\to\infty} \mathbb{F}(p_i)$ using the conditions in Remark 4.31.

For condition (a) we have to show that every point $\rho \in \mathbb{F}(p)$ is a limit of points $\rho_i \in \mathbb{F}(p_i)$. Let us assume that $\|p - p_i\| < 1$ for all $i \in \mathbb{N}$. By Remark 2.39 on page 49 the unitaries $v_i := \operatorname{sgn}(\mathbb{1} - p_i - p)$ satisfy $p = v_i p_i v_i^*$ and

$$\lim_{i\to\infty} v_i = \mathbb{1} - 2p.$$

Given an arbitrary density matrix $\rho \in \mathbb{F}(p)$, we put $\rho_i := v_i^* \rho v_i$ for $i \in \mathbb{N}$. Then each ρ_i is element of the conjugate face $\mathbb{F}(p_i) = v_i^* \mathbb{F}(p) v_i$ by Lemma 4.28 and the limit

$$\lim_{i\to\infty} \rho_i = \lim_{i\to\infty} v_i^* \rho v_i = (\mathbb{1} - 2p)\rho(\mathbb{1} - 2p) = \rho$$

completes the first condition of face convergence.

For condition (b) we have to show that the limit of any convergent sequence (ρ_{i_j}) belongs to $\mathbb{F}(p)$ where $\rho_{i_j} \in \mathbb{F}(p_{i_j})$ for a subsequence (p_{i_j}) of (p_i), $j \in \mathbb{N}$. Let $\rho := \lim_{j\to\infty} \rho_{i_j}$. Each state ρ_{i_j} may be written $\rho_{i_j} = p_{i_j} \rho_{i_j} p_{i_j}$ for the support projector $p_{i_j} := s(\rho_{i_j})$. By joint continuity of matrix multiplication we get

$$\rho = \lim_{i\to\infty} \rho_{i_j} = \lim_{i\to\infty} p_{i_j} \rho_{i_j} p_{i_j} = p\rho p.$$

The support projector $s(p\rho p)$ is the infimum of all projectors $q \in \mathcal{P}$ with $q(p\rho p) = p\rho p$, see (2.46). Thus

$$s(\rho) = s(p\rho p) \leq p$$

and ρ belongs to $\mathbb{F}(p)$ by definition (4.33). This concludes the proof of continuity of the mapping $\mathcal{P} \to \mathcal{F}$, $p \mapsto \mathbb{F}(p)$.

Now the projector lattice $\mathcal{P}$ is compact by Lemma 2.38. Since the lattice isomorphism $\mathbb{F} : \mathcal{P} \to \mathcal{F}$ is continuous and injective (see Corollary 4.20) it is a homeomorphism to its image, to the face lattice $\mathcal{F}$. As described in Lemma 2.38, the projector lattice $\mathcal{P}$ is a union of compact real analytic manifolds, of the conjugation manifold $\mathcal{P}_k$ for multi-indices $k \in \mathbb{N}_0^N$ with $k \leq n$. Each face manifold $\mathcal{F}_k \subset \mathcal{F}$ inherits the the structure of a compact real analytic manifold from $\mathcal{P}_k \subset \mathcal{P}$ through the homeomorphism $\mathbb{F} : \mathcal{P} \to \mathcal{F}$, see Definition 1.6, 1.7 and 1.8 in [Ga] for the definition of a differentiable manifold. qed

Proof of Proposition 4.35. The union of faces in a face manifold $\bigcup \mathcal{F}_k$ is a compact subset of the state space $\overline{S}(A)$. Assume (ρ_i) is an arbitrary sequence in $\bigcup \mathcal{F}_k$. We will find a converging subsequence. The state space is compact by Proposition 4.18. Thus we can assume that (ρ_i) converges and it is sufficient to show that the limit $\rho := \lim_{i\to\infty} \rho_i$ belongs to $\bigcup \mathcal{F}_k$. There is a sequence of faces (F_i) in $\mathcal{F}_k$ with $\rho_i \in F_i$ for all $i \in \mathbb{N}$. By Proposition 4.32 the face manifold $\mathcal{F}_k$ is compact for the Hausdorff distance. Thus we can select a subsequence of faces (F_{i_j}) with limit $F = \lim_{j\to\infty} F_{i_j} \in \mathcal{F}_k$. Then by condition (b) in Remark 4.31 we have $\lim_{j\to\infty} \rho_{i_j} \in F \subset \bigcup \mathcal{F}_k$.

By (4.59), for $d \in \mathbb{N}_0$ the d-skeleton of the state space is a finite union of compact unions over face manifolds. Hence the d-skeleton is compact. The closedness of all skeletons is equivalent to lower semi-continuity of the dimension function, see Remark 4.34 (c). `qed`

Proof of Lemma 4.36. For a proof we use Lemma 3.5 and show that the level sets of the rank function on $M := \mathbb{C}^{k_1 \times k_2}$ are closed. Let $k := \min\{k_1, k_2\}$ and choose $a \in M$. For $1 \leq l \leq k$, an l-minor of a is the number $\det(\widetilde{a})$ where $\widetilde{a}$ is an $l \times l$ matrix obtained from a by deleting columns and rows. It is well-known for $0 \leq m \leq k$ that $\mathrm{rk}(a) \leq m$ if and only if every l-minor of a is zero for $l = m+1, \ldots, k$, see [Fi]. Thus, the level set $\{a \in M : \mathrm{rk}(a) \leq m\}$ is the intersection of finitely many zero sets of polynomials in the coefficients of a. Each of these zero sets is a closed subset of M, hence the level set is closed. `qed`

Proof of Proposition 5.15. By Theorem 18.7 in [Ro] (a combination of Minkowski and Straszewicz Theorem), the compact set sr_V is the closure of the convex hull of exposed extreme points of sr_V,

$$\mathrm{sr}_V = \overline{\mathrm{conv}(E)},$$

where E denotes the set of exposed points of sr_V. By (5.15) an exposed point $x \in E$ is the projection of the face $\mathbb{F}(p)$ for some non-zero $p \in \mathcal{P}_{V,\perp}$. One point is sufficient to cover the zero-dimensional face $\{x\}$,

$$x = \pi_V\left(\frac{p}{\mathrm{tr}(p)}\right).$$

The exposed projector lattice $\mathcal{P}_{V,\perp}$ is derived from the vector space V by (5.16) through calculation of maximal projectors. Hence for some $v \in V$ we have $p = p_+(v)$ with the maximal projector $p_+(v)$ of v. With $v \in B$ also the spectral projectors of v belong to B and in particular $p \in B$. This shows $\frac{p}{\mathrm{tr}(p)} \in \overline{S}(B)$ and therefore

$$E \subset \pi_V(\overline{S}(B)).$$

Since $\pi_V(\overline{S}(B))$ is closed and convex we get $\mathrm{sr}_V \subset \pi_V(\overline{S}(B))$. Since $B \subset A$, the converse inclusion $\pi_V(\overline{S}(B)) \subset \pi_V(\overline{S}(A))$ is obvious. `qed`

Proof of Lemma 5.17. Let $p \in \mathcal{P}$ be a non-zero projector, let $V \subset \mathrm{lin}(\mathbb{F}(p))$ and put $W := (\kappa^p)^{-1}(V) \subset (A^p)^0_{\mathrm{sa}}$. Since κ^p is an isometry, one has for $q \in \mathcal{P}(A^p)$

$$\kappa^p(\mathbb{F}_W(q)) = \kappa^p(\pi_W(\mathbb{F}(q)) = \pi_{\kappa^p(W)}(\kappa^p(\mathbb{F}(q)).$$

Using the composition (5.59) of *-isomorphisms on state spaces, we obtain the result

$$\kappa^p(\mathbb{F}_W(q)) = \pi_V(\mathbb{F}(\kappa^p(q))).$$

This proves the first assertion. In the special case $q := (\kappa^p)^{-1}(p) = \mathbb{1}^p$ we have $\mathbb{F}_W(q) = \mathbb{F}_W(\mathbb{1}^p) = \mathrm{sr}_W$ and $\kappa^p(q) = p$. Then with $V \subset \mathrm{lin}(\mathbb{F}(p))$ we get

$$\kappa^p(\mathrm{sr}_W) = \pi_V(\mathbb{F}(p)) = \pi_V \circ \pi_{\mathrm{lin}(\mathbb{F}(p))}(\mathbb{F}(p)).$$

Since faces of a state space are “large” (5.5), we get

$$\pi_V \circ \pi_{\mathrm{lin}(\mathbb{F}(p))}(\mathbb{F}(p)) = \pi_V \circ \pi_{\mathrm{lin}(\mathbb{F}(p))}(\overline{S}(A)) = \pi_V(\overline{S}(A)) = \mathrm{sr}_V\,. \qquad \texttt{qed}$$

Proof of Proposition 5.20. We establish the left commuting diagram. The other two follow by restriction from this one. Let us discuss the top line first. By definition (4.33) of the face with support p we have $\mathbb{F}(p) = \kappa^p(\overline{S}(A^p))$ and this face has the affine hull (4.36)

$$\mathrm{aff}\,(\mathbb{F}(p)) = \kappa^p\left((A^p)^1_{\mathrm{sa}}\right).$$

The *-monomorphism κ^p restricts to an affine isomorphism $(A^p)^1_{\mathrm{sa}} \longrightarrow \mathrm{aff}(\mathbb{F}(p))$.

The mapping $\pi_{\varsigma^p(V)} : (A^p)^1_{\mathrm{sa}} \to \varsigma^p(V)$ on the left down-arrow of the diagram is onto. Since $\varsigma^p(V) \subset (A^p)^0_{\mathrm{sa}}$ one has

$$\pi_{\varsigma^p(V)}((A^p)^1_{\mathrm{sa}}) = \pi_{\varsigma^p(V)} \circ \pi_{(A^p)^0_{\mathrm{sa}}}((A^p)^1_{\mathrm{sa}}) = \pi_{\varsigma^p(V)}((A^p)^0_{\mathrm{sa}}) = \varsigma^p(V).$$

Hence, in case of existence of a map $\vartheta^p : \mathrm{aff}(\mathbb{F}_V(p)) \longrightarrow \varsigma^p(V)$ which satisfies the relation $\vartheta^p \circ \pi_V \circ \kappa^p = \pi_{\varsigma^p(V)}$ on the domain $(A^p)^1_{\mathrm{sa}}$, the mapping ϑ^p will be surjective. Since $\dim(\mathbb{F}_V(p)) = \dim(\varsigma^p(V))$ by (5.61) injectivity of ϑ^p follows also.

Observe (5.3) that the orthogonal projection of a matrix $v \in A_{\mathrm{sa}}$ to $\mathrm{lin}(\mathbb{F}(p))$ is given by

$$\pi_{\mathrm{lin}(\mathbb{F}(p))}(v) = pvp - \mathrm{tr}(pv)\frac{p}{\mathrm{tr}(p)}.$$

Let $\{x_i\}_{i=1}^r$ respectively $\{y_j\}_{j=1}^s$ be an ONB of V respectively of $\varsigma^p(V)$. Since $\kappa^p(\varsigma^p(V)) = \pi_{\mathrm{lin}(\mathbb{F}(p))}(V)$ there exist real coefficients $\{t_{i,j}\}_{i=0,\ldots,r}^{j=1,\ldots,s}$ such that for $j = 1, \ldots, s$

$$\kappa^p\left(y_j\right) = \sum_{i=1}^r t_{i,j} p x_i p + t_{0,j} p$$

holds. Now define for $v \in V$

$$\vartheta^p(v) := \sum_{j=1}^{s} \big\langle \sum_{k=1}^{r} t_{k,j} x_k, v - \frac{p}{\mathrm{tr}(p)} \big\rangle y_j.$$

Then for $a \in (A^p)^1_{\mathrm{sa}}$ we have

$$\begin{aligned}
\vartheta^p \circ \pi_V \circ \kappa^p\,(a) &= \vartheta^p(\textstyle\sum_{i=1}^{r} \langle x_i, \kappa^p(a)\rangle x_i) \\
= &\textstyle\sum_{j=1}^{s} \big\langle \sum_{k=1}^{r} t_{k,j} x_k,\ \sum_{i=1}^{r} \langle x_i, \kappa^p(a)\rangle x_i - \frac{p}{\mathrm{tr}(p)} \big\rangle y_j \\
= &\textstyle\sum_{j=1}^{s} \big[\big\langle \sum_{k=1}^{r} t_{k,j} x_k, \kappa^p(a) \big\rangle - \big\langle \sum_{k=1}^{r} t_{k,j} x_k, \frac{p}{\mathrm{tr}(p)} \big\rangle \big] y_j \\
= &\textstyle\sum_{j=1}^{s} \big[\langle \kappa^p(y_j), \kappa^p(a)\rangle - \langle \kappa^p(y_j), \frac{p}{\mathrm{tr}(p)} \rangle \big] y_j \\
= &\textstyle\sum_{j=1}^{s} \langle \kappa^p(y_j), \kappa^p(a)\rangle y_j = \sum_{j=1}^{s} \langle y_j, a\rangle y_j = \pi_{\varsigma^p(V)}(a)
\end{aligned}$$

and we have proved that ϑ^p fits into the commuting diagram.

We discuss the metric behavior of the affine isomorphism $\vartheta^p : \mathrm{aff}(\mathbb{F}_V(p)) \to \varsigma^p(V)$. Since $\varsigma^p = (\kappa^p)^{-1} \circ \pi_{\mathrm{lin}(\mathbb{F}(p))}$ we can assume that $\dim\big(\pi_{\mathrm{lin}(\mathbb{F}(p))}(V)\big) \geq 1$. Otherwise ϑ^p is the isometry $\{\pi_V(\frac{p}{\mathrm{tr}(p)})\} \to \{0\}$. In the finite dimensional Euclidean space A_{sa} (or $(A^p)_{\mathrm{sa}}$) we have for a vector x and a linear subspace U of dimension $\dim(U) \geq 1$ the equation

$$\|\pi_U(x)\|_2 = \max_{u \in U, \|u\|_2 = 1} |\langle u, x\rangle|.$$

This follows from the Schwarz inequality (2.3). Moreover, for a normalized vector $u \in U$ with $\pi_U(x) \in \mathbb{R}u$ we have $\|\pi_U(x)\|_2 = |\langle u, x\rangle|$.

Let us prove the expanding property of ϑ^p on two points in $\mathrm{aff}(\mathbb{F}_V(p))$. Let $a, b \in (A^p)^1_{\mathrm{sa}}$, put $a' := \kappa^p(a)$, $b' := \kappa^p(b) \in \mathrm{aff}(\mathbb{F}(p))$ and consider the points

$$\pi_V(a'), \pi_V(b') \in \mathrm{aff}(\mathbb{F}_V(p)).$$

By the commuting diagram and using the isometry κ^p one has

$$\begin{aligned}
&\|\vartheta^p(\pi_V(b')) - \vartheta^p(\pi_V(a'))\|_2 = \|\pi_{\varsigma^p(V)}(b) - \pi_{\varsigma^p(V)}(a)\|_2 \\
= &\max_{\substack{w \in \varsigma^p(V) \\ \|w\|_2 = 1}} |\langle w, b - a\rangle| = \max_{\substack{w' \in \pi_{\mathrm{lin}(\mathbb{F}(p))}(V) \\ \|w'\|_2 = 1}} |\langle w', b' - a'\rangle|.
\end{aligned}$$

Using a normalized vector $v \in V$ with $\pi_V(b') - \pi_V(a') \in \mathbb{R}v$ and $\widetilde{w} := \pi_{\mathrm{lin}(\mathbb{F}(p))}(v)$ one has $\|\widetilde{w}\|_2 \leq 1$. Since $b' - a' \in \mathrm{lin}(\mathbb{F}(p))$ this implies

$$\max_{\substack{w' \in \pi_{\mathrm{lin}(\mathbb{F}(p))}(V) \\ \|w'\|_2 = 1}} |\langle w', b' - a'\rangle| \geq |\langle \widetilde{w}, b' - a'\rangle| = |\langle v, b' - a'\rangle| = \|\pi_V(b') - \pi_V(a')\|_2. \tag{10.7}$$

This shows that ϑ^p is expanding.

We discuss when ϑ^p is an isometry. if $\pi_{\mathrm{lin}(\mathbb{F}(p))}(V) \subset V$ then we find in (10.7) the converse inequality

$$\max_{\substack{w' \in \pi_{\mathrm{lin}(\mathbb{F}(p))}(V) \\ \|w'\|_2 = 1}} |\langle w', b' - a' \rangle| \leq \max_{\substack{v \in V \\ \|v\|_2 = 1}} |\langle v, b' - a' \rangle| = \|\pi_V(b') - \pi_V(a')\|_2.$$

Conversely, notice for a vector $v \in V$ that

- $v \in \mathrm{lin}(\mathbb{F}(p)) \quad \Longrightarrow \quad \pi_{\mathrm{lin}(\mathbb{F}(p))}(v) = v \in V,$
- $v \perp \mathrm{lin}(\mathbb{F}(p)) \quad \Longrightarrow \quad \pi_{\mathrm{lin}(\mathbb{F}(p))}(v) = 0 \in V.$

If $\pi_{\mathrm{lin}(\mathbb{F}(p))}(V) \subsetneq V$ then there exists a normalized vector $v \in V$ such that $v = \widetilde{w} + w$ for non-zero $\widetilde{w} \in \mathrm{lin}(\mathbb{F}(p))$ and non-zero $w \in \mathrm{lin}(\mathbb{F}(p))^\perp$. We can choose two distinct points $a', b' \in \mathrm{aff}(\mathbb{F}(p))$ such that the difference $b' - a'$ is collinear with $\widetilde{w}$. The modulus of $\widetilde{w} = \pi_{\mathrm{lin}(\mathbb{F}(p))}(v)$ is strictly smaller than one, $\|\widetilde{w}\|_2 < \|v\|_2 = 1$. Hence we obtain a strict inequality in (10.7). `qed`

Proof of Corollary 5.23. If $p \leq s(F)$ then by the isomorphism between projector lattice and face lattice of the state space (Corollary 4.20) we have $\mathbb{F}(p) \subset \mathbb{F}(s(F)) = F$ and this shows $\mathrm{lin}(\mathbb{F}(p)) \subset \mathrm{lin}(F)$. Then

$$\varsigma^p(\mathrm{lin}(F)) = (\kappa^p)^{-1} \circ \pi_{\mathrm{lin}(\mathbb{F}(p))}(\mathrm{lin}(F)) = (\kappa^p)^{-1}(\mathrm{lin}(\mathbb{F}(p))) = (A^p)^0_{\mathrm{sa}}$$

by (4.37). If $a \in (A^p)^1_{\mathrm{sa}}$ then by (5.3)

$$\pi_{\varsigma^p(\mathrm{lin}(F))}(a) = \pi_{(A^p)^0_{\mathrm{sa}}}(a) = a - \frac{\mathbb{1}^p}{\mathrm{tr}(\mathbb{1}^p)}$$

holds. Since $p \leq s(F)$ we get

$$\pi_{\mathrm{lin}(F)}(\kappa^p(a)) = s(F)\kappa^p(a)s(F) - \mathrm{tr}(s(F)\kappa^p(a))\frac{s(F)}{\mathrm{tr}(s(F))} = \kappa^p(a) - \frac{s(F)}{\mathrm{tr}(s(F))}.$$

Using the previous two equations and the relation $(\vartheta^p)^{-1} \circ \pi_{\varsigma^p(\mathrm{lin}(F))} = \pi_{\mathrm{lin}(F)} \circ \kappa^p$ from Proposition 5.20 we get

$$(\vartheta^p)^{-1}\left(a - \frac{\mathbb{1}^p}{\mathrm{tr}(\mathbb{1}^p)}\right) = (\vartheta^p)^{-1}\left(\pi_{\varsigma^p(\mathrm{lin}(F))}(a)\right) = \pi_{\mathrm{lin}(F)}(\kappa^p(a)) = \kappa^p(a) - \frac{s(F)}{\mathrm{tr}(s(F))}.$$

Since $\pi_{\varsigma^p(\mathrm{lin}(F))} : (A^p)^1_{\mathrm{sa}} \to \varsigma^p(\mathrm{lin}(F))$ is onto, the claim follows. `qed`

Proof of Corollary 5.25. Let $p \in \mathcal{P}\backslash\{0\}$, $q \in \mathcal{P}$ such that $q \leq p$ and put $r := (\kappa^p)^{-1}(q)$. By definition of a face reflection (5.13) and by composition law for *-isomorphisms (5.59) we have $\mathbb{F}_V(q) = \pi_V(\mathbb{F}(q)) = \pi_V \circ \kappa^p(\mathbb{F}(r))$. An application of Proposition 5.20 gives

$$\mathbb{F}_V(q) = (\vartheta^p)^{-1} \circ \pi_{\varsigma^p(V)}(\mathbb{F}(r)) = (\vartheta^p)^{-1}\left(\mathbb{F}_{\varsigma^p(V)}(r)\right)$$

and we have $\mathbb{F}_V(p) = (\vartheta^p)^{-1}(\mathrm{sr}_{\varsigma^p(V)})$. The face lattices of $\mathbb{F}_V(p)$ and $\mathrm{sr}_{\varsigma^p(V)}$ are isomorphic under ϑ^p, because this is an affine isomorphism preserving the convex structure. qed

Proof of Proposition 5.27. Let $p \in \mathcal{P} \setminus \{0\}$ be an orthogonal projector. The translation vector space of $\mathbb{F}(p)$ is $\mathrm{lin}(\mathbb{F}(p)) = \kappa^p((A^p)^0_{\mathrm{sa}})$ by (4.37). Both sides of the equation in the first statement are included in $\kappa^p((A^p)_{\mathrm{sa}})$, which we decompose into

$$(A^p)_{\mathrm{sa}} = (A^p)^0_{\mathrm{sa}} + \mathbb{R}\mathbb{1}^p = (A^p)^1_{\mathrm{sa}} - (A^p)^1_{\mathrm{sa}} + \mathbb{R}\mathbb{1}^p.$$

Let $x = a - b + \lambda\mathbb{1}^p$ be an element of $(A^p)_{\mathrm{sa}}$ for $a, b \in (A^p)^1_{\mathrm{sa}}$ and $\lambda \in \mathbb{R}$. Then we have $\kappa^p(x) = \kappa^p(a-b) + \lambda p$ and get

$$\begin{aligned}
& x \in \varsigma^p(V)^\perp \iff a - b \perp \varsigma^p(V) \iff \pi_{\varsigma^p(V)}(a-b) = 0 \\
\iff\ & \vartheta^p \circ \pi_V \circ \kappa^p(a) = \vartheta^p \circ \pi_V \circ \kappa^p(b) \\
\iff\ & \pi_V(\kappa^p(a-b)) = 0 \\
\iff\ & \kappa^p(a-b) \perp V \\
\iff\ & \kappa^p(x) \in (V^\perp \cap \mathrm{lin}(\mathbb{F}(p))) + \mathbb{R}p.
\end{aligned}$$

In the second line of the equivalence we use Proposition 5.20.

For the second assertion we choose a subset $M \subset \mathbb{F}(p)$. Since $p \in \mathcal{P}_V$ we have by (5.63) $(\mathbb{F}(p) + V^\perp) \cap \overline{S}(A) = \mathbb{F}(p)$. Since $M \subset \mathbb{F}(p)$, the intersection of this equation with $M + V^\perp$ gives

$$(M + V^\perp) \cap \overline{S}(A) = (M + V^\perp) \cap \mathbb{F}(p).$$

We modify only the right hand side of the last equation. Since $\mathbb{F}(p) = \left(\frac{p}{\mathrm{tr}(p)} + \mathrm{lin}(\mathbb{F}(p))\right) \cap \mathbb{F}(p)$ and since $M - \frac{p}{\mathrm{tr}(p)} \subset \mathrm{lin}(\mathbb{F}(p))$, we get by the modular law for affine spaces the following.

$$\begin{aligned}
(M + V^\perp) \cap \mathbb{F}(p) &= \left(\left(M - \tfrac{p}{\mathrm{tr}(p)}\right) + \left(V^\perp + \tfrac{p}{\mathrm{tr}(p)}\right)\right) \cap \left(\tfrac{p}{\mathrm{tr}(p)} + \mathrm{lin}(\mathbb{F}(p))\right) \cap \mathbb{F}(p) \\
&= \left[\left(M - \tfrac{p}{\mathrm{tr}(p)}\right) + \left(\left(V^\perp + \tfrac{p}{\mathrm{tr}(p)}\right) \cap \left(\tfrac{p}{\mathrm{tr}(p)} + \mathrm{lin}(\mathbb{F}(p))\right)\right)\right] \cap \mathbb{F}(p) \\
&= \left[M + \left(V^\perp \cap \mathrm{lin}(\mathbb{F}(p))\right)\right] \cap \mathbb{F}(p) \\
&= \left[M + \left(V^\perp \cap \mathrm{lin}(\mathbb{F}(p))\right) + \mathbb{R}p\right] \cap \mathbb{F}(p) \\
&= \left[M + \kappa^p\left(\varsigma^p(V)^\perp\right)\right] \cap \mathbb{F}(p).
\end{aligned}$$

In the penultimate equality we have used a simple trace comparison. In the last equality we have used the first assertion of this lemma. qed

Proof of Corollary 5.28. Let $p \in \mathcal{P}_V \setminus \{0\}$ choose $q \in \mathcal{P}$ with $q \leq p$ and put $r := (\kappa^p)^{-1}(q)$. Since $\mathbb{F}(q) \subset \mathbb{F}(p)$ we can apply Proposition 5.27, second assertion. With (5.59) follows

$$\left(\mathbb{F}(q) + V^\perp\right) \cap \overline{S}(A) = \left[\mathbb{F}(q) + \kappa^p\left(\varsigma^p(V)^\perp\right)\right] \cap \mathbb{F}(p) = \kappa^p\left[\left(\mathbb{F}(r) + \varsigma^p(V)^\perp\right) \cap \overline{S}(A^p)\right].$$

The proof is completed by (5.63). qed

Proof of Lemma 5.30. We consider the smallest exposed face (3.46) containing F,

$$\overset{\perp}{F} = \bigcap\{G \in \mathcal{F}_\perp(C) : F \subset G\}.$$

This intersection is an exposed face of C because the exposed face lattice (3.35) is complete. Moreover, the normal cones are equal (3.48)

$$\mathrm{N}(C, F) = \mathrm{N}(C, \overset{\perp}{F}).$$

Notice that $\overset{\perp}{F} \subsetneq C$. Otherwise $\mathrm{N}(C, F) = \mathrm{N}(C, C) = \mathrm{lin}(C)^\perp$ implies $F = C$ by Lemma 3.20. This case is excluded from the definition. Inductively we argue with the pair $(F, \overset{\perp}{F})$ in place of (F, C) and so forth. Since C is finite dimensional, the induction will stop.

If $F \subset G$ is an ordered pair of proper faces of C then by Remark 3.9 (b) the face F is a face of G. Hence a concatenation of access sequences is an access sequence. qed

Proof of Theorem 2. If $(F_1, \ldots, F_m)$ is a sequence of faces of the state reflection sr_V such that $F_1 \in \mathcal{F}_{V,\perp}$ and such that $F_i \in \mathcal{F}_\perp(F_{i-1})$ for $i = 2, \ldots, m$ then F_i is a face of sr_V for $i = 1, \ldots, m$. This follows from transitivity of the face property under inclusion, see Remark 3.9 (b). Thus the condition $F_i \in \mathcal{F}_V$ is redundant in the definition of an access sequence of faces and we will drop it for the first direction of the proof.

Let $(p_1, \ldots, p_m)$ be an access sequence of projectors for sr_V. Then by definition the projector p_1 is a proper exposed projector of the state reflection and we get from (5.15) that the face $\mathbb{F}_V(p_1)$ is a proper exposed face of sr_V. If $p \gneq q$ are two successive elements of the access sequence $(p_1, \ldots, p_m)$ then by definition of an access sequence of projectors the projector $r := (\kappa^p)^{-1}(q)$ satisfies $r \in \mathcal{P}_{\varsigma^p(V),\perp}$. This implies $\mathbb{F}_{\varsigma^p(V)}(r) \in \mathcal{F}_{\varsigma^p(V),\perp}$. Then by Corollary 5.25 the face $\mathbb{F}_V(q)$ is a proper exposed face of $\mathbb{F}_V(p)$. This completes the first direction of the proof.

Let $(F_1, \ldots, F_m)$ be an access sequence of faces for sr_V. Then $s_V(F_1) \in \mathcal{P}_{V,\perp}$ is a proper exposed projector by (5.15). For two successive elements $F \supsetneq G$ in the access sequence $(F_1, \ldots, F_m)$ we set $p := s_V(F)$ and $q := s_V(G)$. Then by the lattice isomorphism $s_V : \mathcal{F}_V \to \mathcal{P}_V$ (see Remark 5.11) the projectors p, q belong to $\mathcal{P}_V$ and we have $p \gneq q$. Thus for $r := (\kappa^p)^{-1}(q)$ we have $r \in \mathcal{P}_{\varsigma^p(V)}$ by Corollary 5.28. In particular, for the support projector $s_{\varsigma^p(V)}(\mathbb{F}_{\varsigma^p(V)}(r)) = r$ holds. In addition, we have by the definition of an access sequence of faces that $\mathbb{F}_V(q) \in \mathcal{F}_\perp(\mathbb{F}_V(p))$ and from this we get from Corollary 5.25 that $\mathbb{F}_{\varsigma^p(V)}(r) \in \mathcal{F}_{\varsigma^p(V),\perp}$. So $r = s_{\varsigma^p(V)}(\mathbb{F}_{\varsigma^p(V)}(r)) \in \mathcal{P}_{\varsigma^p(V),\perp}$ and we conclude with $(\kappa^p)^{-1}(q) \in \mathcal{P}_{\varsigma^p(V),\perp}$ as desired. `qed`

Proof of Corollary 5.31. This follows from Lemma 5.30 and Theorem 2. `qed`

Proof of Corollary 5.32. An ordered pair $p \leq q$ of proper projectors in $\mathcal{P}_V$ corresponds to the ordered pair $\mathbb{F}_V(p) \subset \mathbb{F}_V(q)$ of proper faces of sr_V by the lattice isomorphism $\mathcal{P}_V \to \mathcal{F}_V$ (5.54). By Lemma 5.30 there exists an access sequence of faces for sr_V including both $\mathbb{F}_V(p)$ and $\mathbb{F}_V(q)$. Theorem 2 proves that the corresponding access sequence of projectors for sr_V includes p and q. The case $q \leq p$ is analogue. `qed`

Proof of Lemma 5.37. The dimension of the face reflection for a projector $p \in \mathcal{P}$ is $\dim(\mathbb{F}_V(p)) = \dim(\pi_{\mathrm{lin}(\mathbb{F}(p))}(V)) = \mathrm{rk}(\pi_{\mathrm{lin}(\mathbb{F}(p))}|_V)$ by (5.61). We use the operator norm on the space of linear mappings $V \to A^0_{\mathrm{sa}}$ and prove below that the mapping

$$p \mapsto \pi_{\mathrm{lin}(\mathbb{F}(p))}|_V$$

is continuous. By equivalence of norms on finite-dimensional vector spaces, the continuity remains true in any matrix representation of the space in question. Thus it is sufficient to have the rank function lower semi-continuous on a matrix space, which is proved in Lemma 4.36.

Let $p, q \in \mathcal{P}$ with spectral norm bound $\|p-q\| < 1$. By Lemma 2.38 we have $\mathrm{rk}(p) = \mathrm{rk}(q)$ so $\mathrm{tr}(p) = \mathrm{tr}(q)$ which we assume positive. By the explicit form (5.3) of the orthogonal projection to the translation vector space of a face and by the inequality (2.12) we have for all $a \in A$ the estimate in trace norm

$$\begin{aligned}
&\sup_{\|a\|_1=1} \|\pi_{\mathrm{lin}(\mathbb{F}(p))}(a) - \pi_{\mathrm{lin}(\mathbb{F}(q))}(a)\|_1 = \sup_{\|a\|_1=1} \|pap - \mathrm{tr}(pa)\tfrac{p}{\mathrm{tr}(p)} - qaq + \mathrm{tr}(qa)\tfrac{q}{\mathrm{tr}(q)}\|_1 \\
&\leq \sup_{\|a\|_1=1} \big(\|pap - qap\|_1 + \|qap - qaq\|_1 + \|\mathrm{tr}(pa)\tfrac{p}{\mathrm{tr}(p)} - \mathrm{tr}(pa)\tfrac{q}{\mathrm{tr}(q)}\|_1 \\
&\qquad\qquad + \|\mathrm{tr}(pa)\tfrac{q}{\mathrm{tr}(q)} - \mathrm{tr}(qa)\tfrac{q}{\mathrm{tr}(q)}\|_1\big) \\
&\leq \sup_{\|a\|_1=1} \big(\|p-q\|_1\|p\|_1 + \|q\|_1\|p-q\|_1 + \tfrac{1}{\|p\|_1}\|p-q\|_1 + \|p-q\|_1\big) \\
&= (2\|p\|_1 + \tfrac{1}{\|p\|_1} + 1)\|p-q\|_1.
\end{aligned}$$

Moreover we have the global bound $\operatorname{tr}(p) \leq \operatorname{tr}(\mathbb{1}) < \infty$. Then we get for non-zero projectors p, q in spectral norm distance $\|p - q\| < 1$ the bound

$$2(\|\mathbb{1}\|_1 + 1)\|p - q\|_1$$

for the operator norm distance in trace norm between the mappings $\pi_{\operatorname{lin}(\mathbb{F}(p))}$ and $\pi_{\operatorname{lin}(\mathbb{F}(q))}$. The estimate improves under restriction to V. qed

Proof of Lemma 5.40. Let $(x_i) \subset \operatorname{sr}_V$ be a sequence with limit $x := \lim_{i\to\infty} x_i \in \operatorname{sr}_V$. As the inaugural proof argument we assume that the second alternative

$$\lim_{i\to\infty} \min\{\|x_i - y\|_2 : y \in \operatorname{rb}(F_V(x_i))\} \;=\; 0$$

is wrong. The minima exist because faces of the state reflection are compact and the same is true for their relative boundary. We deduce the first alternative $x \in \operatorname{ri}(\mathbb{F}_V(p))$ for some $p \in \overline{\{s(x_i) : i \in \mathbb{N}\}}$. After transition to a subsequence we fix a positive number $\epsilon > 0$ such that for $i \in \mathbb{N}$

$$B_\epsilon(x_i) \subset F_V(x_i)$$

with $B_\epsilon(x_i)$ a closed ball of radius ϵ about x_i and of dimension $\dim(F_V(x_i))$. By transition to a subsequence we assume the constant dimension $k := \dim(F_V(x_i))$ for $i \in \mathbb{N}$ and the convergence

$$p := \lim_{i\to\infty} s_V(x_i).$$

Assumption of convergence of the support projectors is possible since the projector lattice is compact by Lemma 2.38.

The main idea of the proof is that an accumulation point of the sequence of balls $B_\epsilon(x_i)$ is a k-dimensional ball of radius ϵ about x included in the face reflection $\mathbb{F}_V(p)$. On the other hand, lower semi-continuity in Lemma 5.37 implies

$$\dim(\mathbb{F}_V(p)) \leq \liminf_{i\to\infty} \dim(\mathbb{F}_V(s_V(x_i))) = k.$$

The two arguments combined give $x \in \operatorname{ri}(\mathbb{F}_V(p))$ and complete the proof.

Clear is that an accumulation point of the balls $B_\epsilon(x_i)$ will be a k-dimensional ball with radius ϵ about x. The remaining proof is about existence of a converging subsequence and about inclusion of the limit into the face reflection $\mathbb{F}_V(p)$. We make a detour about the lifted face lattice, where the homeomorphism $\mathcal{P} \to \mathcal{F}$ is available. First of all, to define an affine lifting map from sr_V to the state space for each ball, let us work with smaller balls of radius $\widetilde{\epsilon} := \frac{\epsilon}{k} > 0$ about x_i. Then by Remark 5.39 for each $i \in \mathbb{N}$ we can sandwich a regular k-dimensional simplex Z_i of edge length $\epsilon\sqrt{\frac{2(k+1)}{k}}$ between the boundary spheres of $B_{\widetilde{\epsilon}}(x_i)$ and $B_\epsilon(x_i)$. The result is

$$B_{\widetilde{\epsilon}}(x_i) \subset Z_i \subset B_\epsilon(x_i) \subset F_V(x_i) \in \mathcal{F}_V.$$

For an index $i \in \mathbb{N}$ we construct a lift $\xi_i : Z_i \to L_V(F_V(x_i))$ by restriction of an affine map. Let $\{e_j\}_{j=0}^k$ be the extreme point set of the simplex $Z_i \subset F_V(x_i)$ and choose arbitrary lifts $f_j \in L_V(F_V(x_i))$ such that $\pi_V(f_j) = e_j$ for $j = 0, \ldots, k$. The assignment $e_j \mapsto f_j$ for $j = 0, \ldots, k$ extends to an injective affine mapping $\mathrm{aff}(F_V(x_i)) \to \mathrm{aff}(L_V(F_V(x_i)))$ and we define ξ_i as the restriction of this mapping to Z_i. As a consequence, for $i \in \mathbb{N}$ a lift ξ_i is continuous and the lifted ball

$$L_i := \xi_i(B_{\widetilde{\epsilon}}(x_i))$$

is a compact set, indeed an ellipse. Moreover, since the lifted face $L_V(F_V(x_i))$ is convex we have

$$L_i = \xi_i(B_{\widetilde{\epsilon}}(x_i)) \subset \xi_i(Z_i) = \mathrm{conv}(\{f_j\}_{j=1}^k) \subset L_V(F_V(x_i)) = \mathbb{F}(s_V(x_i)).$$

Let us find a subsequence of $B_{\widetilde{\epsilon}}(x_i)$ converging to a subset of $\mathbb{F}_V(p)$. Since the state space $\overline{S}(A)$ is bounded, we can use Blaschke selection theorem in [Sch], Theorem 1.8.6. This says that the sequence L_i has a converging subsequence. After transition to a converging subsequence we put

$$L := \lim_{i\to\infty} L_i.$$

From the homeomorphism $\mathcal{P} \to \mathcal{F}$ in Proposition 4.32 follows $\lim_{i\to\infty} \mathbb{F}(s_V(x_i)) = \mathbb{F}(p)$ because $p = \lim_{i\to\infty} s_V(x_i)$. Since $L_i \subset \mathbb{F}(s_V(x_i))$ for $i \in \mathbb{N}$ we get from the criteria of Hausdorff convergence in Remark 4.31 the inclusion $L \subset \mathbb{F}(p)$, thus

$$\pi_V(L) \subset \mathbb{F}_V(p).$$

The following proof of $\lim_{i\to\infty} B_{\widetilde{\epsilon}}(x_i) = \pi_V(L)$ follows also from the Hausdorff convergence criteria. To verify criterion (a) let $y \in \pi_V(L)$ and let $z \in L$ with $y = \pi_V(z)$. From criterion (a) applied to the limit $\lim_{i\to\infty} L_i = L$ we obtain a sequence of points $z_i \in L_i$ for $i \in \mathbb{N}$ with $\lim_{i\to\infty} z_i = z$. Then for $y_i := \pi_V(z_i) \in B_{\widetilde{\epsilon}}(x_i)$ we have for $i \in \mathbb{N}$

$$\lim_{i\to\infty} y_i = \lim_{i\to\infty} \pi_V \circ \xi_i(y_i) = \pi_V\left(\lim_{i\to\infty} z_i\right) = \pi_V(z) = y.$$

To verify the convergence criterion (b) let $y_{i_j} \in B_{\widetilde{\epsilon}}(x_{i_j})$ be a convergent sequence for $j \in \mathbb{N}$. Then

$$\lim_{j\to\infty} y_{i_j} = \lim_{j\to\infty} \pi_V \circ \xi_{i_j}(y_{i_j}) = \pi_V\left(\lim_{j\to\infty} \xi_{i_j}(y_{i_j})\right).$$

Since the state space is compact, we can select a convergent subsequence of $\xi_{i_j}(y_{i_j})$. This sequence converges to a point in L by criterion (b) applied to the limit $\lim_{i\to\infty} L_i = L$. Hence $\lim_{j\to\infty} y_{i_j} \in \pi_V(L)$ as demanded. qed

Proof of Theorem 3. Suppose x is an extreme point of the state reflection sr_V. By Straszevicz's theorem there exists a sequence (x_i) of exposed points of sr_V such that $x = \lim_{i\to\infty} x_i$. Consider a limit point p of the sequence $s_V(x_i) \in \mathcal{P}_{V,\perp}$ of exposed support projectors. Under transition to a subsequence $s_V(x_i)$ with limit p we get

$$\pi_V\left(\frac{p}{\mathrm{tr}(p)}\right) = \lim_{i\to\infty} \pi_V\left(\frac{s_V(x_i)}{\mathrm{tr}(s_V(x_i))}\right) = \lim_{i\to\infty} x_i = x.$$

Since x is an extreme point of the state reflection, we get $\mathbb{F}_V(p) = \{x\}$.

Assume that x belongs to the d-skeleton of the state reflection sr_V. By Asplund's theorem there exists a sequence (x_i) of points each one contained in some exposed face $F_i \in \mathcal{F}_{V,\perp}$ of dimension at most d and such that $x = \lim_{i\to\infty} x_i$. Without changing the limit x of the sequence (x_i) we can choose the point x_i in the relative interior $\mathrm{ri}(F_i)$. Then we get for $i \in \mathbb{N}$ the equalities

$$s_V(x_i) = s_V(F_i) \in \mathcal{P}_{V,\perp}$$

for the support projectors. In particular, they are in the exposed projector lattice. The projector lattice $\mathcal{P}$ is compact by Lemma 2.38 so we have by transition to a subsequence the convergence

$$p := \lim_{i\to\infty} s_V(x_i)$$

to a projector $p \in \overline{\mathcal{P}_{V,\perp}}$. If $x \in \mathrm{ri}(\mathbb{F}_V(p))$ or if $x \in \mathrm{ri}(\mathbb{F}_V(s_V(x_i)))$ for some $i \in \mathbb{N}$ then the proof is complete. Otherwise Lemma 5.40 shows that the point x is the limit of a sequence (y_i) of relative boundary points $y_i \in \mathrm{rb}(F_i)$ for $i \in \mathbb{N}$. Clearly, the dimension of a face $F_V(y_i)$ is strictly less than $\dim(F_i) \leq d$. Then the assumed lower semi-continuity of the dimension function gives

$$\dim(F_V(x)) \leq \liminf_{i\to\infty} \dim(F_V(y_i)) < d$$

and we see that x belongs to the $(d-1)$-skeleton. Inductively we put the problem down to the extreme point case treated in the beginning. qed

Proof of Corollary 5.42. One has the disjoint cover (5.49) $\mathrm{sr}_V = \bigcup_{p\in\mathcal{P}_V\setminus\{0\}} \mathrm{ri}(\mathbb{F}_V(p))$. If the projector lattice $\mathcal{P}_V$ is closed then $\overline{\mathcal{P}_{V,\perp}} \subset \mathcal{P}_V$. Assuming stability (Remark 5.47 (a)) of the state reflection sr_V one has by Theorem 3 the cover $\mathrm{sr}_V = \bigcup_{p\in\overline{\mathcal{P}_{V,\perp}}} \mathrm{ri}(\mathbb{F}_V(p))$. qed

Proof of Lemma 5.48. The idea for the first part is to symmetrize an open set in C by the midpoint map before doing the projection. Let $U_1 \subset C$ be open and convex. By symmetry, the reflection $U_2 := r_V(U_1)$ is an open and convex subset of C. Since C is stable the set $\frac{1}{2}(U_1 + U_2)$ is open. Thus the intersection $\frac{1}{2}(U_1 + U_2) \cap V$ is open in $C \cap V$. By symmetry at V and by convexity of the sets $\frac{1}{2}(U_1 + U_2)$ and C we have proved that $\pi_V(\frac{1}{2}(U_1 + U_2)) = \frac{1}{2}(U_1 + U_2) \cap V$ is open in $\pi_V(C) = C \cap V$.

The observation $\pi_V(U_1) = \pi_V(\frac{1}{2}(U_1 + U_2))$ finishes the proof of the open projection. Indeed, given $v_1 \in V$ and $w_1 \in V^\perp$ such that $v_1 + w_1 \in U_1$ we have $v_1 - w_1 \in U_2$. Hence

$$\pi_V(v_1 + w_1) = \pi_V(\frac{1}{2}(v_1 + w_1 + v_1 - w_1)) \subset \pi_V(\frac{1}{2}(U_1 + U_2)).$$

Conversely, for additional points $v_2 \in V$ and $w_2 \in V^\perp$ such that $v_2 + w_2 \in U_2$ we have $v_2 - w_2 \in U_1$ and get

$$\pi_V(\frac{1}{2}(v_1 + w_1 + v_2 + w_2)) = \pi_V(\frac{1}{2}(v_1 + w_1 + v_2 - w_2)) \subset \pi_V(U_1)$$

because U_1 is convex.

We prove that the image $\pi_V(C)$ is stable. Consider open sets $U_1 \cap \pi_V(C)$ and $U_2 \cap \pi_V(C)$ of $\pi_V(C)$ for $U_1, U_2 \subset V$ open. By the modular law for cylinders we have $(i = 1, 2)$

$$U_i \cap \pi_V(C) = \pi_V((U_i + V^\perp) \cap C)$$

and then

$$\frac{1}{2}(U_1 \cap \pi_V(C) + U_2 \cap \pi_V(C)) = \pi_V\left(\frac{1}{2}\left((U_1 + V^\perp) \cap C + (U_2 + V^\perp) \cap C\right)\right).$$

The argument of π_V on the right-hand side is open because C is stable. Since π_V is an open mapping, the left-hand side is open. The open set U_1 and U_2 in $\pi_V(C)$ were arbitrary, hence the midpoint map for $\pi_V(C)$ is open and the set is stable. qed

Proof of Proposition 5.52. Let $C \subset \mathbb{R}^m$ be a convex set and $H \subset \mathbb{R}^m$ a linear hyperplane. Let $s : C \to s(C)$ be a homeomorphic symmetrization map for C at H. The following proof is a generalization of the proof for Lemma 5.48 and the main idea is to symmetrize an open set of $s(C)$ by the reflection r_H.

Let $U_1 \subset C$ be an open set. Then $V_1 := s(U_1)$ is open in $s(C)$ because s is a homeomorphism. By symmetry of $s(C)$ the set $V_2 := r_H(V_1)$ is open in $s(C)$ and $U_2 := s^{-1}(V_2)$ is open in C. As in Lemma 5.48 we have

$$\pi_H(U_1) = \frac{1}{2}(V_1 + V_2) \cap H,$$

in contrast to the lemma the symmetric set $\frac{1}{2}(V_1+V_2)$ may not be open and the open set $\frac{1}{2}(U_1+U_2)$ may not be symmetric at H.

Let us use a special open base for the topology of $s(C)$ and for the topology of C the open base corresponding under the homeomorphism s^{-1}. We assume that $V_1=(B+B_\perp)\cap s(C)$ where $B\subset H$ and $B_\perp\subset H^\perp$ are convex open sets. Then $V_2=(B-B_\perp)\cap s(C)$ and we prove under these assumptions that $\frac{1}{2}(V_1+V_2)\cap H$ is covered by $s\left(\mathrm{conv}(U_1,U_2)\right)$. Then the equality $\pi_H\left(s\left(\mathrm{conv}(U_1,U_2)\right)\right)=\pi_H(U_1)$ shows

$$\pi_H(U_1)=s\left(\mathrm{conv}(U_1,U_2)\right)\cap H.$$

The proof is completed by the observation that the convex hull $\mathrm{conv}(U_1,U_2)$ is open in C because C is stable [Pa].

We discuss the convex symmetric set $\frac{1}{2}(V_1+V_2)\cap H$ in relation to the open set $s\left(\mathrm{conv}(U_1,U_2)\right)$. Notice that

$$\begin{aligned}\tfrac{1}{2}(V_1+V_2)\cap H &= \tfrac{1}{2}\left[((B+B_\perp)\cap s(C))+((B-B_\perp)\cap s(C))\right]\cap H\\ &= \bigcup_{b\in B_\perp}\tfrac{1}{2}\left((B+b)\cap s(C)+(B-b)\cap s(C)\right)\\ &= \bigcup_{b\in B_\perp}((B+b)\cap s(C))-b.\end{aligned}\tag{10.8}$$

In addition, notice the monotonicity for $b\in H^\perp$ and $\lambda\in[0,1]$. One has

$$((B+b)\cap s(C))-b\quad\subset\quad((B+\lambda b)\cap s(C))-\lambda b.$$

The monotonicity follows from symmetry at H of $s(C)$ and from convexity of $s(C)$. In the case $0\in B_\perp$ we find

$$\frac{1}{2}(V_1+V_2)\cap H\quad=\quad B\cap s(C)\quad\subset\quad V_1\quad=\quad s(U_1)\quad\subset\quad s\left(\mathrm{conv}(U_1,U_2)\right).$$

Now we have to restrict to the hyperplane case. If $0\notin B_\perp$ then by (10.8) for a point $v\in\frac{1}{2}(V_1+V_2)\cap H$ there exists a point $b\in B_\perp\setminus\{0\}$ such that $v+b\in V_1$ and $v-b\in V_2$. We put $x_1:=s^{-1}(v+b)\in U_1$ and $x_2:=s^{-1}(v-b)\in U_2$ and define the continuous function

$$\psi: C\to\mathbb{R},\qquad x\mapsto\langle b,s(x)\rangle$$

with $\psi(x_1)=\|b\|_2^2>0$ and $\psi(x_2)=-\|b\|_2^2<0$. There exists a point $\widetilde{x}\in[x_1,x_2]\subset\mathrm{conv}(U_1,U_2)$ such that $\psi(\widetilde{x})=0$. We have $s(\widetilde{x})=v$ because the fiber $\pi_H^{-1}(v)=v+H^\perp$ is one-dimensional.

Finally the proof that the image $\pi_H(C)$ is stable is the same as in Lemma 5.48. qed

Proof of Corollary 5.53. By Proposition 4.35 the skeletons of the state space $\overline{S}(A)$ are all compact. Thus, by Remark 5.47 (a) the state space is stable. Moreover, the state space is compact so the symmetrization at H is a homeomorphism. Then Proposition 5.52 shows that the state reflection $\mathrm{sr}_H = \pi_H(\overline{S}(A))$ is stable. qed

Proof of Lemma 6.22. Let $a \in \mathrm{Cyl} = \mathcal{E} + U^\perp$ and put $M := S(A) \cap (a + U^\perp)$. We want to prove that $\mathcal{E} \cap M$ consists of a single point. Since $\mathbb{1} \in U^\perp$ we can assume without loss of generality that $\mathrm{tr}(a) = 1$. Then we have

$$M = S(A) \cap (a + U^{\perp,0}) \qquad \text{and} \qquad a \in \mathcal{E} + U^{\perp,0}.$$

For abbreviation we put $\mathbb{A} := a + U^{\perp,0}$.

By definition we have $M \neq \emptyset$. The convex manifold $S(A)$ is the relative interior of the state space $\overline{S}(A)$, so the affine space $\mathbb{A}$ and $S(A)$ share a relative interior point. Thus by (3.18) we have $\mathrm{aff}(M) = \mathrm{aff}(S(A)) \cap \mathbb{A}$. The affine hull of $S(A)$ is the affine hull of the state space (4.25), $\mathrm{aff}(S(A)) = A^1_{\mathrm{sa}}$. So $\mathrm{aff}(M) = \mathbb{A}$ and we see that $M = S(A) \cap \mathrm{aff}(M)$ is a convex sub-manifold of $S(A)$ with translation vector space $\mathrm{lin}(M) = U^{\perp,0}$. In particular for any $\sigma \in \mathcal{E} \cap M$ the m-representation of the tangent space $\mathrm{T}_\sigma M$ is $\mathrm{T}^{(m)}_\sigma M = \mathrm{lin}(M) = U^{\perp,0}$. Let $\theta := \ln_0(\sigma) \in \Theta$. The canonical representation of the tangent space $\mathrm{T}_\sigma \mathcal{E}$ is $\mathrm{T}^{(\Theta)}_\theta \mathcal{E} = U$. The BKM-metric (6.23) evaluated under two vectors $u \in \mathrm{T}_\sigma \mathcal{E}$ and $v \in \mathrm{T}_\sigma M$ is (6.11) $\langle u, v \rangle_\sigma = \langle u^{(\Theta)}, v^{(m)} \rangle = 0$ and the dimension equation

$$\dim(\mathrm{T}_\sigma S(A)) = \dim(A^0_{\mathrm{sa}}) = \dim(U) + \dim(U^{\perp,0}) = \dim(\mathrm{T}_\sigma \mathcal{E}) + \dim(\mathrm{T}_\sigma M)$$

completes the proof of the orthogonal direct sum with respect to the BKM-metric

$$\mathrm{T}_\sigma S(A) = \mathrm{T}_\sigma \mathcal{E} \oplus \mathrm{T}_\sigma M. \tag{10.9}$$

Now we prove that the intersection $\mathcal{E} \cap M$ consists of only one point. Since $a \in \mathbb{A} = \mathrm{aff}(M) \subset \mathrm{aff}(S(A))$ and since σ belongs to the relative open set $S(A)$, there is $\epsilon > 0$ such that $\rho := \sigma + \epsilon(a - \sigma) \in \overline{S}(A)$. In particular, if $a \in \overline{S}(A)$ we may choose $\epsilon = 1$ and keep $\rho = a$. The m-geodesic (6.24) $\gamma^{(m)}_{\sigma\rho}$ and the e-geodesic (6.25) $\gamma^{(e)}_{\sigma\tau}$ for arbitrary $\tau \in \mathcal{E}$ meet at $\sigma \in \mathcal{E}$ orthogonally with respect to BKM-metric (10.9). If $\tau \neq \sigma$ then we obtain by the distance like properties (6.19) and the Pythagorean theorem of relative entropy (6.26) the estimate

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

This proves that S_ρ has a unique minimum at σ. In particular, the intersection $\mathcal{E} \cap M$ contains a unique point because $\sigma \in \mathcal{E} \cap M$ was chosen arbitrary. qed

Proof of Lemma 6.25. Let $(x_i) \subset A_{\mathrm{sa}}$ be a sequence with $\lim_{i\to\infty}\|x_i\| = \infty$. We select a subsequence and put $u := \lim_{i\to\infty}\frac{x_i}{\|x_i\|}$. We put $s_i := x_i - \mu_+(x_i)\mathbb{1}$ with the maximal eigenvalue $\mu_+(x_i)$ of x_i for $i \in \mathbb{N}$. Let us first prove the support projector bound for an accumulation point of (e^{s_i}). For convenience we select a subsequence and put $z = \lim_{i\to\infty} e^{s_i}$. For $i \in \mathbb{N}$ we use the perturbation $t_i := \frac{x_i}{\|x_i\|}$ of u where

$$s_i = \|x_i\|(t_i - \mu_+(t_i)\mathbb{1}).$$

Consider for $i \in \mathbb{N}$ the total projectors of t_i which are for $\mu \in \mathrm{spec}(u)$ given by $q_\mu(i) := -\frac{1}{2\pi i}\int_{\gamma_\mu}(t_i - \zeta)^{-1}\,\mathrm{d}\,\zeta$. Here we have used a positive-oriented curve γ_μ in the resolvent set of u enclosing μ and no other eigenvalues of u. With the spectral projector $p_\mu(u)$ of u for an eigenvalue $\mu \in \mathrm{spec}(u)$ one has [KaT]

$$\lim_{i\to\infty} q_\mu(i) = p_\mu(u).$$

For a fixed spectral value $\mu \in \mathrm{spec}(u)$ we define for $i \in \mathbb{N}$ a sequence of hermitean unitaries $v_i := \mathrm{sgn}(\mathbb{1} - q_\mu(i) - p_\mu(u))$ reflecting the projectors on each other, $p_\mu(u) = vq_\mu(i)v$ [Av] and such that (2.50)

$$\lim_{i\to\infty} v_i = \mathbb{1} - 2p_\mu(u).$$

The eigenvalues of $t_i|_{\mathrm{Im}(q_\mu(i))}$ converge to μ for $i \to \infty$, hence $\frac{s_i}{\|x_i\|} = t_i - \mu_+(t_i)\mathbb{1}$ behaves as follows. The eigenvalues of $\frac{s_i}{\|x_i\|}|_{\mathrm{Im}(q_\mu(i))}$ converge to $\mu - \mu_+(u)$ for $i \to \infty$ and $\frac{s_i}{\|x_i\|}q_\mu(i)$ is zero on the complement $\ker(q_\mu(i))$. We assume $\mu \neq \mu_+(u)$ is not the largest eigenvalue of u, then $\epsilon := \frac{1}{2}(\mu_+(u) - \mu) > 0$. For large $i \in \mathbb{N}$ the eigenvalues of $v_i\frac{s_i}{\|x_i\|}q_\mu(i)v_i$ are bounded above by $-\epsilon$ in restriction to $\mathrm{Im}(p_\mu(u))$ and zero on the complement $\ker(p_\mu(u))$. So

$$\mathbb{1} - p_\mu(u) \leq e^{v_i\frac{s_i}{\|x_i\|}q_\mu(i)v_i} \leq \mathbb{1} + (e^{-\epsilon} - 1)p_\mu(u)$$

and then $\mathbb{1} - p_\mu(u) \leq e^{v_i s_i q_\mu(i) v_i} \leq \mathbb{1} + (e^{-\epsilon\|x_i\|} - 1)p_\mu(u)$. We find $\lim_{i\to\infty}\|x_i\| = \infty$ that

$$\lim_{i\to\infty} e^{s_i q_\mu(i)} = \lim_{i\to\infty} v_i e^{v_i s_i q_\mu(i) v_i} v_i = (\mathbb{1} - 2p_\mu(u))(\mathbb{1} - p_\mu(u))(\mathbb{1} - 2p_\mu(u)) = \mathbb{1} - p_\mu(u).$$

Now consider

$$\begin{aligned} z = \lim_{i\to\infty} e^{s_i} = \lim_{i\to\infty} e^{\sum\limits_{\mu\in\mathrm{spec}(u)} s_i q_\mu(i)} &= \prod_{\mu\in\mathrm{spec}(u)} \lim_{i\to\infty} e^{s_i q_\mu(i)} = \prod_{\substack{\mu\in\mathrm{spec}(u)\\ \mu\neq\mu_+(u)}} (\mathbb{1} - p_\mu(u)) \lim_{i\to\infty} e^{s_i q_{\mu_+(u)}(i)} \\ &= p_+(u) \lim_{i\to\infty} e^{s_i q_{\mu_+(u)}(i)}. \end{aligned}$$

This shows $p_+(u)z = z$ and then by (2.46) we have $s(z) \leq p_+(u)$.

The discussion of the support projector bound for an accumulation point ρ of $(\exp_1(s_i)) = (\frac{e^{s_i}}{\mathrm{tr}(e^{s_i})})$ is led by geometry. The points s_i belong to the relative boundary of the negative of the positive cone $-\operatorname{rb}(A^+)$, (4.14). In particular, by the maximal eigenvalue zero the trace $\operatorname{tr}(e^{s_i}) \geq 1$ is bounded below away from zero and one has the inclusion

$$(e^{s_i}) \subset \{z \in A_{\mathrm{sa}} : \|z\| \leq 1\}$$

into the unit ball of A_{sa} which is a compact subset. The limit z of a convergent subsequence of $(e^{s_{i_j}})$ reproduces $\rho = \frac{z}{\operatorname{tr}(z)}$ with the support bound discussed above. qed

Proof of Corollary 6.26. Let $(\rho_i) \subset \mathcal{E}$ be a converging sequence with limit $\rho \notin \mathcal{E}$. For $i \in \mathbb{N}$ we put $\theta_i := \ln_0(\rho_i) \in \Theta$. If a subsequence of $\|\theta_i\|$ is bounded then (θ_i) has a converging subsequence (θ_{i_j}) with limit in Θ and by continuity of $\exp_1$ follows

$$\rho = \lim_{j\to\infty} \rho_{i_j} = \lim_{j\to\infty} \exp_1(\theta_{i_j}) = \exp_1(\lim_{j\to\infty} \theta_{i_j}) \in \mathcal{E}.$$

We see that the assumption $\rho \notin \mathcal{E}$ implies $\lim_{i\to\infty} \|\theta_i\| = \infty$.

By compactness of the unit sphere in A^0_{sa} the sequence $(\frac{\theta_i}{\|\theta_i\|})$ has an accumulation point. We select a converging subsequence and put $u := \lim_{i\to\infty} \frac{\theta_i}{\|\theta_i\|}$. From $\lim_{i\to\infty} \|\theta_i\| = \infty$ follows $u \in \operatorname{lin}(\Theta) = U$ so by (5.16) the maximal projector $p_+(u)$ of u belongs to the exposed projector lattice $\mathcal{P}_{U,\perp}$ of sr_U. We can apply Lemma 6.25 and get $s(\rho) \leq p_+(u)$ because $\rho = \lim_{i\to\infty} \exp_1(\theta_i) = \lim_{i\to\infty} \exp_1(\theta_i - \mu_+(\theta_i)\mathbb{1})$. qed

Proof of Lemma 6.28. Since $f(\overline{K} \setminus K) \cap L = \emptyset$ we have $f(\overline{K}) \cap L = f(K) \cap L$ and $L \setminus f(\overline{K}) = L \setminus f(K)$. This gives

$$L = (f(\overline{K}) \cap L) \cup ((L \setminus f(\overline{K})) \cap L) = (f(K) \cap L) \cup (L \setminus f(\overline{K})).$$

Since $f(K)$ is open and $f(\overline{K})$ is compact, the above union is a disconnection for L unless $L \setminus f(K) = L \setminus f(\overline{K})$ is empty. So $f(K) \supset L$. qed

Proof of Theorem 4. First we prove that the Jacobian of $\pi : \mathcal{E} \to U$ is invertible at every point $\sigma \in \mathcal{E}$. We choose a basis $x_1, \dots, x_p$ of $\mathrm{T}_\sigma \mathcal{E}$ such that the canonical representation $x_1^{(\Theta)}, \dots, x_p^{(\Theta)}$ is an ONB of the canonical tangent space $U = \mathrm{T}^{(\Theta)}_{\ln_0(\sigma)} \mathcal{E}$. Then for $u \in \mathrm{T}_\sigma \mathcal{E}$ we have by (6.10) and by (6.23)

$$\mathrm{D}\,\pi(u)|_\sigma = \pi_U(u^{(m)}) = \sum_{i=1}^{p} \langle x_i^{(\Theta)}, u^{(m)} \rangle x_i^{(\Theta)} = \sum_{i=1}^{p} \langle x_i, u \rangle_\sigma \, x_i^{(\Theta)}.$$

The BKM-metric $\langle \cdot, \cdot \rangle_\sigma$ is non-degenerate hence $\mathrm{T}_\sigma \mathcal{E} \to U$, $u \mapsto \mathrm{D}\,\pi(u)|_\sigma$ is invertible. We can conclude that π is a diffeomorphism through a proof of injectivity. We can prove injectivity of $\pi : \mathcal{E} \to U$ with Lemma 6.22. Let $\pi(\rho) = \pi(\sigma)$ for $\rho, \sigma \in \mathcal{E}$. Then $\rho - \sigma \in U^\perp$ implies $\rho + U^\perp = \sigma + U^\perp$. Thus

$$\{\rho\} = \mathcal{E} \cap (\rho + U^\perp) = \mathcal{E} \cap (\sigma + U^\perp) = \{\sigma\}.$$

We have proved that π is a diffeomorphism.

We want to find out the image of the mean value chart. The image of the diffeomorphism π is an open subset of U. And it is included in the relative interior of the state reflection

$$\pi(\mathcal{E}) \subset \mathrm{ri}(\mathrm{sr}_U)$$

because $\mathcal{E} \subset S(A) = \mathrm{ri}(\overline{S}(A))$. Suppose a state ρ belongs to $\overline{\mathcal{E}} \setminus \mathcal{E}$. It is proved in Corollary 6.26 that the support projector of ρ is dominated by a proper exposed projector $p \in \mathcal{P}_{U,\perp}$, that is $s(\rho) \leq p$. From this it follows by (5.58) that $\pi_V(\rho) \in \mathrm{rb}(\mathrm{sr}_U)$. This shows

$$\pi_U(\overline{\mathcal{E}} \setminus \mathcal{E}) \subset \mathrm{rb}(\mathrm{sr}_U).$$

With $K := \mathcal{E}$, $L := \mathrm{ri}(\mathrm{sr}_U)$ and $f := \pi_U|_{\overline{\mathcal{E}}}$ we meet the conditions of Lemma 6.28 that $\mathcal{E}$ is bounded, $\mathrm{ri}(\mathrm{sr}_U)$ is connected, $\pi_U : \overline{\mathcal{E}} \to \mathrm{sr}_U$ is continuous, $\pi_U(\mathcal{E})$ is open, $\pi_U(\mathcal{E}) \cap \mathrm{ri}(\mathrm{sr}_U) \neq \emptyset$ and $\pi_U\left(\overline{\mathcal{E}} \setminus \mathcal{E}\right) \cap \mathrm{ri}(\mathrm{sr}_U) = \emptyset$. The lemma proves

$$\pi_U(\mathcal{E}) = \mathrm{ri}(\mathrm{sr}_U).$$

By continuity of π_U and by compactness of $\overline{\mathcal{E}}$ we get also $\pi_U(\overline{\mathcal{E}}) = \mathrm{sr}_U$. qed

Proof of Corollary 6.30. This follows immediately from the definition (6.29) of the mean value parametrization and from Theorem 4. qed

Proof of Corollary 6.31. By Theorem 4 we have

$$\mathrm{Cyl} = \mathcal{E} + U^\perp = \pi_U(\mathcal{E}) + U^\perp = \mathrm{ri}(\mathrm{sr}_U) + U^\perp.$$

By equivariance (3.15) of reduction to the relative interior under affine maps we have $\mathrm{ri}(\mathrm{sr}_U) = \mathrm{ri}(\pi_U(\overline{S}(A))) = \pi_U(S(A))$ and obtain $\mathrm{Cyl} = \pi_U(S(A)) + U^\perp = S(A) + U^\perp$ as desired. For the domain expression notice that

$$\mathrm{Dom} = \overline{S}(A) \cap \mathrm{Cyl} = \overline{S}(A) \cap \left(\mathrm{ri}(\mathrm{sr}_U) + U^\perp\right).$$

The argument is completed by (5.58) which says that a state $\rho \in \overline{S}(A)$ projects under π_U to the relative boundary $\mathrm{rb}(\mathrm{sr}_U)$ if and only if the support projector $s(\rho)$ of ρ satisfies $s(\rho) \leq p$ for some proper projector $p \in \mathcal{P}_U$. qed

Proof of Lemma 7.5. At the beginning we estimate the support projector of a possible limit point. We can not apply Lemma 6.25 directly because the largest eigenvalue of $\theta+\lambda u$ may not be zero. Thus we choose a monotone real sequence (λ_i) with $\lim_{i\to\infty}\lambda_i=\infty$ and we study the limit $z=\lim_{i\to\infty}e^{\theta+\lambda_i u}$ for $u\neq 0$. With Lemma 6.25 we have the estimate $s(\widetilde{z})\leq p_+(u)$ for the support projector of an accumulation point $\widetilde{z}$ of

$$\exp(\theta+\lambda_i u-\mu_+(\theta+\lambda_i u)\mathbb{1})$$

because $\lim_{i\to\infty}\frac{\theta+\lambda_i u}{\|\theta+\lambda_i u\|}=\frac{u}{\|u\|}$. Let us prove $s(z)=s(\widetilde{z})$. For $i\in\mathbb{N}$ we have

$$e^{\theta+\lambda_i u-\mu_+(\theta+\lambda_i u)\mathbb{1}}=e^{-\mu_+(\theta+\lambda_i u)}e^{\theta+\lambda_i u}$$

and it is sufficient to show that $e^{-\mu_+(\theta+\lambda_i u)}$ converges to a non-zero number for $i\to\infty$. We show that the maximal eigenvalue $\mu_+(\theta+\lambda_i u)$ is bounded below and monotone decreasing. With H the Hilbert space (2.7) of A, one has by the min-max principle [Ree4] $\mu_+(\theta+\lambda_i u)=\max_{x\in H,\|x\|_2=1}\langle x,(\theta+\lambda_i u)(x)\rangle$. Since $\langle x,u(x)\rangle\leq 0$ for $x\in H$ we see the desired monotonicity. Using a normalized vector $x\in\ker(u)$ we get the lower bound $\mu_+(\theta+\lambda_i u)\geq\langle x,\theta(x)\rangle$. A monotone decreasing and bounded real sequence converges. We have proved $s(z)\leq p_+(u)$.

To calculate z we denote for $\mu\in\operatorname{spec}(u)$ the spectral projector $p_\mu:=p_\mu(u)$ of u. Then $p_0=p_+(u)$ is the kernel projector and the maximal projector of u. We consider for real parameter $c>0$ the perturbed matrix

$$t(c):=u+c\theta.$$

For an eigenvalue $\mu\in\operatorname{spec}(u)$ and a positive-oriented curve γ_μ in the resolvent set of u, enclosing the eigenvalue μ and no other eigenvalues of u, we have the total projector $q_\mu(c):=-\frac{1}{2\pi i}\int_{\gamma_\mu}(t(c)-\zeta)^{-1}\,\mathrm{d}\,\zeta$. The total projector is expanded as[1]

$$q_\mu(c)=p_\mu+cq_\mu^{(1)}+o(c)_{c\to 0}$$

for $q_\mu^{(1)}:=\frac{1}{2\pi i}\int_{\gamma_\mu}(u-\zeta)^{-1}\theta(u-\zeta)^{-1}\,\mathrm{d}\,\zeta$, see [KaT] II-§1.3/1.4. We get

$$q_\mu(c)t(c)q_\mu(c)=\mu p_\mu+\ c\left(\mu(p_\mu q_\mu^{(1)}+q_\mu^{(1)}p_\mu)+p_\mu\theta p_\mu\right)+o(c)_{c\to 0}.$$

In particular for $\mu=0$ one has

$$\frac{1}{c}q_0(c)t(c)q_0(c)=p_0\theta p_0+o(1)_{c\to 0}. \tag{10.10}$$

[1]The equality $f(c)=o(g(c))_{c\to d}$ means that $\lim_{c\to d}\frac{\|f(c)\|}{|g(c)|}=0$ for an operator valued function f, a real function g and $d\in\mathbb{R}\cup\{\pm\infty\}$.

The total projectors $q_\mu(c)$ are pairwise orthogonal and the set $\{q_\mu(c)\}_{\mu\in\mathrm{spec}(u)}$ is a complete set of projectors for small values of c. We expand for small values of c

$$e^{\frac{1}{c}t(c)} = e^{\sum_{\mu\in\mathrm{spec}\,u}\frac{1}{c}q_\mu(c)t(c)q_\mu(c)} = \prod_{\mu\in\mathrm{spec}\,u} e^{\frac{1}{c}q_\mu(c)t(c)q_\mu(c)}.$$

As shown above, the support projector is bounded $s(z)\leq p_0$ and we get

$$z = p_0 z = p_0 \lim_{c\to 0} e^{\frac{1}{c}t(c)} = \lim_{c\to 0} q_0(c)e^{\frac{1}{c}t(c)} = \lim_{c\to 0} q_0(c)e^{\frac{1}{c}q_0(c)t(c)q_0(c)}.$$

The limit is evaluated using (10.10) as $z = p_0 e^{p_0\theta p_0}$. qed

Proof of Lemma 7.6. Let $\theta, u \in A_{\mathrm{sa}}$. We put $\mu := \mu_+(u)$ the maximal eigenvalue of u and $p := p_+(u)$ the maximal projector of u. For $\lambda\in\mathbb{R}$ one has $\exp_1(\theta+\lambda u) = \exp_1(\theta+\lambda(u-\mu\mathbb{1}))$ by invariance of $\exp_1$. Since $p = p_+(u-\mu\mathbb{1})$ we get by Lemma 7.5

$$\lim_{\lambda\to\infty}\exp_1(\theta+\lambda u) = \frac{pe^{p\theta p}}{\mathrm{tr}(pe^{p\theta p})}.$$

The free energy (6.6) is for $\lambda\in\mathbb{R}$

$$F(\theta+\lambda u)-\lambda\mu = F\left(\theta+\lambda(u-\mu\mathbb{1})\right) = \ln\left(\mathrm{tr}\left(e^{\theta+\lambda(u-\mu\mathbb{1})}\right)\right).$$

We get by Lemma 7.5 $\lim_{\lambda\to\infty}\left(F(\theta+\lambda u)-\lambda\mu\right) = \ln\left(\mathrm{tr}\left(pe^{p\theta p}\right)\right)$. Using transport of functional calculus (2.33) under κ^p this gives $\ln\left(\mathrm{tr}\left(pe^{p\theta p}\right)\right) = \ln\left(\mathrm{tr}\left(e^{(\kappa^p)^{-1}(p\theta p)}\right)\right) = F\left((\kappa^p)^{-1}(p\theta p)\right)$. qed

Proof of Theorem 5. To begin, observe that $\mathcal{E}$ is covered by e-geodesics contained in $\mathcal{E}$ and trivially $\mathcal{E} = \kappa^{\mathbb{1}}(\mathcal{E}_{\mathbb{1}})$. Other points $\rho\in\overline{S}(A)\setminus\mathcal{E}$ belong to the closure of an e-geodesic in $\mathcal{E}$ if and only if

$$\rho = \lim_{\lambda\to\infty}\exp_1(\theta+\lambda u) \qquad \text{for some } \theta\in\Theta \text{ and } u\in U\setminus\{0\}.$$

By the limit expression in Lemma 7.6 and by the characterization of the exposed projector lattice $\mathcal{P}_{U,\perp}$ through maximal projectors, see (5.16), the above statement is equivalent to

$$\rho = \frac{pe^{p\theta p}}{\mathrm{tr}(pe^{p\theta p})} \qquad \text{for some } \theta\in\Theta \text{ and a proper } p\in\mathcal{P}_{U,\perp}.$$

Using the invariant parametrization (7.7) of a compression $\kappa^p(\mathcal{E}_p)$, the previous statement is equivalent to

$$\rho\in\kappa^p(\mathcal{E}_p) \qquad \text{for a proper } p\in\mathcal{P}_{U,\perp}.$$

qed

Proof of Corollary 7.7. The state reflection sr_U is the disjoint union of relative interiors of its faces (5.56) $\mathrm{sr}_U = \bigcup_{p\in\mathcal{P}_U\setminus\{0\}} \mathrm{ri}(\mathbb{F}_U(p))$. For non-zero $p \in \mathcal{P}_U$ the relative interior $\mathrm{ri}(\mathbb{F}_U(p))$ is the set of mean values with respect to U of the compression $\kappa^p(\mathcal{E}_p)$ (7.10). By Theorem 5 the compression $\kappa^p(\mathcal{E}_p)$ is covered by closures of e-geodesics in $\mathcal{E}$ if and only if $p \in \mathcal{P}_{U,\perp}$. For all other projectors $p \in \mathcal{P}$ the compression $\kappa^p(\mathcal{E}_p)$ is not met by the closure of any e-geodesic in $\mathcal{E}$. `qed`

Proof of Proposition 7.9. Let $p \in \mathcal{P}_U$ be a non-zero projector. The cylinder on $\mathcal{E}_p$ is $\mathrm{Cyl}_p = \mathcal{E}_p + \varsigma^p(U)^\perp = \mathrm{ri}(\mathrm{sr}_{\varsigma^p(U)}) + \varsigma^p(U)^\perp = S(A) + \varsigma^p(U)^\perp$, see (7.11). This implies the sequence

$$\mathcal{E}_p \subset S(A^p) \subset \mathrm{Dom}_p \subset \mathrm{Cyl}_p$$

ordered by inclusion. Recall that $\mathrm{Dom}_p = \overline{S}(A^p) \cap \mathrm{Cyl}_p$ by definition. For the smallest set $\mathcal{E}_p$ in the sequence follows from the mean value chart for compressions (7.9) the diffeomorphism

$$\pi_p = \pi_{\varsigma^p(U)}|_{\mathcal{E}_p} : \quad \mathcal{E}_p \to \mathrm{ri}(\mathrm{sr}_{\varsigma^p(U)}).$$

The inverse is given by the mean value parametrization $M_p : \mathrm{ri}(\mathrm{sr}_{\varsigma^p(U)}) \to \mathcal{E}_p$, see Corollary 6.30. But the largest set in the sequence $\mathrm{Cyl}_p = \mathcal{E}_p + \varsigma^p(U)^\perp$ has the same mean values. We have verified the left-hand side of the diagram.

It is not reasonable to include $\kappa^p(\mathrm{Cyl}_p)$ into the diagram, see Remark 5.21 (c). We enter the right-hand side of the diagram through the domain Dom_p of $\mathcal{E}_p$. Since $\mathrm{Dom}_p \subset (A^p)^1_{\mathrm{sa}}$ we get from the first diagram in Proposition 5.20 that the whole diagram commutes, except we have to verify the expression for

$$\kappa^p(\mathcal{E}_p) \subset \kappa^p(S(A^p)) \subset \kappa^p(\mathrm{Dom}_p)$$

and $\pi_U(\kappa^p(\mathrm{Dom}_p))$ on the right hand side of the diagram. By (4.38) we have $\kappa^p\left(S(A^p)\right) = \mathrm{ri}\left(\mathbb{F}(p)\right)$. The formula for the domain follows from Proposition 5.27 applied to $\mathrm{ri}(\mathbb{F}(p)) \subset \mathbb{F}(p)$,

$$\begin{aligned}
\kappa^p(\mathrm{Dom}_p) &= \kappa^p\left(\overline{S}(A^p) \cap \mathrm{Cyl}_p\right) = \kappa^p\left(\overline{S}(A^p) \cap \left(S(A^p) + \varsigma^p(U)^\perp\right)\right)\\
&= \mathbb{F}(p) \cap \left(\mathrm{ri}(\mathbb{F}(p)) + \kappa^p\left(\varsigma^p(U)^\perp\right)\right)\\
&= \overline{S}(A) \cap \left(\mathrm{ri}(\mathbb{F}(p)) + U^\perp\right) = \overline{S}(A) \cap \left(\mathrm{ri}(\mathbb{F}_U(p)) + U^\perp\right).
\end{aligned}$$

From this expression follows $\pi_U(\kappa^p(\mathrm{Dom}_p)) \subset \mathrm{ri}(\mathbb{F}_U(p))$. By (7.10) there is a diffeomorphism $\pi_U|_{\kappa^p(\mathcal{E}_p)} : \kappa^p(\mathcal{E}_p) \to \mathrm{ri}(\mathbb{F}_U(p))$. `qed`

Proof of Proposition 7.11. A density matrix $\rho \in \overline{S}(A)$ belongs to $\kappa^p(\mathrm{Dom}_p)$ for a unique non-zero projector $p \in \mathcal{P}_U$ by (7.13). We can show that

$$p = \bigwedge\{q \in \mathcal{P}_U : q \geq s(\rho)\}.$$

Since $\rho \in \mathbb{F}(p)$ one has $p \geq s(\rho)$ (4.33). Hence it remains to prove that p is the minimum of $\{q \in \mathcal{P}_U : q \geq s(\rho)\}$. For all $q \geq s(\rho)$ we have $\rho \in \mathbb{F}(q)$ so $\pi_U(\rho) \in \mathbb{F}_U(q)$. On the other hand, since $\rho \in \kappa^p(\mathrm{Dom}_p)$ we get from Proposition 7.9 that $\pi_U(\rho) \in \mathrm{ri}(\mathbb{F}_U(p))$. Now the relative interior $\mathrm{ri}(\mathbb{F}_U(p))$ meets the face $\mathbb{F}_U(q)$ and Remark 3.9 (e) proves the inclusion $\mathbb{F}_U(p) \subset \mathbb{F}_U(q)$. The lattice isomorphism $\mathcal{F}_U \to \mathcal{P}_U$ (5.53) concludes with $p \leq q$.

The combinatorial extension $\mathcal{E}^{\mathrm{cmb}} = \bigcup_{p \in \mathcal{P}_U \setminus \{0\}} \kappa^p(\mathcal{E}_p)$ is composed of isomorphically labeled parts compared to the state reflection $\mathrm{sr}_U = \bigcup_{p \in \mathcal{P}_U \setminus \{0\}} \mathrm{ri}(\mathbb{F}_U(p))$. On each of these parts, for non-zero $p \in \mathcal{P}_U$, there is a bijection induced by the mean value chart (7.10)

$$\pi_U|_{\kappa^p(\mathcal{E}_p)} : \kappa^p(\mathcal{E}_p) \to \mathrm{ri}(\mathbb{F}_U(p)).$$

The ranges $\mathrm{ri}(\mathbb{F}_U(p))$ are disjoint by the stratification of the state reflection into relative interiors of faces. The domains $\kappa^p(\mathcal{E}_p)$ (included in $\mathrm{ri}(\mathbb{F}(p))$) are disjoint by stratification of the state space in relative interiors of faces. Hence the combination $\pi^{\mathrm{cmb}} : \mathcal{E}^{\mathrm{cmb}} \to \mathrm{sr}_U$ (7.5) of these restricted projections is a bijection. The combinatorial mean value chart $\pi^{\mathrm{cmb}} = \pi_U|_{\mathcal{E}^{\mathrm{cmb}}}$ is a restriction of the projection π_U. Hence for a point $x \in \mathrm{ri}(\mathbb{F}_U(p)) = \pi_U(\kappa^p(\mathcal{E}_p))$ with non-zero $p \in \mathcal{P}_U$ the fiber $x + U^\perp$ intersects $\mathcal{E}^{\mathrm{cmb}}$ at $(\pi^{\mathrm{cmb}})^{-1}(x) = (\pi_U|_{\kappa^p(\mathcal{E}_p)})^{-1}(x)$ and nowhere else. This proves the geometric expression

$$x \mapsto (x + U^\perp) \cap \mathcal{E}^{\mathrm{cmb}}$$

for the inverse $(\pi^{\mathrm{cmb}})^{-1}$ of the combinatorial mean value chart. qed

Proof of Lemma 7.14. By Lemma 6.22 for a state $\rho \in \mathrm{Dom}$ there is a unique minimum of relative entropy (6.18) on $\mathcal{E}$ at $N(\rho) \in \mathcal{E}$. The combinatorial normal projection N^{cmb} restricts to $N = N^{\mathrm{cmb}}|_{\mathrm{Cyl}}$ by definition and the distance $S_\rho(N(\rho))$ is finite since $N(\rho) \in \mathcal{E}$ has full support. It remains to show $S_\rho(\sigma) > S_\rho(N(\rho))$ for $\sigma \in \mathcal{E}^{\mathrm{cmb}} \setminus \mathcal{E}$. In fact we prove $S_\rho(\sigma) = \infty$ for $\sigma \in \mathcal{E}^{\mathrm{cmb}} \setminus \mathcal{E}$. Since $\rho \in \mathrm{Dom}$ one has by Corollary 6.31

$$s(\rho) \not\leq p \text{ for all proper } p \in \mathcal{P}_U.$$

By definition (7.4) of the combinatorial extension $\mathcal{E}^{\mathrm{cmb}}$ and since $\sigma \in \mathcal{E}^{\mathrm{cmb}} \setminus \mathcal{E}$ there is a proper projector $p \in \mathcal{P}_U$ such that $\sigma \in \kappa^p(\mathcal{E}_p)$. This implies $s(\sigma) = p$. Hence $s(\rho) \not\leq s(\sigma)$ and $S_\rho(\sigma) = \infty$. qed

Proof of Lemma 7.15. Let $\theta, u \in A^0_{\mathrm{sa}}$ with non-zero u. The derivative of relative entropy in the canonical chart is (6.21)

$$\frac{\mathrm{d}}{\mathrm{d}\,\lambda} S_\rho\left(\exp_1(\theta + \lambda u)\right) = \langle u, \exp_1(\theta + \lambda u) - \rho \rangle.$$

The face of $\overline{S}(A)$ exposed by u is $F_\perp(u) = \mathbb{F}(p)$ (4.47) for $p := p_+(u)$ the maximal projector of u. If $\rho \in F_\perp(u)$ then for arbitrary $\tau \in \overline{S}(A)$ the inequality $\langle u, \tau - \rho\rangle \leq 0$ (3.34) holds. Since $u \in \operatorname{lin}(\overline{S}(A)) = A^0_{\mathrm{sa}}$ (4.26) one has by Theorem 13.1 in [Ro] for τ in the relative interior $\operatorname{ri}(\overline{S}(A))$ the strict inequality $\langle u, \tau - \rho\rangle < 0$. Points on an e-geodesic are invertible and belong to $\operatorname{ri}(\overline{S}(A)) = S(A)$ (4.27). So we get for $\lambda \in \mathbb{R}$

$$\frac{\mathrm{d}}{\mathrm{d}\,\lambda} S_\rho\left(\exp_1(\theta + \lambda u)\right) < 0.$$

The limit of the e-geodesic is by Lemma 7.6 and by (7.7)

$$\sigma := \lim_{\lambda\to\infty} \exp_1(\theta + \lambda u) = \kappa^p\left(\exp_1\left((\kappa^p)^{-1}(p\theta p)\right)\right).$$

We use the decomposition (6.20) of relative entropy for an invertible density matrix τ

$$S(\rho, \tau) = -S(\rho) - \langle \rho, \ln_0(\tau)\rangle + F(\ln_0(\tau)).$$

The limit σ is not invertible but it can be made invertible by transition to the compression A^p. Let $\rho_p := (\kappa^p)^{-1}(\rho)$ and $\sigma_p := (\kappa^p)^{-1}(\sigma)$. Then $\ln_0(\sigma_p) = (\kappa^p)^{-1}(p\theta p)$ and with the limit of free energy calculated in Lemma 7.6 we have

$$\begin{aligned} S(\rho,\sigma) + S(\rho) &= S(\rho_p, \sigma_p) + S(\rho_p) = -\langle \rho_p, \ln_0(\sigma_p)\rangle + F(\ln_0(\sigma_p)) \\ &= -\langle (\kappa^p)^{-1}(\rho), (\kappa^p)^{-1}(p\theta p)\rangle + F((\kappa^p)^{-1}(p\theta p)) \\ &= -\langle \rho, \theta\rangle + \lim_{\lambda\to\infty}\left[F(\theta + \lambda u) - \lambda \mu_+(u)\right]. \end{aligned}$$

Since $s(\rho) \leq p = p_+(u)$ one has $\langle \rho, u\rangle = \mu_+(u)$ and so

$$\begin{aligned} S(\rho,\sigma) + S(\rho) &= \lim_{\lambda\to\infty}\left[-\langle \rho, \theta + \lambda u\rangle + F(\theta + \lambda u)\right] \\ &= \lim_{\lambda\to\infty}\left[S(\rho, \exp_1(\theta + \lambda u)) + S(\rho)\right]. \end{aligned}$$

The desired equation $S(\rho,\sigma) = \lim_{\lambda\to\infty} S(\rho, \exp_1(\theta + \lambda u))$ follows. `qed`

Proof of Lemma 7.16. Let $p \in \mathcal{P}_U$ and $q \in \kappa^p\left(\mathcal{P}_{\varsigma^p(U),\perp}\right)$ be non-zero projectors with $q \lneq p$. Then $Q := (\kappa^p)^{-1}(q) \in \mathcal{P}_{\varsigma^p(U),\perp}$ is a proper projector. By (5.16) there exists non-zero $\widetilde{u} \in \varsigma^p(U)$ such that $Q = p_+(\widetilde{u})$. First we prove the covering formula. Let $\widetilde{\sigma} \in \mathcal{E}_p$ be arbitrary. Then for $\widetilde{\theta} := \ln_0(\widetilde{\sigma}) \in \varsigma^p(\Theta)$ the e-geodesic

$$\lambda \mapsto g_{\widetilde{\theta}}(\lambda) := \exp_1(\widetilde{\theta} + \lambda \widetilde{u})$$

is included in $\mathcal{E}_p$ it passes through $\widetilde{\sigma}$. Let us prove that the following diagram commutes.

$$\begin{tikzcd} \Theta \arrow[r, "\varsigma^p"] \arrow[d, "\varsigma^q"'] & \varsigma^p(\Theta) \arrow[d, "\kappa^p\left[\lim\limits_{\lambda\to\infty} \exp_1(g_{(\cdot)}(\lambda))\right]"] \\ \varsigma^q(\Theta) \arrow[r, "\kappa^q\circ\exp_1"'] & \kappa^q(\mathcal{E}_q) \end{tikzcd}$$

We have the equality $\widetilde{\theta} = \varsigma^p(\theta)$ for some $\theta \in \Theta$. By Lemma 7.6 one has

$$\kappa^p \left[\lim_{\lambda\to\infty} \exp_1(g_{\widetilde{\theta}}(\lambda)) \right] = \kappa^p \left[\frac{Qe^{Q\widetilde{\theta}Q}}{\operatorname{tr}\left(Qe^{Q\widetilde{\theta}Q}\right)} \right] = \frac{qe^{q\kappa^p(\widetilde{\theta})q}}{\operatorname{tr}\left(qe^{q\kappa^p(\widetilde{\theta})q}\right)}. \tag{10.11}$$

One has (5.60) $\kappa^p(\widetilde{\theta}) = p\theta p - \operatorname{tr}(p\theta)\frac{p}{\operatorname{tr}(p)}$. Since $q \leq p$ and by invariance of (10.11) under addition of multiples of q to $\kappa^p(\widetilde{\theta})$, (10.11) is equal to $\frac{qe^{q\theta q}}{\operatorname{tr}\left(qe^{q\theta q}\right)}$. This is (modulo κ^q) the surjective pullback (7.7) of the canonical parametrization of $\mathcal{E}_q$, so the diagram commutes and $\kappa^q(\mathcal{E}_q)$ is covered by the images of the mappings in the diagram.

Let $\rho \in \mathbb{F}(q)$. We visit the algebra A^p and set $\rho_p := (\kappa^p)^{-1}(\rho) \in \overline{S}(A^p)$. The relative entropy $S_\rho\left(\kappa^p\left(\exp_1(\widetilde{\theta} + \lambda\widetilde{u})\right)\right) = S_{\rho_p}\left(\exp_1(\widetilde{\theta} + \lambda\widetilde{u})\right)$ is strictly monotone decreasing in the parameter λ by Lemma 7.15 because $s(\rho_p) \leq p_+(\widetilde{u})$. Moreover

$$\lim_{\lambda\to\infty} S_{\rho_p}(g_{\widetilde{\theta}}(\lambda)) = S_{\rho_p}(\lim_{\lambda\to\infty} g_{\widetilde{\theta}}(\lambda)) = S_\rho\left(\kappa^p\left(\lim_{\lambda\to\infty} g_{\widetilde{\theta}}(\lambda)\right)\right).$$

Hence the infimum is attained at the limit point. qed

Proof of Theorem 6. At the beginning let us choose $\rho \in \overline{S}(A)$. Then $\rho \in \kappa^p(\mathrm{Dom}_p)$ for the infimum of projectors

$$p := \bigwedge\{r \in \mathcal{P}_U : r \geq s(\rho)\}$$

by Proposition 7.11. If $p = \mathbb{1}$ then $\rho \in \mathrm{Dom} = \overline{S}(A) \cap (S(A) + U^\perp)$ and the statement of the theorem is proved in Lemma 7.14. Let us assume in the following that $p \in \mathcal{P}_U$ is a proper projector. We divide the proof in four paragraphs. Let $q \in \mathcal{P}_U$ be an arbitrary non-zero projector.

(a) S_ρ has a unique minimum on $\kappa^p(\mathcal{E}_p)$ at $N^{\mathrm{cmb}}(\rho)$,

(b) if $q \geq p$ then $\inf_{\sigma\in\mathcal{E}} S_\rho(\sigma) = \inf_{\sigma\in\kappa^q(\mathcal{E}_q)} S_\rho(\sigma)$,

(c) if $q \gneq p$ then S_ρ has no minimum on $\kappa^q(\mathcal{E}_q)$,

(d) if $q \not\geq p$ then $S_\rho(\sigma) = \infty$ for all $\sigma \in \kappa^q(\mathcal{E}_q)$.

These statements are a self-explaining proof of the theorem because $\mathcal{E}^{\mathrm{cmb}} = \bigcup_{q\in\mathcal{P}_U\setminus\{0\}} \kappa^q(\mathcal{E}_q)$ (7.4) and because "$\leq$" is a partial ordering on the projector lattice $\mathcal{P}_U$.

(a) We prove the existence of a minimum of S_ρ on $\kappa^p(\mathcal{E}_p)$. Put $\rho_p := (\kappa^p)^{-1}(\rho) \in \mathrm{Dom}_p$. Using the normal projection N_p for $\mathcal{E}_p$ (7.12) and Lemma 6.22 there is unique minimum of relative entropy on $\kappa^p(\mathcal{E}_p)$,

$$\inf_{\sigma\in\kappa^p(\mathcal{E}_p)} S_\rho(\sigma) = \inf_{\sigma_p\in\mathcal{E}_p} S_{\rho_p}(\sigma_p) = S_{\rho_p}(N_p(\rho_p)).$$

With (7.22) we have $S_{\rho_p}(N_p(\rho_p)) = S_\rho(\kappa^p \circ N_p(\rho_p)) = S_\rho(N^{\mathrm{cmb}}(\rho))$.

(b) We show for $q \geq p$ the equality of infima $\inf_{\sigma\in\mathcal{E}} S_\rho(\sigma) = \inf_{\sigma\in\kappa^q(\mathcal{E}_q)} S_\rho(\sigma)$. If $q = \mathbb{1}$ then nothing is to show. Otherwise if q is proper then by Corollary 5.32 there is an access sequence of projectors for sr_U including both p and q. We can use for some $m \in \mathbb{N}$ an access sequence of projectors for sr_U

$$\mathbb{1} \gneq p_1 \gneq \cdots \gneq p_m$$

such that $p = p_m$ and $q = p_i$ for some $i \in \{1,\dots,m\}$. We put $p_0 := \mathbb{1}$. By Definition 5.29 of an access sequence one has for $i \in \{0,\dots,m-1\}$ and two successive members $p_i \gneq p_{i+1}$ of the sequence the proper projector $(\kappa^{p_i})^{-1}(p_{i+1}) \in \mathcal{P}_{\varsigma^{p_i}(U),\perp}$. Further, since $\rho \in \kappa^p(\mathrm{Dom}_p)$ one has $\rho \in \mathbb{F}(p)$ and so $\rho \in \mathbb{F}(p_i)$ for $i \in \{1,\dots,m\}$. Now, for each $i \in \{0,\dots,m-1\}$ there is by Lemma 7.16 a collection of e-geodesics $g_{\widetilde{\sigma}} : \mathbb{R} \to \mathcal{E}_{p_i}$ labeled by $\widetilde{\sigma} \in \mathcal{E}_{p_i}$, included in $\mathcal{E}_{p_i}$ and passing through $\widetilde{\sigma}$. Then the lemma proves the equality

$$\begin{aligned}\inf_{\widetilde{\sigma}\in\mathcal{E}_{p_i}} S_\rho(\kappa^{p_i}(\widetilde{\sigma})) &= \inf_{\widetilde{\sigma}\in\mathcal{E}_{p_i}} \inf_{\tau\in g_{\widetilde{\sigma}}} S_\rho(\kappa^{p_i}(\tau)) = \inf_{\widetilde{\sigma}\in\mathcal{E}_{p_i}} S_\rho\left(\kappa^{p_i}\left(\lim_{\lambda\to\infty} g_{\widetilde{\sigma}}(\lambda)\right)\right)\\ &= \inf_{\sigma\in\mathcal{E}_{p_{i+1}}} S_\rho(\kappa^{p_{i+1}}(\sigma)).\end{aligned}$$

By induction we get the desired equality for q.

(c) We prove for $q \gneq p$ that S_ρ has no minimum on $\kappa^q(\mathcal{E}_q)$. Here we have to include the case $q = \mathbb{1}$. We can use the setup of **(b)** with the modification that q may be equal to $p_0 = \mathbb{1}$. Since $q \gneq p$ one has $q = p_i$ for some $i < m$. When we consider an arbitrary point $\widetilde{\sigma} \in \mathcal{E}_{p_i}$ then by Lemma 7.16 the relative entropy S_ρ is strictly monotone decreasing along the curve $\kappa^{p_i} \circ g_{\widetilde{\sigma}}$. Thus a minimum on $\kappa^{p_i}(\mathcal{E}_{p_i})$ is impossible.

(d) We prove for $q \not\geq p$ that $S_\rho(\sigma) = \infty$ for all $\sigma \in \kappa^q(\mathcal{E}_q)$. Since p is defined as the infimum $p = \bigwedge\{r \in \mathcal{P}_U : r \geq s(\rho)\}$ the inequality $q \geq s(\rho)$ implies $q \geq p$. Hence the relation $q \not\geq s(\rho)$ holds. A point $\sigma \in \kappa^q(\mathcal{E}_q)$ has support $s(\sigma) = q$ so $s(\sigma) \not\geq s(\rho)$ and the relative entropy (6.17) is $S_\rho(\sigma) = \infty$. qed

Proof of Corollary 7.17. The reverse information closure of $\mathcal{E}$ is defined by (7.2)

$$\mathrm{cl}_{rI}(\mathcal{E}) := \{\rho \in \overline{S}(A) : \inf_{\sigma\in\mathcal{E}} S(\rho,\sigma) = 0\}.$$

For arbitrary $\rho \in \overline{S}(A)$ we have by Theorem 6

$$\inf_{\sigma \in \mathcal{E}} S(\rho, \sigma) = S(\rho, N^{\mathrm{cmb}}(\rho)).$$

Hence $\rho \in \mathrm{cl}_{rI}(\mathcal{E})$ if and only if $S(\rho, N^{\mathrm{cmb}}(\rho)) = 0$, which is equivalent to $\rho = N^{\mathrm{cmb}}(\rho)$ (6.19). By Remark 7.13 one has $N^{\mathrm{cmb}}(\overline{S}(A)) = \mathcal{E}^{\mathrm{cmb}}$ hence $\rho = N^{\mathrm{cmb}}(\rho)$ if and only if $\rho \in \mathcal{E}^{\mathrm{cmb}}$. qed

Proof of Example 7.21. For $\mathcal{E}$ we use the parametrization with $t, \varphi \in \mathbb{R}$

$$\rho(t, \varphi) := \exp_1 \left[t((\sigma_1 \oplus 0)\sin(\varphi) + (\sigma_2 \oplus 1)\cos(\varphi)) \right].$$

With $\mathbb{1}_2$ the identity in M_2, this is by (4.8)

$$\rho(t, \varphi) = \frac{(\mathbb{1}_2 \cosh(t) + b(\varphi)\widehat{\sigma}\sinh(t)) \oplus e^{t\cos(\varphi)}}{2\cosh(t) + e^{t\cos(\varphi)}}$$

which we want to write for large t as

$$\rho(t, \varphi) = \frac{\frac{1}{2}\left(\frac{2\cosh(t)}{e^t}\mathbb{1}_2 + \frac{2\sinh(t)}{e^t}b(\varphi)\widehat{\sigma}\right) \oplus e^{t(\cos(\varphi)-1)}}{\frac{2\cosh(t)}{e^t} + e^{t(\cos(\varphi)-1)}}$$

because $\frac{2\cosh(t)}{e^t} \to 1$ and $\frac{2\sinh(t)}{e^t} \to 1$ for $t \to \infty$. We parametrize the relative interior of the segment g with $\lambda \in (0,1)$ by

$$\tau(\lambda) := \frac{\lambda}{2}p + (1-\lambda)q = (1 - \frac{\lambda}{2})p_+(\sigma_2) \oplus \frac{\lambda}{2}.$$

For $t > 0$ we use the angles

$$\varphi(t) := \sqrt{\frac{2}{t}\ln\left(\frac{2-\lambda}{\lambda}\right)}$$

that satisfy $\lim_{t\to\infty}\varphi(t) = 0$ and

$$e^{t(\cos(\varphi(t))-1)} = e^{-\frac{t}{2}\varphi(t)^2(1+o(1)_{t\to\infty})} = \left(\frac{\lambda}{2-\lambda}\right)^{1+o(1)_{t\to\infty}}.$$

Then

$$\lim_{t\to\infty}\rho(t, \varphi(t)) = \frac{\frac{1}{2}(\mathbb{1}_2 + \sigma_2) \oplus \frac{\lambda}{2-\lambda}}{1 + \frac{\lambda}{2-\lambda}} = \tau(\lambda).$$

This proves that $g \subset \overline{\mathcal{E}}$. qed

Proof of Lemma 7.22. Let $\rho \neq \sigma$ in $\overline{\mathcal{E}}$ and put $x := \pi_U(\rho) = \pi_U(\sigma) \in \mathrm{sr}_U$. Then, with $a := \langle \sigma - \rho, \rho \rangle$ and $b := \langle \sigma - \rho, \sigma \rangle$ one has $b = a + \|\sigma - \rho\|_2^2 > a$ and we obtain a real segment $[a, b]$ of positive length.

Let $n \in \mathbb{N}$. There exist $\widehat{\rho}, \widehat{\sigma} \in \mathcal{E}$ such that $\|\rho - \widehat{\rho}\|_2 < \frac{1}{n}$ and $\|\sigma - \widehat{\sigma}\|_2 < \frac{1}{n}$. We put $y := \pi_U(\widehat{\rho})$ and $z := \pi_U(\widehat{\sigma})$ and define the curve for $\lambda \in [0, 1]$

$$r(\lambda) := (1 - \lambda)y + \lambda z.$$

By Theorem 4 the curve r belongs to the relative interior $\mathrm{ri}(\mathrm{sr}_U)$ of the state reflection. We can use the mean value parametrization for $\mathcal{E}$ established in Corollary 6.30 and define a curve in $\mathcal{E}$ for $\lambda \in [0, 1]$

$$\tau(\lambda) := M \circ r(\lambda).$$

Using the inequality $\|x - y\|_2 = \|\pi_U(\rho) - \pi_U(\widehat{\rho})\|_2 \leq \|\rho - \widehat{\rho}\|_2 < \frac{1}{n}$ and the analogue $\|x - z\|_2 < \frac{1}{n}$ we have for $\lambda \in [0, 1]$ the inequality $\|x - r(\lambda)\|_2 < \frac{1}{n}$. This proves that the curve τ is included in the cylinder

$$\{u \in U : \|x - u\|_2 < \frac{1}{n}\} + U^{\perp}. \tag{10.12}$$

The curve τ leads from $\widehat{\rho}$ to $\widehat{\sigma}$. By the Schwarz inequality (2.3) and the bound $\sqrt{2}$ of the state space in HS norm (Remark 4.5 (b)) one has

$$|\langle \sigma - \rho, \widehat{\rho} \rangle - a| \leq \|\sigma - \rho\|_2 \|\widehat{\rho} - \rho\|_2 \leq \frac{\sqrt{2}}{n}$$

and similarly $|\langle \sigma - \rho, \widehat{\sigma} \rangle - b| \leq \frac{\sqrt{2}}{n}$. Hence, by the intermediate value theorem for a continuous real function, for each $\xi \in [a + \frac{\sqrt{2}}{n}, b - \frac{\sqrt{2}}{n}]$ there exists $\lambda \in [0, 1]$ such that $\langle \sigma - \rho, \tau(\lambda) \rangle = \xi$. We can define for each $\xi \in [a + \frac{\sqrt{2}}{n}, b - \frac{\sqrt{2}}{n}]$

$$\rho_n(\xi) := \tau(\lambda) \qquad \text{such that} \qquad \langle \sigma - \rho, \rho_n(\xi) \rangle = \xi.$$

Now let us fix $\xi \in (a, b)$. Then for sufficiently large $n \in \mathbb{N}$, a sequence of density matrices $\rho_n(\xi)$ is defined. By compactness of the variation closure $\overline{\mathcal{E}}$ there is a converging subsequence of $\rho_n(\xi)$ with limit $\rho_\infty(\xi) \in \overline{\mathcal{E}}$. By (10.12) one has $\pi_U(\rho_\infty(\xi)) = x$. On the other hand $\langle \sigma - \rho, \rho_\infty(\xi) \rangle = \xi$ proves that limit points $\rho_\infty(\xi)$ for mutually distinct values of ξ in the segment (a, b) are mutually distinct. qed

Proof of Lemma 7.24. If $p \in \overline{\mathcal{P}_U}$ is a non-zero projector then we can assume for a sequence $(p_i) \subset \mathcal{P}_U$ of non-zero projectors that $p = \lim_{i \to \infty} p_i$. For every $\rho \in \kappa^p(\mathcal{E}_p)$

$$\rho = \frac{p e^{p\theta p}}{\mathrm{tr}(p e^{p\theta p})}$$

holds for some $\theta \in \Theta$ with the canonical parameter space Θ of $\mathcal{E}$ (7.7). Similarly, for $i \in \mathbb{N}$ the density matrix

$$\rho_i := \frac{p_i e^{p_i \theta p_i}}{\mathrm{tr}(p_i e^{p_i \theta p_i})}$$

belongs to $\kappa^{p_i}(\mathcal{E}_{p_i})$ and then by Corollary 7.17 on obtains $(\rho_i) \subset \mathrm{cl}_{rI}(\mathcal{E})$. The limit of the sequence is $\rho = \lim_{i\to\infty} \rho_i$. By the inclusion $\mathrm{cl}_{rI}(\mathcal{E}) \subset \overline{\mathcal{E}}$ (7.27) and since $\overline{\mathcal{E}}$ is closed one obtains $\rho \in \overline{\mathcal{E}}$. `qed`

Proof of Proposition 7.26. Corollary 7.17 proves equality of rI-closure and combinatorial extension, $\mathrm{cl}_{rI}(\mathcal{E}) = \bigcup_{p\in\mathcal{P}_U\setminus\{0\}} \kappa^p(\mathcal{E}_p)$. On the other hand we know from Lemma 7.24 that $\overline{\mathcal{E}} \supset \bigcup_{p\in\overline{\mathcal{P}_U}\setminus\{0\}} \kappa^p(\mathcal{E}_p)$. Since the exponential families $\kappa^p(\mathcal{E}_p)$ and $\kappa^q(\mathcal{E}_q)$ are disjoint for distinct non-zero projectors $p, q \in \mathcal{P}$, the equality $\overline{\mathcal{E}} = \mathrm{cl}_{rI}(\mathcal{E})$ gives $\overline{\mathcal{P}_U} \subset \mathcal{P}_U$. So $\mathcal{P}_U$ is closed in this case. `qed`

Proof of Theorem 7. If $\overline{\mathcal{E}} \neq \mathrm{cl}_{rI}(\mathcal{E})$ then by (7.28) the reverse information closure

$$\mathrm{cl}_{rI}(\mathcal{E}) = \{\rho \in \overline{S}(A) : S_{\mathcal{E}}(\rho) = 0\}$$

is not closed in norm topology. Hence we can choose a sequence $\rho_i \in \mathrm{cl}_{rI}(\mathcal{E})$ with limit $\rho := \lim_{i\to\infty} \rho_i$ outside of $\mathrm{cl}_{rI}(\mathcal{E})$, that is $S_{\mathcal{E}}(\rho) > 0$ while $S_{\mathcal{E}}(\rho_i) = 0$ for $i \in \mathbb{N}$. As ρ belongs to the compact state space $\overline{S}(A)$, it is a point of discontinuity for $S_{\mathcal{E}}$.

Conversely, under the assumption $\overline{\mathcal{E}} = \mathrm{cl}_{rI}(\mathcal{E})$ we can prove in the first step that entropy distance is lower semi-continuous. Since $\overline{\mathcal{E}}$ is a compact subset of A_{sa}, lower semi-continuity of relative entropy (Remark 6.15) yields lower semi-continuity of the minimum

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

This can be proved using a covering of $\overline{\mathcal{E}}$ by open balls (Theorem 2 on page 116 in [Ber]). This minimum function is entropy distance. Indeed: Theorem 6 proves for $\rho \in \overline{S}(A)$ that $S_{\mathcal{E}}(\rho) = \min_{\sigma\in\mathcal{E}^{\mathrm{cmb}}} S(\rho,\sigma)$, Corollary 7.17 shows $\mathcal{E}^{\mathrm{cmb}} = \mathrm{cl}_{rI}(\mathcal{E})$ and with the assumption $\mathrm{cl}_{rI}(\mathcal{E}) = \overline{\mathcal{E}}$ one has

$$S_{\mathcal{E}}(\rho) = \min\{S(\rho,\sigma) : \sigma \in \overline{\mathcal{E}}\}.$$

Having established lower semi-continuity of entropy distance we follow the proof of Lemma 4.2 in [Ay02] (except the argument for lower semi-continuity) and we get continuity of entropy distance. This is as follows. We choose a density matrix $\rho \in \overline{S}(A)$ and an approximating sequence $\rho_n \in \overline{S}(A)$ such that $\rho = \lim_{n\to\infty} \rho_n$. For arbitrary $\sigma \in \mathcal{E} \subset S(A)$ and $n \in \mathbb{N}$ we have $S_{\mathcal{E}}(\rho_n) \leq S(\rho_n, \sigma)$. From continuity of $S(\cdot,\sigma)$ with the invertible density matrix σ follows $\lim_{n\to\infty} S(\rho_n, \sigma) = S(\rho, \sigma)$. The previous two equations combined give

$$\limsup_{n\to\infty} S_{\mathcal{E}}(\rho_n) \leq \limsup_{n\to\infty} S(\rho_n, \sigma) = S(\rho, \sigma).$$

The infimum over all $\sigma \in \mathcal{E}$ of the previous inequality is $\limsup_{n\to\infty} S_{\mathcal{E}}(\rho_n) \leq S_{\mathcal{E}}(\rho)$. Together with the lower semi-continuity of $S_{\mathcal{E}}$ established above one has

$$\limsup_{n\to\infty} S_{\mathcal{E}}(\rho_n) \leq S_{\mathcal{E}}(\rho) \leq \liminf_{n\to\infty} S_{\mathcal{E}}(\rho_n).$$

This shows $\lim_{n\to\infty} S_{\mathcal{E}}(\rho_n) = S_{\mathcal{E}}(\rho)$ and proves continuity of $S_{\mathcal{E}}$. qed

Proof of Proposition 7.29. Let $\rho \in \overline{S}(A)$ be fixed in the proof. We confine attention to the subset $K := F(\rho) \cap (\rho + U^{\perp})$ of the state space. The combinatorial normal projection is constant on K, for $\tau \in K$ we have (7.21) $N^{\mathrm{cmb}}(\tau) = N^{\mathrm{cmb}}(\rho)$. Thus, by Theorem 6 one has

$$S_{\mathcal{E}}(\tau) = S(\tau, N^{\mathrm{cmb}}(\rho))$$

and the function

$$K \to \mathbb{R}, \qquad \tau \mapsto S_{\mathcal{E}}(\tau) = -S(\tau) - \mathrm{tr}\left(\tau \ln\left(N^{\mathrm{cmb}}(\rho)\right)\right) \tag{10.13}$$

is strictly convex on K, because von Neumann entropy is strictly concave, see Remark 6.13. Let us assume that ρ is a local maximizer of $S_{\mathcal{E}}$. Then ρ is a local maximizer of $S_{\mathcal{E}}|_K$. Since $S_{\mathcal{E}}|_K$ is strictly convex on K and since the maximizer ρ belongs to the relative interior $\mathrm{ri}(K)$, one has $K = \{\rho\}$. This implies

$$\dim(F(\rho)) + \dim(U^{\perp}) \leq \dim(A_{\mathrm{sa}}) = \dim(U) + \dim(U^{\perp}).$$

The exponential family $\mathcal{E}$ has the same dimension as the canonical tangent space U. qed

Proof of Lemma 7.31. We set $q := s(F)$ for the support projector of F. Since $\dim(F) = \dim(\overline{S}(A)) - 1$ the complementary projector $q' := \mathbb{1} - q$ has rank one. As a consequence, the canonical parameter space of $\mathcal{E}$ is the translate of $\mathrm{lin}(F)$ by a multiple of $q' - q$. We can parametrize $\mathcal{E}$ for a fixed scalar $\lambda \in \mathbb{R}$ by

$$\rho : (A^q)^0_{\mathrm{sa}} \to \mathcal{E}, \qquad \vartheta \mapsto \exp_1(\kappa^q(\vartheta) + \lambda q').$$

For $\vartheta \in (A^q)^0_{\mathrm{sa}}$ we define $\sigma := \exp_1(\vartheta)$ and $c := \frac{\mathrm{tr}(e^{\vartheta})}{\mathrm{tr}(e^{\vartheta}) + e^{\lambda}}$, then

$$\rho(\vartheta) = c\kappa^q(\sigma) + (1 - c)q'$$

holds. We consider the density matrices for $\vartheta \in (A^q)^0_{\mathrm{sa}}$ in the relative interior of the face F

$$\tau(\vartheta) := \pi_{\mathrm{aff}(F)}(\rho(\vartheta)) = c\kappa^q(\sigma) + (1-c)\frac{q}{\mathrm{tr}(q)}.$$

The projection to the affine hull of the face is explained in (5.4). The mean value chart in Theorem 4 is a diffeomorphism $\pi : \mathcal{E} \to \mathrm{ri}(\mathrm{sr}_{\mathrm{lin}(F)})$. Then by the equation $\mathrm{sr}_{\mathrm{lin}(F)} = F - \frac{q}{\mathrm{tr}(q)}$ (5.5) there is diffeomorphism $\pi_{\mathrm{aff}(F)}|_{\mathcal{E}} : \mathcal{E} \to \mathrm{ri}(F)$. Thus

$$\tau : (A^q)^0_{\mathrm{sa}} \to \mathrm{ri}(F)$$

is a diffeomorphism. For each $\vartheta \in (A^q)^0_{\mathrm{sa}}$ the density matrix $\tau(\vartheta)$ has the same mean value with respect to $\mathrm{lin}(F)$ as the density $\rho(\vartheta) \in \mathcal{E}$ by construction. Thus we have determined the normal projection (6.27) $N(\tau(\vartheta)) = \rho(\vartheta)$ and the entropy distance from $\mathcal{E}$ is given by Lemma 6.22 as $S_{\mathcal{E}}(\tau(\vartheta)) = S(\tau(\vartheta), \rho(\vartheta))$. We introduce the abbreviations

$$f := (\kappa^q)^{-1}(\tau(\vartheta)) \qquad \text{and} \qquad x := \frac{1-c}{c}.$$

Then entropy distance is given for $\vartheta \in (A^q)^0_{\mathrm{sa}}$ by

$$a(\vartheta) := S_{\mathcal{E}}(\tau(\vartheta)) = \mathrm{tr}(f \ln(\mathbb{1}^q + \frac{x}{\mathrm{tr}(q)}\sigma^{-1})). \tag{10.14}$$

Here $\mathbb{1}^q$ denotes the identity in the compression A^q. We can show that the function $a : (A^q)^0_{\mathrm{sa}} \to \mathbb{R}$ has a local maximum at zero by calculation of two derivatives. We use the derivative of the logarithm (10.15) and of the normalized exponential (6.10). We have the derivatives along $u \in (A^q)^0_{\mathrm{sa}}$ at $\vartheta \in (A^q)^0_{\mathrm{sa}}$

$$\begin{aligned}
\tfrac{\partial}{\partial t}|_{t=0} c(\vartheta + tu) &= c(1-c)\,\mathrm{tr}(u\sigma), \\
\tfrac{\partial}{\partial t}|_{t=0} x(\vartheta + tu) &= -x\,\mathrm{tr}(u\sigma), \\
\tfrac{\partial}{\partial t}|_{t=0} f(\vartheta + tu) &= c\left(\int_0^1 \sigma^y u \sigma^{1-y}\,\mathrm{d}\,y - \mathrm{tr}(u\sigma) f\right)
\end{aligned}$$

and these expressions can be used to verify the first derivative of a

$$\frac{\partial}{\partial t}|_{t=0} a(\vartheta + tu) = c\,\mathrm{tr}((u\sigma - \mathrm{tr}(u\sigma)f)\ln(\mathbb{1}^q + \frac{x}{\mathrm{tr}(q)}\sigma^{-1})) - \mathrm{tr}(uf(\mathbb{1}^q + \frac{\mathrm{tr}(q)}{x}\sigma)^{-1}).$$

In particular, $\vartheta = 0$ is a critical point of a with $\sigma = \exp_1(\vartheta) = \frac{\mathbb{1}^q}{\mathrm{tr}(\mathbb{1}^q)}$. The second derivative of a at zero along $u, v \in (A^q)^0_{\mathrm{sa}}$ is

$$\frac{\partial^2}{\partial s\,\partial t}|_{s=t=0} a(su + tv) = -\frac{c(1-c)}{\mathrm{tr}(q)}\,\mathrm{tr}(uv).$$

The Hessian form is negative definite at zero, so a has a local maximum at zero, that is, entropy distance from $\mathcal{E}$ in restriction to the relative interior of F has a local maximum at the centroid $\frac{s(F)}{\mathrm{tr}(s(F))}$.

Since $\dim(F) = \dim(\overline{S}(A)) - 1$ it is easy to extend the local maximum $\frac{s(F)}{\mathrm{tr}(s(F))}$ on $\mathrm{ri}(F)$ to a local maximum on the state space $\overline{S}(A)$. On each of the segments for $\vartheta \in (A^q)^0_{\mathrm{sa}}$

$$K(\vartheta) := (\tau(\vartheta) + \mathrm{lin}(F)^\perp) \cap \overline{S}(A)$$

the entropy distance is a strictly convex function (10.13), because it is essentially the negative von Neumann entropy. On the interval

$$[\rho(\vartheta), \tau(\vartheta)] \subset K(\vartheta)$$

entropy distance is strictly monotone increasing because it is zero at $\rho(\vartheta)$ and positive for other points of the segment. `qed`

Proof of Lemma 8.2. Let $U \subset A^0_{\mathrm{sa}}$ be a vector space and put $\mathcal{E} := \exp_1(U)$. For a point $x \in \mathrm{ri}(\mathrm{sr}_U)$ we consider the problem of maximizing von Neumann entropy on $K := \overline{S}(A) \cap (x + U^\perp)$.

The von Neumann entropy is a strictly concave and continuous function on the state space $S : \overline{S}(A) \to \mathbb{R}$, see Remark 6.13 [Weh]. Since x belongs to the relative interior of the state reflection the set K intersects the relative interior of the state space because ri interchanges with affine mappings (3.15). Then we obtain that $\mathrm{ri}(K)$ is the intersection of K with $\mathrm{ri}(\overline{S}(A))$ (3.17). The relative interior of the state space consists of the invertible density matrices $S(A)$ (4.27) hence

$$\mathrm{ri}(K) = S(A) \cap K.$$

Since von Neumann entropy S is strictly concave on $\overline{S}(A)$, a local maximum of S on $\mathrm{ri}(K)$ is a unique global maximum on K. We can prove that a local maximum σ of S exists on $\mathrm{ri}(K)$. We prove that σ belongs to the exponential family $\mathcal{E} := \exp_1(U)$. Then by Corollary 6.30 it follows that σ is given by the mean value parametrization $\sigma = M(x)$.

We consider the relative interior $\mathrm{ri}(A^+)$ of the positive cone $A^+ = \{a \in A : a \geq 0\}$ (2.10) consisting of the invertible and positive matrices (4.14). This is an open subset of the space of self-adjoint matrices A_{sa} (4.11). The von Neumann entropy on this extended domain $S : \mathrm{ri}(A^+) \to \mathbb{R}$ is an analytic function. For arbitrary $a \in \mathrm{ri}(A^+)$ and self-adjoint $u \in A_{\mathrm{sa}}$ we have with (3.6) and (3.3) in [Lie]

$$\frac{\partial}{\partial t} \ln(a + tu)|_{t=0} = \int_0^\infty (a + s\mathbb{1})^{-1} u (a + s\mathbb{1})^{-1} \,\mathrm{d}\, s. \qquad (10.15)$$

This gives

$$\frac{\partial}{\partial t}S(a+tu)|_{t=0} = -\operatorname{tr}(u\ln(a)) - \operatorname{tr}(ua\int_0^\infty (a+s\mathbb{1})^{-2}\,\mathrm{d}\,s).$$

Using for $\mu > 0$ the equality $\int_0^\infty (\mu+s)^{-2}\,\mathrm{d}\,s = \frac{1}{\mu}$ the derivative of von Neumann entropy becomes

$$\frac{\partial}{\partial t}S(a+tu)|_{t=0} = -\langle u, \ln(a))\rangle - \operatorname{tr}(u). \tag{10.16}$$

By the mean value chart, see Lemma 6.22, there exists a density matrix $\sigma \in \mathcal{E} \cap \operatorname{ri}(K)$. Then $v := \ln_0(\sigma)$ belongs to U. The tangent space $\mathrm{T}_\sigma^{(m)} \operatorname{ri}(K) = U^\perp \cap A_{\mathrm{sa}}^0$ in m-representation consists of the traceless matrices in $U^\perp$. For each $u \in U^\perp \cap A_{\mathrm{sa}}^0$ we have

$$\frac{\partial}{\partial t}S(\sigma+tu)|_{t=0} = -\langle u, \ln(\sigma)\rangle - \operatorname{tr}(u) = -\langle u, v - F(v)\mathbb{1}\rangle = 0.$$

This proves that σ is a local maximum of S on $\operatorname{ri}(K)$. qed

Proof of Remark 8.3. Let $\{a_i\}_{i=1}^k \subset A_{\mathrm{sa}}$ be a finite number of self-adjoint matrices such that $a_1, \ldots, a_k, \mathbb{1}$ is a linear independent set. For an invertible density matrix $\rho \in S(A)$ with mean values $\xi_i := \langle a_i, \rho\rangle$ for $i = 1, \ldots, k$ we consider the constraint set

$$K := \{\sigma \in \overline{S}(A) : \langle a_j, \sigma\rangle = \xi_j \text{ for } j = 1, \ldots, k\}.$$

We consider an invertible local maximum $\sigma \in \operatorname{ri}(K)$ of von Neumann entropy S. (The relative interior of K is the set of invertible density matrices in K, see the proof of Lemma 8.2.) The constraint set K is described by the equations $g(a) = 0$ and $h_i(a) = 0$ for $i = 1, \ldots k$ where $g(a) := \operatorname{tr}(a) - 1$ and $h_i(a) := \langle a_i, a\rangle - \xi_i$ for $a \in A_{\mathrm{sa}}$. In addition there are the constraints defining the positive cone A^+ but since $\sigma \in \operatorname{ri}(A^+)$ is a relative interior point none of the inequalities defining the positive cone are active.

We set $x_0 := \frac{\mathbb{1}}{\sqrt{\operatorname{tr}(\mathbb{1})}}$ and take $x_1, \ldots, x_n$ an ONB of the space of traceless matrices A_{sa}^0. Then $x_0, \ldots, x_n$ is an ONB of A_{sa} for $n = \dim(A_{\mathrm{sa}}) - 1$. The gradient of a differentiable real function f defined on an open subset $Y \subset A_{\mathrm{sa}}$ at a point $y \in M$ is the vector $\nabla f(y) = \sum_{i=0}^n \frac{\partial}{\partial t}|_{t=0} f(y + tx_i)x_i \in A_{\mathrm{sa}}$.

The gradients of the constraint functions $\nabla g(\sigma) = \mathbb{1}$ and $\nabla h_i(\sigma) = a_i$ for $i = 1, \ldots, k$ are linearly independent by assumption. Thus we meet the conditions of Theorem 5.8 in [Ja] which states that a local maximum σ of von Neumann entropy satisfies for some real β_i, $i = 1, \ldots, k$ and real μ the equation

$$\nabla S(\sigma) + \mu \nabla g(\sigma) - \sum_{i=1}^k \beta_i \nabla h_i(\sigma) = 0. \tag{10.17}$$

By custom in thermodynamics we take the multipliers β_i negative. The gradient of von Neumann entropy is $\nabla S(\sigma) = -\ln(\sigma) - \mathbb{1}$ by (10.16) on page 241. Then by (10.17) we get $\ln(\sigma) = -\sum_{i=1}^{k}\beta_i a_i + (\mu - 1)\mathbb{1}$ and we find that a maximum must be of the form

$$\sigma = \exp\left(-\sum_{i=1}^{k}\beta_i a_i\right) e^{\mu-1}.$$

Since $\mathrm{tr}(\sigma) = 1$ we obtain (8.2),

$$\sigma = \frac{e^{-\sum_{i=1}^{k}\beta_i a_i}}{\mathrm{tr}\left(e^{-\sum_{i=1}^{k}\beta_i a_i}\right)}.$$

qed

Proof of Theorem 8. Given $x \in \mathrm{sr}_U$ there is a unique non-zero projector $p \in \mathcal{P}_U$ such that $x \in \mathrm{ri}(\mathbb{F}_U(p))$ (5.56). Let us choose a density matrix $\rho \in \overline{S}(A) \cap (x + U^\perp)$. Then $\overline{S}(A) \cap (x + U^\perp) = \overline{S}(A) \cap (\rho + U^\perp)$. By Proposition 7.9 the set $\overline{S}(A) \cap (x + U^\perp)$ belongs to $\kappa^p(\mathrm{Dom}_p)$. We can put $\rho_p := (\kappa^p)^{-1}(\rho)$ and consider maximization of von Neumann entropy on

$$\overline{S}(A^p) \cap (\rho_p + \varsigma^p(U)^\perp) = (\kappa^p)^{-1}\left(\overline{S}(A) \cap (\rho + U^\perp)\right) \subset \mathrm{Dom}_p$$

instead. The previous equation is proved by Proposition 5.27 and (4.33). The density matrix of maximum von Neumann entropy in $\overline{S}(A^p) \cap (\rho_p + \varsigma^p(U)^\perp) \subset \mathrm{Dom}_p$ is

$$\sigma_p := M_p\left(\pi_{\varsigma^p(U)}(\rho_p)\right)$$

by Lemma 8.2. We obtain from Proposition 7.9 and (7.20)

$$\begin{aligned}\kappa^p(\sigma_p) &= \kappa^p \circ M_p \circ \pi_{\varsigma^p(U)}(\rho_p)\\ &= \left(\pi_U|_{\kappa^p(\mathcal{E}_p)}\right)^{-1} \circ \pi_U \circ \kappa^p(\rho_p)\\ &= M^{\mathrm{cmb}} \circ \pi_U(\rho) = M^{\mathrm{cmb}}(x).\end{aligned}$$

qed

Proof of Proposition 8.6. We assume that $\rho = \lim_{i\to\infty}\rho_i$ for a sequence $(\rho_i) \subset \mathcal{E}$ and put $\theta_i := \ln_0(\rho_i) \in \Theta$ for the canonical parameter values, $i \in \mathbb{N}$. Then for $i \in \mathbb{N}$

$$\rho_i = \frac{\exp(\theta_i)}{\mathrm{tr}(\exp(\theta_i))}$$

holds. For a non-zero projector $p \in \mathcal{P}$, we define a sequence of trace one matrices for $i \in \mathbb{N}$

$$\sigma_i := \frac{p\exp(\theta_i)}{\mathrm{tr}(p\exp(\theta_i))}. \tag{10.18}$$

Using the fact that the algebra is abelian, the σ_i's are density matrices and with (7.7) they belong to the compression $\kappa^p(\mathcal{E}_p)$, for $i \in \mathbb{N}$

$$\sigma_i = \frac{p \exp(p\theta_i p)}{\mathrm{tr}(p \exp(p\theta_i p))} \in \kappa^p(\mathcal{E}_p) \tag{10.19}$$

holds. We use two arguments for the proof of the proposition. The first one works in an arbitrary matrix algebra; we use the σ_i's in (10.18). We prove under the assumption $p \geq s(\rho)$ that $\lim_{i\to\infty} \sigma_i = \rho$. The second argument works only for an abelian algebra; we use the σ_i's in (10.19). We show under the assumption $\pi_U(\rho) \in \mathrm{ri}(\mathbb{F}_U(p))$ that $\lim_{i\to\infty} \sigma_i \in \kappa^p(\mathcal{E}_p)$. This is done by controlling the σ_i's through their mean values. Let us argue how the two arguments combine to a complete proof of the proposition. The projector

$$p^* := \bigwedge \{q \in \mathcal{P}_U : q \geq s(\rho)\}$$

satisfies by Proposition 7.11 and Proposition 7.9 the assumptions of both arguments. We obtain

$$\rho \in \kappa^{p^*}(\mathcal{E}_{p^*}) \subset \mathrm{cl}_{rI}(\mathcal{E}),$$

as Corollary 7.17 proves $\mathcal{E}^{\mathrm{cmb}} = \mathrm{cl}_{rI}(\mathcal{E})$. We conclude with the Pinsker-Csiszár inequality (7.27) that $\mathrm{cl}_{rI}(\mathcal{E}) = \overline{\mathcal{E}}$.

Assuming $p \geq s(\rho)$ we prove that the σ_i's converge to ρ for $i \to \infty$. One has for $i \in \mathbb{N}$

$$\sigma_i = \frac{\mathrm{tr}(\exp(\theta_i))}{\mathrm{tr}(p \exp(\theta_i))} p\rho_i.$$

Using the expression $\theta_i = \ln(\rho_i) - \frac{\mathrm{tr}(\ln(\rho_i))}{\mathrm{tr}(\mathbb{1})}\mathbb{1}$ and the equation $\exp(\mathrm{tr}(a)) = \det(\exp(a))$ for arbitrary $a \in A$ (Satz 8.28 in [Kn01]), one obtains

$$\exp(\theta_i) = \rho_i \det(\rho_i)^{-\frac{1}{\mathrm{tr}(\mathbb{1})}}$$

and this gives us

$$\frac{\mathrm{tr}(\exp(\theta_i))}{\mathrm{tr}(p \exp(\theta_i))} = \frac{1}{\mathrm{tr}(p\rho_i)}.$$

Hence

$$\sigma_i = \frac{1}{\mathrm{tr}(p\rho_i)} p\rho_i \stackrel{i\to\infty}{\longrightarrow} \frac{1}{\mathrm{tr}(p\rho)} p\rho = \rho.$$

Secondly, assuming convergence of the σ_i's and provided that $\pi_U(\rho) \in \mathrm{ri}(\mathbb{F}_U(p))$, we show that the limit of the σ_i's is included in $\kappa^p(\mathcal{E}_p)$. The converging sequence $(\pi_U(\sigma_i))$ of mean values is finally in a compact subset C of $\mathrm{ri}(\mathbb{F}_U(p))$ (with $\pi_U(\rho) \in C$). Using the diffeomorphism $\pi_U|_{\kappa^p(\mathcal{E}_p)} : \kappa^p(\mathcal{E}_p) \to \mathrm{ri}(\mathbb{F}_U(p))$ remarked in (7.10), the continuity of the inverse (combinatorial mean value parametrization) shows that $\lim_{i\to\infty} \sigma_i \in \kappa^p(\mathcal{E}_p)$. `qed`

Proof of Corollary 8.7. This follows from Proposition 8.6 and Theorem 7. qed

Proof of Lemma 8.11. This is a proof by induction. The first formula is trivial for one element sets $I = \{i\}$. Assume the formula to be true for two disjoint non-empty subsets $I, J \subset [N]$ and $a_i \in A_i$ for $i \in I \cup J$. Since the matrices $(a \cdot I) \otimes \mathbb{1}_J$ and $(a \cdot J) \otimes \mathbb{1}_I$ commute one has by (8.7)

$$\exp(a \cdot (I \cup J)) = \exp((a \cdot I) \otimes \mathbb{1}_J) \exp((a \cdot J) \otimes \mathbb{1}_I).$$

By distributive law we have $\exp((a \cdot I) \otimes \mathbb{1}_J) = \exp(a \cdot I) \otimes \mathbb{1}_J$, so the induction step is complete and we get the proposed result. For non-empty $I \subset [N]$ and $a_i \in A_i$ follows with (8.5) the equation $\exp_1(a \cdot I) = \frac{\bigotimes_{i \in I} \exp(a_i)}{\operatorname{tr}(\bigotimes_{i \in I} \exp(a_i))} = \frac{\bigotimes_{i \in I} \exp(a_i)}{\prod_{i \in I} \operatorname{tr}(\exp(a_i))} = \bigotimes_{i \in I} \exp_1(a_i)$. qed

Proof of Lemma 8.14. Recall from Definition 7.20 that the variation closure of an exponential family is the closure in norm topology. For $I = \{i\}$ and $i \in [N]$ the statement is $\overline{S(A_i)} = \overline{S}(A_i)$ and this is proved in (4.27).

For larger sizes of I we first prove that the Kronecker product is jointly continuous. This follows by induction from the following statement. If $I, J \subset [N]$ are disjoint then for $a, c \in A_I$ and $b, d \in A_J$ we have

$$\begin{aligned} \|a \otimes b - c \otimes d\|_2 &= \|a \otimes (b - d) + (a - c) \otimes d\|_2 \\ &\leq \|a \otimes (b - d)\|_2 + \|(a - c) \otimes d\|_2 \\ &= \|a\|_2 \|b - d\|_2 + \|a - c\|_2 \|d\|_2. \end{aligned}$$

Let us prove for $I \subset [N]$ that $\{\rho_I : \rho_i \in \overline{S}(A_i), i \in I\} \subset \overline{\mathcal{F}(I)}$. A matrix in the left-hand side can be written $\sigma = \rho_I$ for $\rho_i \in \overline{S}(A_i)$, $i \in I$. For $i \in I$ each of the factors ρ_i is the limit of a sequence $(\rho_j)_i \subset S(A_i)$, $j \in \mathbb{N}$, and $\rho_I = \lim_{j \to \infty} (\rho_j)_I$ by joint continuity of the Kronecker product. Since $(\rho_j)_I$ belong to $\mathcal{F}(I)$ the first inclusion is proved.

The inclusion $\overline{\mathcal{F}(I)} \subset \{\rho_I \ : \ \rho_i \in \overline{S}(A_i), \ i \in I\}$ follows from joint continuity of the Kronecker product. If $\sigma \in \overline{\mathcal{F}(I)}$ then $\sigma = \lim_{j \to \infty} \sigma_j$ for a sequence $\sigma_j \subset \mathcal{F}(I)$, $j \in \mathbb{N}$. By definition (8.8) of the factorizable family one has $\sigma_j = (\rho_j)_I$ for $(\rho_j)_i \in S(A_i)$, $i \in I$, $j \in \mathbb{N}$. For $i \in I$, each of the sequences $(\rho_j)_i$ has an accumulation point ρ_i in $\overline{S}(A_i)$ and by transition to a suitable subsequence we can assume convergence $(\rho_j)_i \overset{j \to \infty}{\longrightarrow} \rho_i \in \overline{S}(A_i)$. This proves $\sigma = \lim_{j \to \infty} \sigma_j = \lim_{j \to \infty} (\rho_j)_I = \rho_I$. qed

Proof of Proposition 8.15. For a density matrix $\rho \in \overline{S}(A)$ it is equivalent by Theorem 5 whether ρ belongs to the closure of an e-geodesic in $\mathcal{F}(I)$ or ρ belongs to a compression $\kappa^p(\mathcal{F}(I)_p)$ for a non-zero projector $p \in \mathcal{P}_{\chi(I),\perp}$.

One direction of the proof $\overline{\mathcal{F}(I)} = \bigcup_{\substack{p\in\mathcal{P}_{\chi(I),\perp}\\ p\neq 0}} \kappa^p(\mathcal{F}(I)_p)$ is obtained by the observation that the closure of an e-geodesic in $\mathcal{F}(I)$ belongs to the variation closure $\overline{\mathcal{F}(I)}$. For the converse, it is sufficient to prove that every point $\sigma \in \overline{\mathcal{F}(I)}$ belongs to the closure of an e-geodesic in $\mathcal{F}(I)$. For a one-element set $I = \{i\} \subset [N]$ this follows from the regularity of the exponential family $S(A_i)$ of invertible density matrices described in Remark 7.27 (b). In the general case a density matrix $\sigma \in \overline{\mathcal{F}(I)}$ can be written $\sigma = \rho_I$ for $\rho_i \in \overline{S}(A_i)$, $i \in I$, by Lemma 8.14. As discussed in the remark, for each $i \in I$ there exist $\theta_i, u_i \in (A_i)^0_{\mathrm{sa}}$ such that $\rho_i = \lim_{\lambda\to\infty} \exp_1(\theta_i + \lambda u_i)$. One has $u \cdot I, \theta \cdot I \in \chi(A_I)$ so the curve

$$\mathbb{R} \to S(A_I), \qquad \lambda \mapsto \exp_1(\theta \cdot I + \lambda(u \cdot I))$$

is an e-geodesic in $\mathcal{F}(I)$. Since $\theta \cdot I + \lambda(u \cdot I) = (\theta + \lambda u) \cdot I$ by (8.6), one obtains from the functional equation of the exponential function (proved in Lemma 8.11) and from joint continuity of the Kronecker product that

$$\lim_{\lambda\to\infty} g(\lambda) = \lim_{\lambda\to\infty} \bigotimes_{i\in I} \exp_1(\theta_i + \lambda u_i) = \bigotimes_{i\in I} \rho_i = \sigma. \qquad \texttt{qed}$$

Proof of Lemma 8.16. We use spectral decomposition. One has

$$a_I = \bigotimes_{i\in I} \sum_{\mu_i\in\mathrm{spec}(a_i)} \mu_i p_{\mu_i}(a_i) = \sum_{\substack{\mu_i\in\mathrm{spec}(a_i)\\ i\in I}} (\prod_{j\in I} \mu_j) \bigotimes_{j\in I} p_{\mu_j}(a_j).$$

The orthogonal projectors $\left\{\bigotimes_{j\in I} p_{\mu_j}(a_j)\right\}_{\mu_j\in\mathrm{spec}(a_j), j\in I}$ are mutually orthogonal, and one of them contributes to the above sum if and only if $\prod_{j\in I} \mu_j > 0$. So

$$s(a_I) = \sum_{\substack{\mu_i\in\mathrm{spec}(a_i)\setminus\{0\}\\ i\in I}} \bigotimes_{j\in I} p_{\mu_j}(a_j) = \bigotimes_{i\in I} \sum_{\mu_i\in\mathrm{spec}(a_i)\setminus\{0\}} p_{\mu_i}(a_i) = \bigotimes_{i\in I} s(a_i).$$

The Kronecker sum transforms as

$$a \cdot I = \sum_{i\in I} \sum_{\substack{\mu_j\in\mathrm{spec}(a_j)\\ j\in I}} \mu_i \bigotimes_{k\in I} p_{\mu_k}(a_k) = \sum_{\substack{\mu_j\in\mathrm{spec}(a_j)\\ j\in I}} (\sum_{i\in I} \mu_i) \bigotimes_{k\in I} p_{\mu_k}(a_k).$$

Then $\max_{\substack{\mu_j\in\mathrm{spec}(a_j)\\ j\in I}} (\sum_{i\in I} \mu_i) = \sum_{i\in I} \max_{\mu_i\in\mathrm{spec}(a_i)} \mu_i$ completes the proof. qed

Proof of Lemma 8.17. The two projector lattices $\mathcal{P}_{\chi(I),\perp}$ and $\mathcal{P}_{\chi(I)}$ are equal by Proposition 8.15 and Remark 7.27. We can calculate the exposed projector lattice $\mathcal{P}_{\chi(I),\perp}$ with (5.16), it consists of the maximal projectors of elements of $\chi(I)$ and 0. The maximal projectors are calculated in Lemma 8.16. qed

Proof of Proposition 8.18. Let $a_i \in (A_i)^0_{\rm sa}$ for $i \in I$. Then by (5.3) we have

$$\begin{aligned}\pi_{\mathrm{lin}(\mathbb{F}(p_I))}(a \cdot I) &= p_I(a \cdot I)p_I - \tfrac{\mathrm{tr}(p_I(a \cdot I))}{\mathrm{tr}(p_I)} p_I \\ &= \sum_{i \in I} \left(p_i a_i p_i \otimes \bigotimes_{j \in I \setminus \{i\}} p_j - \tfrac{\mathrm{tr}(p_i a_i)}{\mathrm{tr}(p_i)} \bigotimes_{j \in I} p_j \right) \\ &= p_I \sum_{i \in I} \pi_{\mathrm{lin}(\mathbb{F}(p_i))}(a_i) \otimes \bigotimes_{j \in I \setminus \{i\}} \mathbb{1}_j .\end{aligned}$$

For all $i \in I$ we have $\pi_{\mathrm{lin}(\mathbb{F}(p_i))}((A_i)^0_{\rm sa}) = \mathrm{lin}(\mathbb{F}(p_i))$ so

$$\pi_{\mathrm{lin}(\mathbb{F}(p_I))}(\chi(I)) = \{ p_I(\theta \cdot I) \ : \ \theta_i \in \mathrm{lin}(\mathbb{F}(p_i)), \ i \in I \}$$

is proved. For the compression $B_i := A_i^{p_i}$ and $\theta_i \in \mathrm{lin}(\mathbb{F}(p_i))$ we put $\vartheta_i := (\kappa^{p_i})^{-1}(\theta_i) \in (B_i)^0_{\rm sa}$, $i \in I$. Then using the *-isomorphism $\kappa^{p_I} := \bigotimes_{i \in I} \kappa^{p_i} : B_I \to p_I A_I p_I$ one has

$$\begin{aligned}(\kappa^{p_I})^{-1}(p_I(\theta \cdot I)) &= \left(\textstyle\bigotimes_{i \in I} (\kappa^{p_i})^{-1} \right) \left(\textstyle\sum_{j \in I} p_{I \setminus \{j\}} \otimes \theta_j \right) \\ &= \textstyle\sum_{j \in I} \left(\textstyle\bigotimes_{i \in I \setminus \{j\}} (\kappa^{p_j})^{-1}(p_j) \right) \otimes (\kappa^{p_j})^{-1}(\theta_j) = \vartheta \cdot I\end{aligned}$$

with the last Kronecker sum calculated in the algebra B_I. Since $(\kappa^{p_i})^{-1}(\mathrm{lin}(\mathbb{F}(p_i))) = (A_i^{p_i})^0_{\rm sa} = (B_i)^0_{\rm sa}$ (4.37) one has

$$\varsigma^{p_I}(\chi(A_I)) = \left\{ \vartheta \cdot I \ : \ \vartheta_i \in (B_i)^0_{\rm sa}, \ i \in I \right\} = \chi(B_I).$$

As a result, $\mathcal{F}(A_I)_{p_I} = \exp_1(\varsigma^{p_I}(\chi(A_I))$ is a factorizable family. Elements of the image $\kappa^{p_I}(\mathcal{F}(A_I)_{p_I})$ may be expressed (7.7) independent of the choice of *-monomorphisms κ^{p_i} with $\theta_i \in \mathrm{lin}(\mathbb{F}(p_i))$, $i \in I$, by

$$\frac{p_I e^{p_I(\theta \cdot I)}}{\mathrm{tr}\left(p_I e^{p_I(\theta \cdot I)}\right)}.$$

Since p_I commutes with $\theta \cdot I$ one has

$$p_I \exp(\theta \cdot I) \exp(p_I(\theta \cdot I))^{-1} = p_I \exp\left((\mathbb{1} - p_I)(\theta \cdot I)\right) = p_I$$

and $p_I e^{p_I(\theta \cdot I)}$ simplifies to $p_I e^{\theta \cdot I}$. qed

Proof of Lemma 8.20. Let $I \subset [N]$, let $a \in A_I$ and for $i \in I$ let $u_i \in (A_i)^0_{\rm sa}$. If for $i \in I$ we set $b_i := \mathrm{tr}_{\{i\}}\left(\pi_{(A_I)^1_{\rm sa}}(a)\right)$ then one has $\mathrm{tr}(b_i) = 1$ and through (8.11) one obtains

$$\begin{aligned}\mathrm{tr}\left((u \cdot I)a\right) &= \textstyle\sum_{i \in I} \mathrm{tr}\left((\mathbb{1}_{I \setminus \{i\}} \otimes u_i)a\right) = \textstyle\sum_{i \in I} \mathrm{tr}\left(u_i \,\mathrm{tr}_{\{i\}}(a)\right) \\ &= \textstyle\sum_{i \in I} \mathrm{tr}\left(u_i \pi_{(A_i)^1_{\rm sa}}(\mathrm{tr}_{\{i\}}(a))\right) \prod_{j \in I \setminus \{i\}} \mathrm{tr}(b_j) \\ &= \mathrm{tr}\left(\textstyle\sum_{i \in I} u_i b_i \bigotimes_{j \in I \setminus \{i\}} b_j\right) = \mathrm{tr}\left(\textstyle\sum_{i \in I}(u_i \otimes \mathbb{1}_{I \setminus \{i\}})(\bigotimes_{j \in I} b_j)\right) \\ &= \mathrm{tr}\left((u \cdot I)b_I\right).\end{aligned}$$

Notice that $\pi_{(A_i)^1_{\rm sa}}(\mathrm{tr}_{\{i\}}(a))$ differs from $\mathrm{tr}_{\{i\}}(a)$ by a multiple of the identity $\mathbb{1}_i$ in A_i. qed

Proof of Lemma 8.21. The density matrix $\sigma := \bigotimes_{i\in I} \mathrm{tr}_{\{i\}}(\rho)$ belongs to the combinatorial closure $\mathcal{F}(I)^{\mathrm{cmb}}$ of the factorizable family by Lemma 8.14 and by Proposition 8.15. By Lemma 8.20 one has equality of mean values with respect to the canonical tangent space $\chi(I)$

$$\pi_{\chi(I)}(\sigma) = \pi_{\chi(I)}(\rho).$$

Density matrices in $\overline{S}(A)$ with the same mean value have the same value of combinatorial normal projection (7.14), so $N^{\mathrm{cmb}}(\rho) = \sigma$. qed

Proof of Theorem 9. Let $I \subset [N]$ and $\rho \in \overline{S}(A_I)$. We notice by Theorem 6 that the infimum of relative entropy $\inf_{\sigma\in\mathcal{F}(I)} S(\rho,\sigma)$ is assumed for the combinatorial normal projection $\tau := N^{\mathrm{cmb}}(\rho)$. By Lemma 8.21 this density matrix is the factorizable matrix

$$\tau = \bigotimes_{i\in I} \mathrm{tr}_{\{i\}}(\rho).$$

Before evaluation of relative entropy we have to discuss the kernels of the matrices ρ and τ. Since τ is the combinatorial normal projection of ρ, the support projector is by Proposition 7.11 equal to $p := s(\tau) = \bigwedge\{q \in \mathcal{P}_{\chi(I)} : q \geq s(\rho)\}$. In particular, $s(\tau) \geq s(\rho)$ makes the relative entropy $S(\rho,\tau)$ a finite number. The support projector p of τ is actually factorizable, by Lemma 8.16 it is

$$p = \bigotimes_{i\in I} s(\mathrm{tr}_{\{i\}}(\rho)).$$

Let us set $q_i := s(\mathrm{tr}_{\{i\}}(\rho)) \in \mathcal{P}(A_i)$ for $i \in I$, then $p = q_I$. Next we discuss a technical detail. We choose trace preserving *-monomorphism $\kappa^{q_i} : (A_i)^{q_i} \to A_i$ for $i \in I$ and we put $\kappa^p := \bigotimes_{i\in I} \kappa^{q_i}$. It is easy to prove for a factorizable matrix $a_I \in A_I$ defined by $a_i \in A_I$ for $i \in I$ that

$$\mathrm{tr}_{\{i\}} \circ (\kappa^p)^{-1}(a_I) = (\kappa^{q_i})^{-1} \circ \mathrm{tr}_{\{i\}}(a_I)$$

holds for $i \in I$. This is a linear equation in a_I hence it generalizes to arbitrary matrices in A_I. One has

$$S(\rho,\tau) + S(\rho) = -\,\mathrm{tr}\left((\kappa^p)^{-1}(\rho)\ln\left((\kappa^p)^{-1}(\tau)\right)\right).$$

We can discuss the logarithm with Lemma 8.11 and find

$$\ln\left((\kappa^p)^{-1}(\tau)\right) = \ln\left(\bigotimes_{i\in I}(\kappa^{q_i})^{-1}\left(\mathrm{tr}_{\{i\}}(\rho)\right)\right) = \sum_{i\in I} \mathbb{1}_{I\setminus\{i\}} \otimes \ln\left((\kappa^{q_i})^{-1}\left(\mathrm{tr}_{\{i\}}(\rho)\right)\right).$$

Then using (8.11) we obtain that $S(\rho,\tau) + S(\rho)$ is equal

$$\begin{aligned} & -\textstyle\sum_{i\in I} \mathrm{tr}\left((\kappa^p)^{-1}(\rho)\left(\mathbb{1}_{I\setminus\{i\}} \otimes \ln\left((\kappa^{q_i})^{-1}(\mathrm{tr}_{\{i\}}(\rho))\right)\right)\right) \\ = & \ -\textstyle\sum_{i\in I} \mathrm{tr}\left(\mathrm{tr}_{\{i\}}((\kappa^p)^{-1}(\rho))\ln\left((\kappa^{q_i})^{-1}(\mathrm{tr}_{\{i\}}(\rho))\right)\right) \\ = & \ \textstyle\sum_{i\in I} S\left((\kappa^{q_i})^{-1}(\mathrm{tr}_{\{i\}}(\rho))\right) = \sum_{i\in I} S(\mathrm{tr}_{\{i\}}(\rho)). \end{aligned}$$

qed

Proof of Lemma 8.26. The projection $\pi_U|_{\mathrm{aff}(\mathcal{E})} : \mathrm{aff}(\mathcal{E}) \to U$ is an affine mapping. It is surjective because $\mathrm{ri}(\mathrm{sr}_U) = \pi_U(\mathcal{E})$ by Theorem 4 and since the state reflection sr_U has non-empty interior in U. Since $\mathcal{E}$ is convex one has $\dim(\mathrm{aff}(\mathcal{E})) = \dim(\mathcal{E}) = \dim(U)$. This shows that $\pi_U|_{\mathrm{aff}(\mathcal{E})}$ is an affine isomorphism.

For the proof of $\mathcal{E} = \mathrm{aff}(\mathcal{E}) \cap S(A)$ we use the equation $\pi_U(\mathcal{E}) = \mathrm{ri}(\mathrm{sr}_U) = \pi_U(S(A))$ where the second equality is true by definition $\mathrm{sr}_U = \pi_U(\overline{S}(A))$ of the sate reflection and because taking a relative interior commutes with π_U (5.17). We can use the affine isomorphism from the first part and obtain for arbitrary $\rho \in \mathrm{aff}(\mathcal{E})$

$$\rho \in \mathcal{E} \qquad \Longleftrightarrow \qquad \pi_U(\rho) \in \mathrm{ri}(\mathrm{sr}_U) \qquad \Longleftrightarrow \qquad \rho \in S(A).$$

For a proof of $\overline{\mathcal{E}} = \mathrm{aff}(\mathcal{E}) \cap \overline{S}(A)$ we observe that the relative open convex sets $S(A)$ and $\mathrm{aff}(\mathcal{E})$ share a point. So $\mathcal{E}$ is a relative open convex set and the closure distributes over intersection (3.17),

$$\overline{\mathcal{E}} = \overline{\mathrm{aff}(\mathcal{E}) \cap S(A)} = \mathrm{aff}(\mathcal{E}) \cap \overline{S}(A).$$

qed

Proof of Proposition 8.27. All faces of the state space $\overline{S}(A)$ are exposed by Proposition 4.18. We have found in Lemma 8.26 that

$$\overline{\mathcal{E}} = \mathrm{aff}(\mathcal{E}) \cap \overline{S}(A)$$

is the intersection of an affine space with the sate space. Under these assumptions it is proved in Lemma 3.50 that all faces of the compact convex set $\overline{\mathcal{E}}$ are exposed. In Theorem 4 we can prove that $\mathrm{sr}_U = \pi_U(\overline{\mathcal{E}})$. Since $\overline{\mathcal{E}} \subset \mathrm{aff}(\mathcal{E})$ the affine isomorphism $\pi_U|_{\mathrm{aff}(\mathcal{E})} : \mathrm{aff}(\mathcal{E}) \to U$ established in Lemma 8.26 restricts to $\overline{\mathcal{E}}$ and the state reflection sr_U is affinely isomorphic to $\overline{\mathcal{E}}$. Thus it follows that all faces of the state reflection sr_U are exposed, one has equality of the projector lattices

$$\mathcal{P}_U = \mathcal{P}_{U,\perp}.$$

Another consequence of the restricted affine isomorphism $\pi_U|_{\overline{\mathcal{E}}} : \overline{\mathcal{E}} \to \mathrm{sr}_U$ is its injectivity. Following the discussion in Remark 7.23 (c) one obtains equality of $\overline{\mathcal{E}} = \mathrm{cl}_{rI}(\mathcal{E})$. The equality $\mathrm{cl}_{rI}(\mathcal{E}) = \mathcal{E}^{\mathrm{cmb}}$ from Corollary 7.17 proves $\mathrm{cl}_{rI}(\mathcal{E}) = \bigcup_{p \in \mathcal{P}_{U,\perp} \setminus \{0\}} \kappa^p(\mathcal{E}_p)$. qed

Proof of Lemma 8.29. Assume $\theta \in \Theta$ is such that $[\theta, u] = 0$ for all $u \in U$ and that $U + \mathbb{R}\mathbb{1}$ contains the spectral projectors of all matrices in U. We prove that $\mathcal{E}$ is closed under convex combinations by showing for arbitrary $u, v \in U$ and $\lambda \in [0,1]$ that the density matrix

$$\rho := \lambda \exp_1(\theta + u) + (1-\lambda) \exp_1(\theta + v)$$

belongs to $\mathcal{E}$. In the canonical chart (6.2) this means $\ln_0(\rho) \in \Theta$. Since $[\theta, u] = [\theta, v] = 0$ one has $[e^\theta, e^u] = 0$ by Remark 8.28 (b). The functional equations for commuting matrices of exponential function and logarithm give

$$\ln(\rho) = \ln\left(\lambda \frac{e^{\theta+u}}{\operatorname{tr}(e^{\theta+u})} + (1-\lambda)\frac{e^{\theta+v}}{\operatorname{tr}(e^{\theta+v})}\right) = \theta + \ln\left(\lambda \frac{e^{u}}{\operatorname{tr}(e^{\theta+u})} + (1-\lambda)\frac{e^{v}}{\operatorname{tr}(e^{\theta+v})}\right).$$

One has $\ln(\rho) - \theta \in U + \mathbb{R}\mathbb{1}$ and then $\ln_0(\rho) - \theta \in U$, that is $\ln_0(\rho) \in \Theta$. qed

Proof of Proposition 8.31. Let $u \in U$ and assume a matrix $\theta \in \Theta$ commutes with u. Then the maximal projector $p := p_+(u)$ of u commutes with θ by Remark 8.28 (b). We consider the traceless self-adjoint matrix $\widetilde{u} := p \operatorname{tr}(p') - p' \operatorname{tr}(p)$. Since $[\theta, \widetilde{u}] = 0$ we obtain from Remark 8.30 that the e-geodesic

$$\widetilde{g} : \mathbb{R} \to S(A), \qquad \lambda \mapsto \exp_1(\theta + \lambda \widetilde{u})$$

is convex, it is a straight line segment in the space $S(A)$ of invertible density matrices. The e-geodesic

$$g : \mathbb{R} \to \mathcal{E}, \qquad \lambda \mapsto \exp_1(\theta + \lambda u)$$

is included in $\mathcal{E}$. With Lemma 7.6 one can compute the limit

$$\widetilde{g}(\infty) := \lim_{\lambda\to\infty} \widetilde{g}(\lambda) = \lim_{\lambda\to\infty} g(\lambda) \in \overline{\mathcal{E}}$$

which is the same for the two geodesics because u and $\widetilde{u}$ have the same maximal projector.

The closure $\overline{\mathcal{E}}$ of the convex family $\mathcal{E}$ is convex (Theorem 6.2 in [Ro]). The closed segment $\overline{\widetilde{g}}$ intersects $\overline{\mathcal{E}}$ in the distinct points $\widetilde{g}(0)$ and $\widetilde{g}(\infty)$ thus

$$[\widetilde{g}(0), \widetilde{g}(\infty)] \subset \overline{\mathcal{E}}.$$

In particular $\widetilde{g}(1) \in \overline{\mathcal{E}}$ and this implies $\widetilde{g}(1) \in \mathcal{E}$. Otherwise by the mean value chart in Theorem 4 follows from $\widetilde{g}(1) \in \overline{\mathcal{E}} \setminus \mathcal{E}$ that $\pi_U(\widetilde{g}(1)) \in \operatorname{rb}(\operatorname{sr}_U)$. This is a contradiction because the density matrix $\widetilde{g}(1)$ is invertible and has mean value in the relative interior $\operatorname{ri}(\operatorname{sr}_U)$ by Remark 5.5 (e). Now $\widetilde{g}(1) \in \mathcal{E}$ implies that

$$\theta + \widetilde{u} = \ln_0(\widetilde{g}(1)) \in \Theta,$$

so $p \operatorname{tr}(p') - p' \operatorname{tr}(p) = \widetilde{u} \in U$ and then $p \in U + \mathbb{R}\mathbb{1}$.

We have proved above that the maximal projector $p_+(u)$ belongs to the vector space $U + \mathbb{R}\mathbb{1}$. It follows by induction on the cardinality of the spectrum of u that all spectral projectors of u belong to $U + \mathbb{R}\mathbb{1}$. qed

Proof of Lemma 9.5. To prove surjectivity of J let $p \in P(\Omega \times \Omega)$ and consider the first marginal p_1 of p. For each $\omega' \in \Omega \setminus \operatorname{supp}(p_1)$ we choose an arbitrary probability distribution $p^{\omega'} \in P(\Omega)$ and set for $\omega \in \Omega$

$$k(\omega|\omega') := p^{\omega'}(\omega).$$

For $\omega' \in \operatorname{supp}(p_1)$ and $\omega \in \Omega$ we set

$$k(\omega|\omega') := \frac{p(\omega, \omega')}{p_1(\omega')}.$$

Then k is a Markov transition kernel on Ω with $J(p_1, k) = p$.

To determine the fibers of J let $p, p' \in P(\Omega)$ and $k, k' \in K(\Omega)$ with $J(p, k) = J(p', k')$. Then one has $p = J(p, k)_2 = J(p', k')_2 = p'$ and for $\omega' \in \operatorname{supp}(p)$ and arbitrary $\omega \in \Omega$ follows the equality

$$k(\omega|\omega') = \frac{p(\omega')k(\omega|\omega')}{p(\omega')} = \frac{J(p,k)(\omega,\omega')}{p(\omega')} = k'(\omega|\omega').$$

These necessary conditions on p, p' and k, k' are sufficient for the equality $J(p, k) = J(p', k')$ by the discussion in the first paragraph.

We discuss the stationary case. Let $(p, k) \in P(\Omega) \times K(\Omega)$. Since $J(p, k)_2 = p$ one has the equivalence

$$\begin{aligned} &(p,k) \text{ is stationary} \quad \iff \quad J(p,k)_1 = p \\ \iff \quad &J(p,k)_1 = J(p,k)_2 \quad \iff \quad J(p,k) \in C(\Omega). \end{aligned}$$

Since $J(p, k)$ belongs to $P(\Omega \times \Omega)$ and since the Kirchhoff polytope is the intersection $\operatorname{Kirch}(\Omega) = P(\Omega \times \Omega) \cap C(\Omega)$ with the cycle space, the previous statements are equivalent to $J(p, k) \in \operatorname{Kirch}(\Omega)$. qed

Proof of Lemma 9.11. The Kirchhoff polytope $\operatorname{Kirch}(\Omega)$ is a polytope because it is the intersection of the joint probability simplex $P(\Omega \times \Omega)$ with the (linear) cycle space $C(\Omega)$ [Zi]. The elementary probability cycles belong trivially to $\operatorname{Kirch}(\Omega)$. The face relation is proved by shifting the face lattice to the joint probability simplex. In detail, let $x, y \in \operatorname{Kirch}(\Omega)$. Since $\operatorname{Kirch}(\Omega) = P(\Omega \times \Omega) \cap C(\Omega)$, the face of y is obtained by intersection:

$$F(\operatorname{Kirch}(\Omega), y) = F(P(\Omega \times \Omega), y) \cap C(\Omega).$$

This is proved in (3.25). Hence $x \in F(\operatorname{Kirch}(\Omega), y)$ is equivalent to $x \in F(P(\Omega \times \Omega), y)$. The latter condition is equivalent by (4.45) to the support projector relation $s(x) \leq s(y)$ which is the same as $\operatorname{supp}(x) \subset \operatorname{supp}(y)$. qed

Proof of Lemma 9.12. Let $x \in C(\Omega)$ be a cycle. We assume that c is an elementary cycle on Ω given by a non-empty set $U \subset \Omega$ and a cyclic permutation $\pi : U \to U$. If $\mathrm{supp}(x) \subset \mathrm{supp}(c)$ then every branch of the complex $\mathrm{supp}(x)$ is of the form $b = (\pi(\omega), \omega)$ for some $\omega \in U$. Using marginals we get

$$x(\pi(\omega), \omega) = x_2(\omega) = x_1(\omega) = x(\omega, \pi^{-1}(\omega)).$$

Since π acts transitively on $\mathrm{supp}(c)$, the vector x must be constant on $\mathrm{supp}(c)$. For branches b not in c we have $x(b) = 0$.

Let $x \in C(\Omega)$ be a non-zero cycle with non-negative coefficients. We construct an elementary cycle dominated by x. Let (ω_2, ω_1) be a branch in the non-empty complex $\mathrm{supp}(x)$. By assumption we have $x(\omega_2, \omega_1) > 0$. Then

$$x_2(\omega_2) = x_1(\omega_2) \geq x(\omega_2, \omega_1) > 0$$

shows that there must be a node $\omega_3 \in \Omega$ with $x(\omega_3, \omega_2) > 0$, that is $(\omega_3, \omega_2) \in \mathrm{supp}(x)$. Inductively we find a ray of branches $\dots, (\omega_3, \omega_2), (\omega_2, \omega_1)$ in the complex $\mathrm{supp}(x)$. Since Ω is finite there exists a smallest index $k \in \mathbb{N}$ such that $\omega_k = \omega_1$. Then with $U := \{\omega_1, \dots, \omega_{k-1}\}$ and the cyclic permutation

$$\pi : U \to U, \qquad \omega_1 \mapsto \omega_2, \quad \dots \quad \omega_{k-1} \mapsto \omega_k$$

we obtain the elementary cycle $c(U, \pi)$ which is dominated by $\mathrm{supp}(x)$. `qed`

Proof of Lemma 9.13. At first we show that the extreme points of $\mathrm{Kirch}(\Omega)$ are the elementary probability cycles. Let ζ be an elementary probability cycle. If there is $x \in F(\mathrm{Kirch}(\Omega), \zeta)$ then by Lemma 9.11 we have $\mathrm{supp}(x) \subset \mathrm{supp}(\zeta)$. Then Lemma 9.12 shows that $x = \lambda\zeta$ for some $\lambda \in \mathbb{C}$. Since x and ζ are probability distributions we get $\lambda = 1$. Conversely we show that every extreme point of $x \in \mathrm{Kirch}(\Omega)$ is an elementary probability cycle. Since x is non-negative and non-zero, there exists an elementary cycle c such that $\mathrm{supp}(c) \subset \mathrm{supp}(x)$ (Lemma 9.12). The corresponding normalized elementary probability cycle ζ belongs to $\mathrm{Kirch}(\Omega)$ and the support inclusion $\mathrm{supp}(\zeta) \subset \mathrm{supp}(x)$ proves that $\zeta \in F(\mathrm{Kirch}(\Omega), x)$ (Lemma 9.11). If x is an extreme point then this implies $\zeta = x$.

We discuss how extreme points of a face can be described. As a general principle, see Remark 3.11 (c) on page 55, the extreme points of a face of a convex set $C \in \mathbb{R}^n$ are these extreme points of C that belong to the face. Returning to the case of the Kirchhoff polytope we consider $x \in \mathrm{Kirch}(\Omega)$. We can decide with Lemma 9.11 that a cycle $\zeta \in Z(\Omega)$ belongs to the face $F(\mathrm{Kirch}(\Omega), x)$ if and only if $\mathrm{supp}(\zeta) \subset \mathrm{supp}(x)$.

Under the assumption $x \in \mathrm{Kirch}(\Omega)$ we write an arbitrary vector $y \in C(\mathrm{supp}(x))$ as a complex linear combination of elementary probability cycles dominated by x. We use a decomposition[2]

$$y = (\Re(y) + \lambda x) - \lambda x + i(\Im(y) + \mu x) - i\mu x$$

for non-negative scalars $\mu, \lambda \in \mathbb{R}$ such that $\Re(y) + \lambda x \geq 0$ and $\Im(y) + \mu x \geq 0$. These choices can be made because $\mathrm{supp}(y) \subset \mathrm{supp}(x)$ and we can assume that $y \geq 0$. A non-zero and non-negative cycle is decomposed into elementary cycles using successive support reduction (Lemma 9.12). The elementary cycles in the decomposition can be replaced by the corresponding normalized elementary probability cycles. qed

Proof of Proposition 9.14. Let $x \in \mathrm{Kirch}(\Omega)$ and $y \in P(\Omega \times \Omega)$. Then

$$\begin{aligned} y \in F(\mathrm{Kirch}(\Omega), x) \quad &\iff \quad y \in C(\Omega) \text{ and } \mathrm{supp}(y) \subset \mathrm{supp}(x) \\ \iff \quad & y \in C(\mathrm{supp}(x)). \end{aligned}$$

The first equivalence follows from the definition $\mathrm{Kirch}(\Omega) = P(\Omega \times \Omega) \cap C(\Omega)$ of the Kirchhoff polytope (9.7) and from the equivalence for points $z \in \mathrm{Kirch}(\Omega)$ that $z \in F(\mathrm{Kirch}(\Omega), x)$ if and only if $\mathrm{supp}(z) \subset \mathrm{supp}(x)$ (Lemma 9.11). The second equivalence above is the definition of the cycle space of a complex in Definition 9.8. This proves that $F(\mathrm{Kirch}(\Omega), x) = P(\Omega \times \Omega) \cap C(\mathrm{supp}(x))$.

By Minkowski theorem a compact convex subset of $\mathbb{R}^n$ is the convex hull of its extreme points. The extreme points of $F(\mathrm{Kirch}(\Omega), x)$ are the elementary probability cycles dominated by $\mathrm{supp}(x)$ by Lemma 9.13, hence

$$F(\mathrm{Kirch}(\Omega), x) = \mathrm{conv}\left(\{\zeta \in Z(\Omega) : \mathrm{supp}(\zeta) \subset \mathrm{supp}(x)\}\right).$$

The elementary cycles have all trace one. So their affine hull does not contain zero and therefore

$$\begin{aligned} &\dim(F(\mathrm{Kirch}(\Omega), x)) = \dim\left(\mathrm{aff}\left(\{\zeta \in Z(\Omega) : \mathrm{supp}(\zeta) \subset \mathrm{supp}(x)\}\right)\right) \\ &= \dim_{\mathbb{R}}\left(\mathrm{Lin}_{\mathbb{R}}\left(\{\zeta \in Z(\Omega) : \mathrm{supp}(\zeta) \subset \mathrm{supp}(x)\}\right)\right) - 1 \\ &= \dim_{\mathbb{C}}\left(\mathrm{Lin}_{\mathbb{C}}\left(\{\zeta \in Z(\Omega) : \mathrm{supp}(\zeta) \subset \mathrm{supp}(x)\}\right)\right) - 1 \end{aligned}$$

with the last equality because elementary cycles are real. The latter vector space is the cycle space $C(\mathrm{supp}(x))$ by Lemma 9.13. It has dimension equal the cyclomatic number $\mu(\mathrm{supp}(x))$ (9.12). qed

[2] $\Re(z)$ and $\Im(z)$ denote real and imaginary part of a complex number z.

Proof of Proposition 9.17. For a detailed proof we introduce a finer factorization of the algebra $A := A_{[N]}$. We use $B_j := \mathbb{C}^{\Omega_j}$ for $j \in [N]$. Then

$$A = B_{\{1,2\}\times[N]} = \mathbb{C}^{\Omega} \otimes \mathbb{C}^{\Omega}.$$

The following partial traces will be considered. If $x \in B_{\{1,2\}\times[N]}$ then we use

$\mathrm{tr}_{\{1,2\}\times\{j\}}(x)$	$\in$	$\mathbb{C}^{\Omega_j} \otimes \mathbb{C}^{\Omega_j}$	the "j-th unit marginal", $j \in [N]$,
$\mathrm{tr}_{\{1\}\times[N]}(x)$	$\in$	$\mathbb{C}^{\Omega}$	the "1st global marginal",
$\mathrm{tr}_{\{2\}\times[N]}(x)$	$\in$	$\mathbb{C}^{\Omega}$	the "2nd global marginal",
$\mathrm{tr}_{(1,j)}(x)$	$\in$	$\mathbb{C}^{\Omega_j}$	the "1st local marginal for the unit j", $j \in [N]$,
$\mathrm{tr}_{(2,j)}(x)$	$\in$	$\mathbb{C}^{\Omega_j}$	the "2nd local marginal for the unit j", $j \in [N]$.

We consider the local cycle spaces $C(\Omega_j) \subset A_j$ for $j \in [N]$ and the global cycle space $C(\Omega) \subset A$ (9.6). Local and global cycle spaces are connected.

(a) If $y_j \in C(\Omega_j) \subset A_j$ for $j \in [N]$ then the Kronecker product $y_{[N]}$ belongs to $C(\Omega)$. We can write $y_j = \sum_{\alpha\in I_j} x^{\alpha}_{(1,j)} \otimes x^{\alpha}_{(2,j)}$ with finite index sets I_j for $j \in [N]$ where $x^{\alpha_j}_{(i,j)} \in \mathbb{C}^{\Omega_j}$ for $\alpha_j \in I_j$, $i = 1, 2$ and $j \in [N]$. To apply the distributive law we introduce a multi-index $\alpha := (\alpha_1, \ldots, \alpha_N)$ in $I := \times_{j\in[N]} I_j$ and define $z^{\alpha}_{(i,j)} := x^{\alpha_j}_{(i,j)}$ for $i = 1, 2$ and $j \in [N]$. Then

$$\begin{aligned} y_{[N]} &= \textstyle\bigotimes_{j\in[N]} \sum_{\alpha\in I_j} x^{\alpha}_{(1,j)} \otimes x^{\alpha}_{(2,j)} \\ &= \textstyle\sum_{\alpha\in I} \bigotimes_{j\in[N]} (x^{\alpha_j}_{(1,j)} \otimes x^{\alpha_j}_{(2,j)}) \\ &= \textstyle\sum_{\alpha\in I} \bigotimes_{j\in[N]} (z^{\alpha}_{(1,j)} \otimes z^{\alpha}_{(2,j)}) \\ &= \textstyle\sum_{\alpha\in I} z^{\alpha}_{\{1,2\}\times[N]}. \end{aligned}$$

For $j \in [N]$ the assumption that $y_j \in C(\Omega_j) \subset A_j$ reads $\mathrm{tr}_{\{1\}}(y_j) = \mathrm{tr}_{\{2\}}(y_j)$ in the algebra $\mathbb{C}^{\Omega_j} \otimes \mathbb{C}^{\Omega_j}$ and this is

$$\sum_{\alpha\in I_j} x^{\alpha}_{(1,j)} \,\mathrm{tr}(x^{\alpha}_{(2,j)}) = \sum_{\alpha\in I_j} x^{\alpha}_{(2,j)} \,\mathrm{tr}(x^{\alpha}_{(1,j)}).$$

These equations imply

$$\begin{aligned} \textstyle\sum_{\alpha\in I} z^{\alpha}_{\{1\}\times[N]} \,\mathrm{tr}(z^{\alpha}_{\{2\}\times[N]}) &= \textstyle\sum_{\alpha\in I} \bigotimes_{j\in[N]} z^{\alpha}_{(1,j)} \,\mathrm{tr}(z^{\alpha}_{(2,j)}) \\ &= \textstyle\bigotimes_{j\in[N]} \sum_{\alpha\in I_j} x^{\alpha}_{(1,j)} \,\mathrm{tr}(x^{\alpha}_{(2,j)}) \\ &= \textstyle\sum_{\alpha\in I} z^{\alpha}_{\{2\}\times[N]} \,\mathrm{tr}(z^{\alpha}_{\{1\}\times[N]}). \end{aligned}$$

Then finally one obtains

$$\mathrm{tr}_{\{1\}\times[N]}(y_{[N]}) = \mathrm{tr}_{\{1\}\times[N]}(\sum_{\alpha\in I} z^{\alpha}_{\{1,2\}\times[N]}) = \sum_{\alpha\in I} z^{\alpha}_{\{1\}\times[N]} \,\mathrm{tr}(z^{\alpha}_{\{2\}\times[N]}) = \mathrm{tr}_{\{2\}\times[N]}(y_{[N]}).$$

(b) If $x \in C(\Omega)$ then for $j \in [N]$ the j-th unit marginal $x_{\{1,2\}\times\{j\}}$ belongs to $C(\Omega_j)$. Since $x \in A = B_{\{1,2\}\times[N]}$ we can write $x = \sum_{\alpha\in I} x^\alpha_{\{1,2\}\times[N]}$ for a finite index set I and $x^\alpha_{(i,j)} \in \mathbb{C}^{\Omega^j}$ for each $j \in [N]$. Since $x \in C(\Omega)$ one has $\mathrm{tr}_{\{1\}\times[N]}(x) = \mathrm{tr}_{\{2\}\times[N]}(x)$, that is

$$\sum_{\alpha\in I} x^\alpha_{\{1\}\times[N]}\,\mathrm{tr}(x^\alpha_{\{2\}\times[N]}) = \mathrm{tr}_{\{1\}\times[N]}(\sum_{\alpha\in I} x^\alpha_{\{1,2\}\times[N]}) = \mathrm{tr}_{\{1\}\times} = \sum_{\alpha\in I} x^\alpha_{\{2\}\times[N]}\,\mathrm{tr}(x^\alpha_{\{1\}\times[N]}).$$

For $j \in [N]$ that the j-th unit marginal $x_{\{1,2\}\times\{j\}}$ belongs to $C(\Omega_j)$ means that $\mathrm{tr}_{(1,j)}(x) = \mathrm{tr}_{(2,j)}(x)$. Taking the j-th partial trace in the algebra $A_{[N]}$ in step (*) one has

$$\begin{array}{rcl} \mathrm{tr}_{(1,j)}(x) & = & \mathrm{tr}_{(1,j)}(\sum_{\alpha\in I} x^\alpha_{\{1,2\}\times[N]}) \\ & = & \sum_{\alpha\in I} x^\alpha_{(1,j)}\,\mathrm{tr}(x^\alpha_{\{1\}\times([N]\setminus\{j\})})\,\mathrm{tr}(x^\alpha_{\{2\}\times[N]}) \\ & \overset{(*)}{=} & \sum_{\alpha\in I} \mathrm{tr}_{\{j\}}(x^\alpha_{\{1\}\times[N]})\,\mathrm{tr}(x^\alpha_{\{2\}\times[N]}) \\ & = & \sum_{\alpha\in I} \mathrm{tr}_{\{j\}}(x^\alpha_{\{1\}\times[N]}\,\mathrm{tr}(x^\alpha_{\{2\}\times[N]})) \\ & = & \mathrm{tr}_{(2,j)}(x). \end{array}$$

For a factorizable family the closure and *rI*-closure are equal by Remark 8.23 (a), *rI*- and combinatorial closure coincide by Corollary 7.17. We can prove that the combinatorial normal projection to $\mathcal{F}([N])^{\mathrm{cmb}}$ maps $\mathrm{Kirch}(\Omega)$ into $\mathrm{Kirch}(\Omega)$. If Q is a probability distribution of the Kirchhoff polytope $\mathrm{Kirch}(\Omega)$ then $N^{\mathrm{cmb}}(Q) = \bigotimes_{i\in[N]} \mathrm{tr}_{\{1,2\}\times\{j\}}(Q)$, this is a result for factorizable families proved in Lemma 8.13. The paragraph (b) above gives $\mathrm{tr}_{\{1,2\}\times\{j\}}(Q) \in C(\Omega_j)$ for $j \in [N]$ and this implies by paragraph (a) that the product $N^{\mathrm{cmb}}(Q)$ belongs to $\mathrm{Kirch}(\Omega)$.

We can prove the equality $\overline{\mathcal{F}([N])} \cap \mathrm{Kirch}(\Omega) = \{Q_{[N]} \;:\; Q_i \in \mathrm{Kirch}(\Omega_i),\; i \in [N]\}$. Lemma 8.14 proves that a probability distribution $Q \in \overline{\mathcal{F}([N])}$ is of the form $Q_{[N]}$ for $Q_i \in P(\Omega_i \times \Omega_i)$, $i \in [N]$. Since $Q = Q_{[N]}$ belongs to the global cycle space $C(\Omega)$ we get from paragraph (b) above that the j-the unit-marginal $\mathrm{tr}_{\{1,2\}\times\{j\}}(Q) = Q_j$ belongs to $\mathrm{Kirch}(\Omega_j)$ for $j \in [N]$. This shows the inclusion "$\subset$". Conversely, if $Q_i \in \mathrm{Kirch}(\Omega_j)$ for $j \in [N]$ then obviously $Q_{[N]}$ belongs to $P(\Omega \times \Omega)$ and by the paragraph (a) above $Q_{[N]} \in C(\Omega)$ holds. This completes equality.

The equation $\mathcal{F}([N]) \cap \mathrm{Kirch}(\Omega) = \{Q_{[N]} \;:\; Q_i \in \mathrm{ri}(\mathrm{Kirch}(\Omega_i)),\; i \in [N]\}$ follows from the previous formula. On one hand the exponential family $\mathcal{F}([N])$ consists of the invertible density matrices in $\overline{\mathcal{F}([N])}$ on the other hand $\mathrm{ri}(\mathrm{Kirch}(\Omega))$ consists of the invertible density matrices in $\mathrm{Kirch}(\Omega)$ because the vector space $C(\Omega)$ intersects the relative interior of $P(\Omega \times \Omega)$ in the identical distribution $\frac{\mathbb{1}}{\mathrm{tr}(\mathbb{1})}$ (4.27) and (3.17). `qed`

Bibliography

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

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

[Am01] S. Amari, Information geometry on hierarchy of probability distributions, IEEE Trans. Inf. Theory 47, No. 5, 1701–1711 (2001).

[Asl] H. Aslaksen and A. B. Sletsjøe, Generators of matrix algebras in dimension 2 and 3, Linear Algebra Appl. 430, No. 1, 1–6 (2009).

[Asp] E. Asplund, A k-extreme point is the limit of k-exposed points, Isr. J. Math. 1, 161–162 (1963).

[Av] J. Avron, R. Seiler and B. Simon, The index of a pair of projections, J. Funct. Anal. 120, No. 1, 220–237 (1994).

[Ay01] N. Ay, Information Geometry on Complexity and Stochastic Interaction, MPI MIS, Preprint Nr. 95/2001 (2001).

[Ay02] N. Ay, An Information-Geometric Approach to a Theory of Pragmatic Structuring, The Annals of Probability, vol. 30, No. 1, 416–436, (2002).

[Ay03] N. Ay and T. Wennekers, Dynamical Properties of Strongly Interacting Markov Chains, Neural Networks 16, 1483–1497 (2003).

[AK] N. Ay and A. Knauf, Maximizing Multi-Information, Kybernetika 42(5), 517–538 (2006).

[Bam] P. Bamberg and S. Sternberg, *A Course in Mathematics for Students of Physics*, Cambridge University Press, (1990).

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

[BaM] H. Bauer, *Maß- und Integrationstheorie*, de Gruyter (1990).

[BaW] H. Bauer, *Wahrscheinlichkeitstheorie*, de Gruyter (1990).

[Ben] F. Benatti, *Deterministic Chaos in Infinite Quantum Systems*, Springer-Verlag (1993).

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

[Bel] E. G. Beltrametti and G. Cassinelli, *The logic of quantum mechanics*, Addison-Wesley Publishing Company (1981).

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

[Bi] G. Birkhoff, *Lattice Theory*, Third Edition, AMS Colloquium Publications (1973).

[Bj05] I. Bjelaković, J.-D. Deuschel, T. Krüger, R. Seiler, R. Siegmund-Schultze and A. Szkoła, A quantum version of Sanov's theorem, Commun. Math. Phys. 260, No. 3, 659–671 (2005).

[BS] I. Bjelaković and A. Szkoła, The data compression theorem for ergodic quantum information sources, Quantum Inf. Process. 4, No. 1, 49–63 (2005).

[Bo] S. Borac, On the algebra generated by two projections, J. Math. Phys. 36, No. 2, 863–874 (1995).

[Bra] O. Bratteli und D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics 1*, Springer-Verlag (1979).

[Brø] A. Brøndsted, *An Introduction to Convex Polytopes*, GTM 90, Springer-Verlag (1983).

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

[Co] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications (1991).

[Cs03] I. Csiszár and F. Matúš, Information Projections Revisited, IEEE Trans. Inf. Theory, 49, No. 6, 1474–1490 (2003).

[Cs05] I. Csiszár and F. Matúš, Closures of Exponential Families, Ann. Probab. 33, No. 2, 582–600 (2005).

[Cs08] I. Csiszár and F. Matúš, Generalized maximum likelihood estimates for exponential families, Probab. Theory Relat. Fields 141, No. 1-2, 213–246 (2008).

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

[Du] N. Dunford and J. T. Schwartz, *Linear operators. Part I: General theory*, John Wiley & Sons Ltd. (1988).

[Ef] B. Efron, Maximum likelihood and decision theory, Ann. Stat. 10, 340–356 (1982).

[Er] I. Erb and N. Ay, Multi-information in the thermodynamic limit, J. Stat. Phys. 115, No. 3-4, 949–976 (2004).

[Fi] G. Fischer, *Lineare Algebra*, Vieweg, 15. Auflage (2005).

[Ga] S. Gallot, D. Hulin and J. Lafontaine, *Riemannian geometry*, 3rd ed., Universitext, Springer (2004).

[Ge] D. Geiger, C. Meek and B. Sturmfels, On the toric algebra of graphical models, Ann. Stat. 34, No. 3, 1463–1492 (2006).

[Gra] M. Grasselli and R. F. Streater, On the Uniqueness of the Chentsov Metric in Quantum Information Geometry, Inf. Analysis, 4, 173–182 (2001).

[Grü] B. Grünbaum, *Convex polytopes*, John Wiley & Sons (1967).

[Hal] P. R. Halmos, *Finite-Dimensional Vector Spaces*, Springer-Verlag (1974).

[Har] J. Harris, *Algebraic geometry*, Springer-Verlag (1992).

[Hi] M. W. Hirsch, *Differential Topology*, Springer-Verlag (1976).

[Ho] A. S. Holevo, *Statistical Structure of Quantum Theory*, Springer-Verlag (2001).

[Ib] B. Ibinson, N. Linden, and A. Winter, All inequalities for the relative entropy, Commun. Math. Phys. 269, No. 1, 223–238 (2007).

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

[Ja] J. Jahn, *Introduction to the Theory of Nonlinear Optimization*, Springer-Verlag (1994).

[KaG] G. Kalmbach, *Orthomodular Lattices*, London Mathematical Society Monographs, 18 (1983).

[KaS] S. L. Kalpazidou, *Cycle representations of Markov processes*, 2nd ed., Applications of Mathematics 28, Springer (2006).

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

[Kl] W. Klingenberg, *Lineare Algebra und Geometrie*, Springer-Verlag (1984).

[Kn01] A. Knauf, Vorlesung Lineare Algebra und Analytische Geometrie II (SS 2001).

[Kn07] A. Knauf and R. Seiler, Vorlesung 'Statistische Mechanik' Mathematische Physik II (WS 06/07).

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

[La] S. Lang, *Algebra*, 3. ed., Addison Wesley (1993).

[Lei] K. Leichtweiß, *Konvexe Mengen*, Springer (1980).

[Lev] J. S. Levinger and R. M. Lichtenstein, Intuitive Understanding of Pauli's σ Matrices, Am. J. Phys. 47, 744(1979).

[Lie] E. H. Lieb, Convex Trace Functions and the Wigner-Yanase-Dyson Conjecture, Adv. in Math. 11, 267–288 (1973).

[Lin] R. Linsker, Self-organization in a perceptual network, IEEE Computer 21, 105–117 (1981).

[Ma] H. Matsuda, K. Kudo, N. Kiyoshi, R. Nakamura, O. Yamakawa and T. Murata, Mutual information of Ising systems, Int. J. Theor. Phys. 35, No.4, 839–845 (1996).

[MA] F. Matúš and N. Ay, On maximization of the information divergence from an exponential family, Proceedings of WUPES'03 (2004).

[Mu] G. J. Murphy, *C^*-Algebras and Operator Theory*, Academic Press Inc. (1990).

[MR] M. K. Murray and J. W. Rice, *Differential geometry and statistics*, Monographs on Statistics and Applied Probability 48, Chapman & Hall (1993).

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

[Pa] S. Papadopoulou, On the Geometry of Stable Compact Convex Sets, Math. Ann. 229, 193–200 (1977).

[Pe94] D. Petz, Geometry of Canonical Correlation on the State Space of a Quantum System, J. Math. Phys. 35, 780–795 (1994).

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

[Ree1] M. Reed and B. Simon, *Functional Analysis*, Methods of Modern Mathematical Physics I, Academic Press Inc. (2005).

[Ree4] M. Reed and B. Simon, *Analysis of Operators*, Methods of Modern Mathematical Physics IV, Academic Press Inc. (2005).

[Rem] R. Remmert and G. Schumacher, *Funktionentheorie 1*, Springer-Verlag (2002).

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

[Ru] D. Ruelle, *Statistical Mechanics*, W. A. Benjamin (1969).

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

[Sh] C. E. Shannon, A Mathematical Theory of Communication, Bell System Technical Journal, vol. 27, 379–423 & 623–656 (1948).

[St] S. Straszewicz, Über exponierte Punkte abgeschlossener Punktmengen, Fundam. Math. 24, 139–143 (1935).

[Th3] W. Thirring, *Lehrbuch der Mathematischen Physik 3: Quantenmechanik von Atomen und Molekülen*, Springer-Verlag (1979).

[Th4] W. Thirring, *Lehrbuch der Mathematischen Physik 4: Quantenmechanik großer Systeme*, Springer-Verlag (1980).

[To] G. Tononi, O. Sporns and G. M. Edelmann, A measure for brain complexity: Relating functional segregation and integration in the nervous system, Proc. Nat. Acad. Sci. U.S.A. 91, 5033–5037 (1994).

[Va] V. S. Varadarajan, Geometry of quantum theory, vol. I, D. van Nostrand Company Inc. (1968).

[Vo] L. Volkmann, Estimations for the number of cycles in a graph, Period. Math. Hung. 33, No. 2, 153–161 (1996).

[Weh] A. Wehrl, General Properties of Entropy, Review of Modern Physics, vol. 50, No. 2 (1978).

[Wer] D. Werner, *Funktionalanalysis*, Springer (2007).

[Wey] J. M. Weyman, *Cohomology of vector bundles and syzygies*, Cambridge University Press (2003).

[Wic] E. H. Wichmann, Density Matrices Arising From Incomplete Measurements, J. Math. Phys., vol. 4, No. 7 (1963).

[Wil] J. M. Wills, Symmetrisierung an Unterräumen, Arch. Math. 20, 169–172 (1969).

[Zi] G. Ziegler, *Lectures on polytopes*, Springer-Verlag (1999).

Index of symbols

Index of terms