International Journal of Approximate Reasoning 52 (2011) 613–626

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Support sets in exponential families and oriented matroid theory

Johannes Rauh [a,*], Thomas Kahle [a,c], Nihat Ay [a,b]

[a] Max Planck Institute for Mathematics in the Sciences, Inselstr. 22, D-04103 Leipzig, Germany
[b] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
[c] Institut Mittag-Leffler, Auravägen 17, 182 60 Djursholm, Sweden

ARTICLE INFO

Article history:
Available online 4 February 2011

Keywords:
Exponential families
Oriented matroids
Algebraic statistics
Polytopes
Moment map
Support sets

ABSTRACT

The closure of a discrete exponential family is described by a finite set of equations corresponding to the circuits of an underlying oriented matroid. These equations are similar to the equations used in algebraic statistics, although they need not be polynomial in the general case. This description allows for a combinatorial study of the possible support sets in the closure of an exponential family. If two exponential families induce the same oriented matroid, then their closures have the same support sets. Furthermore, the positive cocircuits give a parameterization of the closure of the exponential family.

1. Introduction

In this paper we study exponential families $\mathcal{E}$ on a finite set $\mathcal{X}$, which are well known statistical models with many nice properties. We are interested in the closure $\overline{\mathcal{E}}$ (with respect to the usual topology) and the set

$$\{\operatorname{supp}(p) \subseteq \mathcal{X} : p \in \overline{\mathcal{E}}\} \tag{1}$$

of all possible support sets occurring in the closure $\overline{\mathcal{E}}$. In this paper, we find a characterization of the support sets and draw connections to various other approaches, thereby generalizing and exposing scattered ideas that deserve wider attention.

Finding the possible support sets and the boundary $\overline{\mathcal{E}} \setminus \mathcal{E}$ of an exponential family $\mathcal{E}$ is an important problem in statistics. An early discussion of the case where $\mathcal{X}$ is finite is due to Barndorff-Nielsen [4]. In the general case, when $\mathcal{X}$ is infinite, there are different possible notions of closure. This was studied by Csiszár and Matúš in a series of papers, see [8]; earlier results are due to Chentsov [6]. The problem is related to characterizing the convex core and convex support of $\mathcal{E}$. In the finite case, the notions of convex core and convex support agree, and the convex support is a polytope, called the marginal polytope in the case of hierarchical models, a particular kind of exponential families. Knowing the convex core is important for studying properties of the generalized maximum likelihood estimate, see [4,9].

Computing the support sets of an exponential family is equivalent to determining the face lattice of the convex support, which can lead to hard combinatorial problems. For example, the so-called CUT-polytopes appear naturally when studying hierarchical models (see [20] for the relation). In [10] it is shown that deciding whether a given point lies in the CUT-polytope is NP-complete. Nevertheless, a local exploration of the face lattice is possible using the results presented here. For instance, one of the authors discusses support sets of small cardinality in hierarchical models [17]. In the present paper we find a concise characterization of the support sets in general exponential families with the help of oriented matroids. Furthermore, we show how to describe the closure of an exponential family parametrically and implicitly. We hope that this will allow for further theoretical results in this direction. For an illustration how these results can be applied see [27].

* Corresponding author.
E-mail addresses: jarauh@gmx.net, rauh@mis.mpg.de (J. Rauh), nay@mis.mpg.de (N. Ay).
URL: http://www.thomas-kahle.de (T. Kahle).

0888-613X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2011.01.013

Although slightly hidden, the connection to oriented matroid theory is very natural. The starting point, and another focus of the presentation, is the implicit description of exponential families in Theorem 4, which is inspired by so called Markov bases [14]. We study the—not necessarily polynomial—equations that define the closure of the exponential family and relate them to the oriented matroid of the sufficient statistics of the model. In the case of a rational valued sufficient statistics, our observations reduce to the fact that the non-negative real part of a toric variety is described by a circuit ideal. We emphasize how the proof of this fact uses arguments from oriented matroid theory.

The oriented matroid also plays a role when one tries to parametrize the closure of an exponential family. The usual exponential parametrization naturally extends to a part of the boundary if the parameters are replaced by their exponentials. With the help of the positive cocircuits of the oriented matroid one can construct a parametrization of the whole closure.

This paper is organized as follows. In Section 2 we prove an implicit representation of exponential families which is analogue to and inspired by algebraic statistics [14]. In contrast to the toric case we do not require the sufficient statistics to take integer values and thereby leave the realm of commutative algebra. What remains is the theory of oriented matroids. We discuss how answers to the support set problem look like in the language of oriented matroids and discuss examples coming from cyclic polytopes. These polytopes are well known in combinatorial convexity for their extremal properties, as stated, for instance, in the Upper Bound Theorem. In 2.3 we show how to obtain a surjective parametrization of the closure of an exponential family. In Section 3 we discuss the basics of the theory of oriented matroids and reformulate statements from Section 2 in this language, making the connection as clear as possible.

2. Exponential families

We assume a finite set $\mathcal{X} := \{1, \ldots, m\}$ and denote $\mathcal{P}(\mathcal{X})$ the open simplex of probability measures with full support on $\mathcal{X}$. The closure of any set $M \subseteq \mathbb{R}^{\mathcal{X}}$, in the standard topology of $\mathbb{R}^n$, is denoted by $\overline{M}$. Any vector $n \in \mathbb{R}^{\mathcal{X}}$ can be decomposed into its positive and negative part, that is $n = n^+ - n^-$ via $n^+(x) := \max(n(x), 0)$ and $n^-(x) := \max(-n(x), 0)$. For any two vectors $n, p \in \mathbb{R}^{\mathcal{X}}$ we define

$$p^n := \prod_{x \in \mathcal{X}} p(x)^{n(x)}, \tag{2}$$

whenever this product is well defined (e.g. when n and p are both non-negative). Here $0^0 = 1$ by convention.

Let q be a positive measure on $\mathcal{X}$ with full support, and let $A \in \mathbb{R}^{d \times m}$ be a matrix of width m. We denote a_x, $x \in \mathcal{X}$, the columns of A. Then we have

Definition 1. The *exponential family* associated with the reference measure q and the matrix A is the set of probability measures

$$\mathcal{E}_{q,A} := \left\{ p_\theta \in \mathcal{P}(\mathcal{X}) : p_\theta(x) = \frac{q(x)}{Z_\theta} \exp\left(\theta^T a_x\right), \theta \in \mathbb{R}^d \right\}, \tag{3}$$

where $Z_\theta := \sum_{x \in \mathcal{X}} q(x) \exp\left(\theta^T a_x\right)$ ensures normalization. If $q(x) = 1$ for all $x \in \mathcal{X}$, i.e. if q is the uniform measure on $\mathcal{X}$, then the corresponding exponential family is abbreviated with $\mathcal{E}_A$.

In the following we always assume that the matrix A has the vector $(1, \ldots, 1)$ in its row span. This means that there exists a dual vector $l_1 \in \left(\mathbb{R}^d\right)^*$ which satisfies $l_1(a_x) = 1$ for all $x \in \mathcal{X}$. There is no loss of generality in this assumption as we can always add an additional row $(1, \ldots, 1)$ to A without changing the exponential family.

Remark 2. Under the assumption that the row span of A contains the vector $(1, \ldots, 1)$ the exponential family depends on A only through its row span. Different matrices with the same row span lead to different parametrizations of the same exponential family. In the following it will be convenient to fix one parametrization, hence we work with matrices A instead of vector spaces.

The geometrical structure of the boundary of $\overline{\mathcal{E}_{q,A}}$ is encoded in the polytope of possible values that the map $A \colon \overline{\mathcal{P}}(\mathcal{X}) \to \mathbb{R}^d$, $x \mapsto Ax$ takes:

Definition 3. The *convex support* of $\mathcal{E}_{q,A}$ is the polytope

$$\mathrm{cs}(\overline{\mathcal{E}_{q,A}}) := \mathrm{conv}\left\{a_x : x \in \mathcal{X}\right\}. \tag{4}$$

If $\mathcal{E}_{q,A}$ is a hierarchical model, then the convex support is also called a *marginal polytope*.

We will see later that the faces of $\mathrm{cs}(\overline{\mathcal{E}_{q,A}})$ are in a one-to-one correspondence with the different support sets occurring in $\overline{\mathcal{E}_{q,A}}$. Even more is true: The linear mapping of A, which is called the *moment map* restricts to a homeomorphism

$\overline{\mathcal{E}_{q,A}} \cong \mathrm{cs}(\overline{\mathcal{E}_{q,A}})$ which maps every probability measure $p \in \overline{\mathcal{E}_{q,A}}$ into the face corresponding to its support, see for example [4].

The parametrization in (3) does not extend to the boundary. One might try to overcome this problem by allowing parameters to become infinite. More elegantly one may replace the parameters θ_i by their exponentials $\xi_i := e^{\theta_i}$, introducing the *monomial parametrization*, and allowing $\xi_i = 0$. However, the image of this parametrization depends on the matrix A, and not only the row space of A. We will discuss this in Section 2.3.

2.1. Implicit representations of exponential families

The problems with parametrizations are the main motivation to move on to an implicit description of the exponential family $\mathcal{E}_{q,A}$.

Theorem 4. *A distribution p is an element of the closure of $\mathcal{E}_{q,A}$ if and only if all the equations*

$$p^{n^+}q^{n^-} = p^{n^-}q^{n^+}, \quad \text{for all } n \in \ker A, \tag{5}$$

hold for p.

Remark 5. This theorem is a direct generalization of Theorem 3.2 in [14]. There only the polynomial equations among (5) are studied under the additional assumption that A has only integer entries. Moreover, only the uniform reference measure is considered. It turns out that their proof generalizes without any major problem, and the proof of our theorem that we present below needs one step less, since we do not need to show the reduction to the polynomial equations. The different flavor of the results will be made more precise in Remark 14 later.

Our proof closely follows [14], but we want to explicitly point out how matroid-type arguments are used, the first example being Lemma 7.

We first state a couple of auxiliary results which are of independent interest. The matrix A and derived objects are fixed for the rest of the considerations. A set C is a *face* of a polytope P if either $C = P$ or C is the intersection of the polytope with an affine hyperplane H, such that all $x \in P$ with $x \notin H$ lie on one side of the hyperplane. Faces of maximal dimension are called *facets*. It is a fundamental result that every polytope can equivalently be described as the convex hull of a finite set or as a finite intersection of closed linear half-spaces (corresponding to its facets), see [29].

In particular we are interested in the face structure of $\mathrm{cs}(\overline{\mathcal{E}_{q,A}})$. Since we assumed that all columns of A lie in the affine hyperplane $l_1 = 1$, we can replace every affine hyperplane H by an equivalent central hyperplane (which passes through the origin). For the convex support $\mathrm{cs}(\overline{\mathcal{E}_{q,A}})$ we want to know which points from $\{a_x : x \in \mathcal{X}\}$ lie on each face. This motivates the following

Definition 6. A set $F \subseteq \mathcal{X}$ is called *facial* if there exists a vector $c \in \mathbb{R}^d$ such that

$$c^T a_y = 0 \quad \forall y \in F, \qquad c^T a_z \geq 1 \quad \forall z \notin F. \tag{6}$$

Lemma 7. *Fix a subset $F \subseteq \mathcal{X}$. Then we have:*

- *F is facial if and only if for any $u \in \ker A$:*

$$\mathrm{supp}(u^+) \subseteq F \Leftrightarrow \mathrm{supp}(u^-) \subseteq F. \tag{7}$$

- *If p is a solution to (5), then* $\mathrm{supp}(p)$ *is facial.*

Proof. One direction of the first statement is direct: Let $u \in \ker A$ and suppose that $\mathrm{supp}(u^+) \subseteq F$. Then $\sum_{x\in F} u(x)a_x = -\sum_{x\notin F} u(x)a_x$, so $0 = \sum_{x\in F} u(x)c^T a_x = -\sum_{x\notin F} u(x)(c^T a_x)$. Since $c^T a_x > 1$ and $u(x) \leq 0$ for $x \notin F$ it follows that $u(x) = 0$ for $x \notin F$, proving one direction of the first statement.

The opposite direction is a bit more complicated. Here, we present a proof using elementary arguments from polytope theory (see, e.g., [29]). For an alternative proof using Farkas' Lemma see [14]. Assume that F is not facial. Let F' be the smallest facial set containing F. Let P_F and $P_{F'}$ be the convex hulls of $\{a_x : x \in F\}$ and $\{a_x : x \in F'\}$. Then P_F contains a point q from the relative interior of $P_{F'}$. Therefore q can be represented as $q = \sum_{x\in F} \alpha(x)a_x = \sum_{x\in F'} \beta(x)a_x$, where $\alpha(x) \geq 0$ for $x \in F$ and $\beta(x) > 0$ for $x \in F'$. Hence $u(x) := \alpha(x) - \beta(x)$ (where $\alpha(x) := 0$ for $x \notin F$ and $\beta(x) := 0$ for $x \notin F'$) defines a vector $u \in \ker A$ such that $\mathrm{supp}(u^+) \subseteq F$ and $\mathrm{supp}(u^-) \cap (\mathcal{X} \setminus F) = F' \setminus F \neq \emptyset$.

The second statement now follows immediately: If p satisfies (5) for some $u \in \ker A$, then the left hand side of (5) vanishes if and only if the right hand side vanishes, and by the first statement this implies that $\mathrm{supp}(p)$ is facial. $\square$

Now we are ready for the proof of Theorem 4.

Proof of Theorem 4. The first thing to note is that it is enough to prove the theorem when $q(x) = 1$ for all x. To see this note that $p \in \overline{\mathcal{E}_A}$ if and only if $\lambda qp \in \overline{\mathcal{E}_{q,A}}$, where $\lambda > 0$ is a normalizing constant, which does not appear in Eq. (5) since they are homogeneous.

Let Z_A be the set of solutions of (5). We first show that $\mathcal{E}_A$ satisfies the equations defining Z_A. Let $p \in \mathcal{E}_A$, using the parametrization we can write $p(x) = e^{\theta^T a_x}$, for some vector of parameters $\theta \in \mathbb{R}^d$ and a_x the x-column of A. Then, for each $u, v \in \mathbb{R}^m$ with $Au = Av$, we find

$$p^u = \prod_{x\in\mathcal{X}} p(x)^{u(x)} = \prod_{x\in\mathcal{X}} \left(e^{\theta^T a_x}\right)^{u(x)} = e^{\theta^T Au} = e^{\theta^T Av} = p^v. \tag{8}$$

Thus $\mathcal{E}_A \subseteq Z_A$, and also $\overline{\mathcal{E}_A} \subseteq \overline{Z}_A = Z_A$.

Next, let $p \in Z_A \setminus \mathcal{E}_A$ and put $F := \mathrm{supp}(p)$. We construct a sequence $p_{(\mu)}$ in $\mathcal{E}_A$ that converges to p as $\mu \to -\infty$. We claim that the system of equations

$$b^T a_x = \log p(x) \quad \text{for all } x \in F. \tag{9}$$

in the variables $b = (b_1, \ldots, b_d)$ has a solution. Otherwise we can find numbers $v(x), x \in \mathcal{X}$, such that $\sum_{x\in F} v(x) \log p(x) \neq 0$ and $\sum_{x\in\mathcal{X}} v(x)a_x = 0$. This leads to the contradiction $p^{v^+} \neq p^{v^-}$.

Fix a vector $c \in \mathbb{R}^d$ with property (6). For any $\mu \in \mathbb{R}$ define

$$p_{(\mu)} := p_{\mu c+b} = \frac{1}{Z_{\mu c+b}} \left(e^{\mu c^T a_1} e^{b^T a_1}, \ldots, e^{\mu c^T a_m} e^{b^T a_m}\right) \in \mathcal{E}_A.$$

By (6) and (9) it follows that $\lim_{\mu\to-\infty} p_{(\mu)} = p$. This proves the theorem. □

We now see that the last statement of Lemma 7 can be generalized (cf. [14, Lemma A.2]):

Proposition 8. *The following are equivalent for any set $F \subseteq \mathcal{X}$:*

(1) *F is facial.*
(2) *The truncation of q to F,*

$$q_F(x) := \begin{cases} \frac{q(x)}{\sum_{x\in F} q(x)}, & \text{if } x \in F, \\ 0 & \text{else,} \end{cases}$$

lies in $\overline{\mathcal{E}_{q,A}}$.
(3) *There is a vector with support F in $\overline{\mathcal{E}_{q,A}}$.*

According to Theorem 4, in order to test whether p is an element of the closure of $\mathcal{E}_{q,A}$, we have to test all the Eq. (5). The next theorem shows that it is actually enough to check finitely many equations. For this, we need the following notion from matroid theory:

Definition 9. A *circuit vector* of a matrix A is a nonzero vector $n \in \ker A \subseteq \mathbb{R}^m$ with inclusion minimal support, i.e if $n' \in \ker A$ satisfies $\mathrm{supp}(n') \subseteq \mathrm{supp}(n)$, then $n' = \lambda n$ for some $\lambda \in \mathbb{R}$. Equivalently, n is the vector of coefficients of a nontrivial linear relation $\sum_x n(x)a_x = 0$ of the columns of A with inclusion minimal support. A *circuit* is the support set of a circuit vector. A *circuit basis* of $\ker A$ is a subset of $\ker A$ containing precisely one circuit vector for every circuit.[1]

The minimality condition implies that the circuit determines its corresponding circuit vectors up to a multiple. If we replace n by a nonzero multiple of n then Eq. (5) is replaced by an equation which is equivalent over the non-negative reals. This means that all systems of equations corresponding to any circuit basis C are equivalent.

Theorem 10. *Let $\mathcal{E}_{q,A}$ be an exponential family. Then $\overline{\mathcal{E}_{q,A}}$ equals the set of all probability distributions that satisfy*

$$p^{c^+} q^{c^-} = p^{c^-} q^{c^+} \text{ for all } c \in C, \tag{10}$$

where C is a circuit basis of A.

[1] It is easy to see that a circuit basis of $\ker A$ spans $\ker A$. However, in general the circuit vectors are not linearly independent.

The proof is based on the following two well-known lemmas. For detailed proofs see [5]. For convenience we sketch the proofs:

Lemma 11. *For every vector $n \in \ker A$ there exists a sign-consistent circuit vector $c \in \ker A$, i.e. if $c(x) \neq 0 \neq n(x)$ then $\operatorname{sgn} c(x) = \operatorname{sgn} n(x)$ for all $x \in \mathcal{X}$.*

Proof. Let c be a vector with inclusion-minimal support that is sign-consistent with n and satisfies $\operatorname{supp}(c) \subseteq \operatorname{supp}(n)$. If c is not a circuit vector, then there exists a circuit vector c' with $\operatorname{supp}(c') \subseteq \operatorname{supp}(c)$. A suitable linear combination $c + \alpha c'$, $\alpha \in \mathbb{R}$ gives a contradiction to the minimality of c. □

Lemma 12. *Every vector $n \in \ker A$ is a finite sign-consistent sum of circuit vectors $n = \sum_{i=1}^{r} c_i$, i.e. if $c_i(x) \neq 0$ then $\operatorname{sgn} c_i(x) = \operatorname{sgn} n(x)$ for all $x \in \mathcal{X}$.*

Proof. Use induction on the size of $\operatorname{supp}(n)$. In the induction step, use a sign-consistent circuit vector, as in the last lemma, to reduce the support. □

Proof of Theorem 10. Again, we can assume that $q(x) = 1$ for all $x \in \mathcal{X}$. By Theorem 4 it suffices to show: If $p \in \mathbb{R}^{\mathcal{X}}$ satisfies (10), then it also satisfies $p^{n^+} = p^{n^-}$ for all $n \in \ker A$. Write $n = \sum_{i=1}^{r} c_i$ as a sign-consistent sum of circuit vectors c_i, as in the last lemma. Without loss of generality we can assume $c_i \in C$ for all i. Then $n^+ = \sum_{i=1}^{r} c_i^+$ and $n^- = \sum_{i=1}^{r} c_i^-$. Hence p satisfies

$$p^{n^+} - p^{n^-} = p^{\sum_{i=2}^{r} c_i^+}\left(p^{c_1^+} - p^{c_1^-}\right) + \left(p^{\sum_{i=2}^{r} c_i^+} - p^{\sum_{i=2}^{r} c_i^-}\right)p^{c_1^-}, \tag{11}$$

so the theorem follows easily by induction. □

Example 13. Consider the following sufficient statistics:

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -\alpha & 1 & 0 & 0 \end{pmatrix}, \tag{12}$$

where $\alpha \notin \{0, 1\}$ is arbitrary. The kernel is then spanned by

$$v_1 = (1, \alpha, -1, -\alpha)^T \quad \text{and} \quad v_2 = (1, \alpha, -\alpha, -1)^T, \tag{13}$$

but these two vectors do not form a circuit basis: They correspond to the two relations

$$p(1)p(2)^{\alpha} = p(3)p(4)^{\alpha} \quad \text{and} \quad p(1)p(2)^{\alpha} = p(3)^{\alpha}p(4). \tag{14}$$

It follows immediately that

$$p(3)p(4)^{\alpha} = p(3)^{\alpha}p(4). \tag{15}$$

If $p(3)p(4)$ is not zero, then we conclude $p(3) = p(4)$. However, on the boundary this does not follow from Eq. (14): Possible solutions to these equations are given by

$$p_a = (0, a, 0, 1 - a) \quad \text{for } 0 \leq a < 1. \tag{16}$$

However, p_a does not lie in the closure of the exponential family $\overline{\mathcal{E}_A}$, since all members of $\mathcal{E}_A$ do satisfy $p(3) = p(4)$.

A circuit basis of A is given by the following vectors:

$$(0, 0, 1, -1)^T \qquad p(3) = p(4), \tag{17a}$$

$$(1, \alpha, 0, -1-\alpha)^T \qquad p(1)p(2)^{\alpha} = p(4)^{1+\alpha}, \tag{17b}$$

$$(1, \alpha, -1-\alpha, 0)^T \qquad p(1)p(2)^{\alpha} = p(3)^{1+\alpha}. \tag{17c}$$

By Theorem 10 these three equations characterize $\overline{\mathcal{E}_A}$.

Remark 14 (*Relation to algebraic statistics*)**.** In the case where the vector space $\ker A$ has a basis with integer components (for example, if A is an integer matrix), every circuit vector is proportional to an integer circuit vector. In this case the corresponding equations (5) are polynomial, and the theorem implies that $\overline{\mathcal{E}_{q,A}}$ is the non-negative real part of a *projective variety*, i.e. the solution set of homogeneous polynomials (see [7] for an introduction to commutative algebra and algebraic geometry). If we want to use the tools of commutative algebra, then circuit vectors are not the right objects to consider: For example, proportional circuit vectors only yield equivalent equations over the non-negative reals, but we may obtain a different solution set if we allow negative real solutions or complex solutions. Therefore, different integer circuit bases do not yield equivalent equations over $\mathbb{C}$. This may greatly increase the running time of many algorithms of computational commutative algebra. One way out is to look at the *ideal* I_c generated by the binomials of all integer valued circuit vectors. Equivalently, I_c corresponds to a "prime circuit basis" C such that the components of any $n \in C$ are integers with greatest common divisor one. I_c is called the *circuit ideal*, and the following discussion shows that the variety V_c of I_c equals the *Zariski closure* of $\mathcal{E}_{q,A}$, i.e. the smallest variety containing $\mathcal{E}_{q,A}$.

Using I_c is still not the best solution, since in general the circuit ideal is not *radical*. This means that there are polynomials that vanish on V_c but which do not lie in I_c. This may also increase the running time of algebraic algorithms. By [13, Proposition 8.7] the radical of I_c is the *toric ideal* I_t generated by the binomials corresponding to all $u \in \ker_{\mathbb{Z}} A$. The fact that I_t is the ideal of the Zariski closure of $\mathcal{E}_{q,A}$ was first noted in [14]. It follows from Theorem 10, knowing that I_t is prime (see [13]) and contains I_c. For further results on the relation between toric ideals and circuit ideals we refer to [3].

Hence, if we want to use algebraic tools, it is best to work with a *Markov basis*, which can be defined as a finite subset of $\ker_{\mathbb{Z}} A$ such that the corresponding binomials generate I_t. One major application of Markov bases makes use of the following fact, which was first noted and applied in statistics by Diaconis and Sturmfels [11]: *A finite set $\mathcal{B} \subseteq \ker A$ is a Markov basis if and only if the following holds: For all $h, h' \in \mathbb{N}_0^{\mathcal{X}}$ such that $Ah = Ah'$ there exists a sequence $(b_i)_{i=1}^s \subseteq \pm\mathcal{B}$ such that $h' = h + \sum_{i=1}^s b_i$ and such that $h + \sum_{i=1}^r b_i \in \mathbb{N}_0^{\mathcal{X}}$ for all $1 \leq r \leq s$.* Therefore, Markov bases elements can be used as moves to explore integer points in polytopes using a Markov Chain Monte Carlo algorithm. A circuit basis is a natural generalisation, in the following sense: *Let h, h' be two non-negative vectors in $\mathbb{R}^{\mathcal{X}}$ such that $Ah = Ah'$, and let $\mathcal{C}$ be a circuit basis. Then there exists a sequence $(c_i)_{i=1}^s \subseteq \mathcal{C}$ and real numbers $\alpha_i \in \mathbb{R}$ such that $h' = h + \sum_{i=1}^s \alpha_i c_i$ and such that $h + \sum_{i=1}^r \alpha_i c_i$ is non-negative for all $1 \leq r \leq s$* (this follows from Lemma 12). Therefore, in principle, a circuit basis could be used to explore polytopes using a Markov Chain Monte Carlo algorithm, drawing the elements $c_i \in \mathcal{C}$ and the coefficients α_i randomly. Note that the converse of the statement on circuit bases does not hold in general. This is related to the fact that circuit bases do not correspond to minimal systems of implicit equations characterizing the exponential family. Another viewpoint is that circuit bases are more related to Graver bases, see [15].

Finding a Markov basis or a circuit basis is in general a non-trivial task. Hemmecke and Malkin [23] and Malkin [16] discuss algorithms for both tasks, which are implemented in the open source software package 4ti2 [1]. Markov basis computations tend to depend on the size of the entries of the matrix: If A has only small entries, then one may hope that there are "enough" vectors in $\ker_{\mathbb{Z}} A$ with small entries, corresponding to polynomials of low degree. The Markov bases algorithm is related to Buchberger's algorithm, whose speed depends on the degrees of the starting polynomials. Circuit computations do not depend essentially on the size of the entries of A, but the number of circuits tends to be much larger than the number of Markov basis elements. In our experience Markov basis computations are faster when A has only "small" entries (which is the most important case for applications), and circuit computations are faster when A has "large" entries.

Example 15. For an example of a Markov basis which contains noncircuits consider the matrix

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix}. \tag{18}$$

A quick calculation with the software 4ti2 gives the following circuit basis of A:

$$\begin{aligned} u_1 &:= (0, 1, -2, 1), \quad u_2 := (1, -2, 1, 0), \\ u_3 &:= (1, 0, -3, 2), \quad u_4 := (2, -3, 0, 1). \end{aligned} \tag{19}$$

However, any Markov basis contains $v := (1, -1, -1, 1)$, which is obviously not a circuit vector. v corresponds to the binomial $p_1p_4 - p_2p_3$. The computation

$$(p_1p_4 - p_2p_3)^2 = p_4(p_1^2p_4 - p_2^3) + p_2^2(p_3^2 - p_2p_4) - 2p_2p_4(p_1p_3 - p_2^2)$$

illustrates that the radical of the circuit ideal is the toric ideal, i.e. any solution of the toric ideal also solves $p_1p_4 - p_2p_3 = 0$.

It is not easy to find a *hierarchical* model whose Markov basis does not consist of circuit vectors. In [2], Aoki and Takemura give an example. Interestingly, neither the full Markov basis nor a circuit basis of this model are known.

Remark 16. Using arguments from matroid theory the number of circuits can be shown to be less than or equal to $\binom{m}{r+2}$, where $m = |\mathcal{X}|$ is the size of the state space and r is the dimension of $\mathcal{E}_{q,A}$, see [12]. This gives an upper bound on the number of implicit equations describing $\overline{\mathcal{E}_{q,A}}$. Note that $\binom{m}{r+2}$ is usually much larger than the codimension $m - r - 1$ of $\mathcal{E}_{q,A}$ in the probability simplex. In contrast, if we only want to find an implicit description of all probability distributions of $\mathcal{E}_{q,A}$, which have full support, then $m - r - 1$ equations are enough: We can test $p \in \mathcal{E}_{q,A}$ by checking whether $\log(p/q)$ lies in the row span of A. This amounts to checking whether $\log(p/q)$ is orthogonal to $\ker A$, which is equivalent to $m - r - 1$ equations after choosing a basis of $\ker A$.

It turns out that even in the boundary the number of equations can be further reduced: In general we do not need all circuits for the implicit description of $\overline{\mathcal{E}_{q,A}}$. For instance, in Example 13, Eqs. (17b) and (17c) are equivalent given (17a), i.e. we only need two of the three circuits to describe $\overline{\mathcal{E}_{q,A}}$. Unfortunately we do not know how to find a minimal subset of circuits that characterizes the closure of the exponential family. In the algebraic case discussed in the previous remark this question is equivalent to determining a minimal generating set of the circuit ideal among the circuit vectors.

2.2. Support sets of exponential families

Now we focus on the following problem: Given a set $S \subseteq \mathcal{X}$, is there a probability distribution $p \in \overline{\mathcal{E}_{q,A}}$ satisfying $\operatorname{supp}(p) = S$? In other words, we want to characterize the set

$$\mathcal{S}_{q,A} := \{\operatorname{supp}(p) : p \in \overline{\mathcal{E}_{q,A}}\} \subseteq 2^{\mathcal{X}}. \tag{20}$$

Proposition 8 and Lemma 7 give the following characterization: A nonempty set $S \subseteq \mathcal{X}$ is the support set of some distribution $p \in \overline{\mathcal{E}_{q,A}}$ if and only if the following holds for all circuit vectors $n \in \ker A$:

- $\operatorname{supp}(n^+) \subseteq S$ if and only if $\operatorname{supp}(n^-) \subseteq S$.

Obviously, this condition does not depend on the circuits themselves, but only on the supports of their positive and negative part.

Definition 17. A *signed subset* (M, N) of $\mathcal{X}$ is a pair of disjoint subsets $M, N \subseteq \mathcal{X}$. Alternatively, a signed subset (M, N) can be represented as a sign vector $X \in \{-1, 0, +1\}^{\mathcal{X}}$, where

$$X(x) = \begin{cases} +1, & \text{if } x \in M, \\ -1, & \text{if } x \in N, \\ 0, & \text{else.} \end{cases} \tag{21}$$

As a slight abuse of notation, we do not make a difference between these two representations in the following.

Consider the map

$$\operatorname{sgn}\colon\ n \mapsto (\operatorname{supp}(n^+), \operatorname{supp}(n^-)),$$

which associates to each vector a signed subset of $\mathcal{X}$. In the sign vector representation, sgn corresponds to the usual signum mapping applied componentwise to vectors. The signed subset $\operatorname{sgn}(c)$ corresponding to a circuit vector $c \in \ker A$ shall be called a *signed circuit*. The set of all signed circuits is denoted by

$$\mathcal{C}(A) := \pm\operatorname{sgn}(C) = \{\operatorname{sgn}(c) : c \in C \text{ or } c \in -C\}, \tag{22}$$

where C is a circuit basis of A.

Note that the set of signed circuits is twice as large as a circuit basis, so the set of signed circuits carries a lot of redundant information, which should be removed when doing calculations with $\mathcal{C}(A)$. However, for theoretical purposes it is advantageous to work with the symmetric set $\mathcal{C}(A)$.

We immediately have the following

Theorem 18. *Let S be a nonempty subset of $\mathcal{X}$. Then $S \in \mathcal{S}_{q,A}$ if and only if the following holds for all signed circuits $(M, N) \in \mathcal{C}(A)$:*

$$M \subseteq S \quad \Leftrightarrow \quad N \subseteq S. \tag{23}$$

Corollary 19. *If two matrices A_1, A_2 satisfy $\mathcal{C}(A_1) = \mathcal{C}(A_2)$ then the possible support sets of the corresponding exponential families $\mathcal{E}_{q_1,A_1}$ and $\mathcal{E}_{q_2,A_2}$ coincide.*

According to Remark 16, Theorem 18 gives up to $\binom{m}{r+2}$ conditions on the support. Usually, some of these conditions are redundant, but it is not easy to see a priori, which conditions are essential. Of course, a necessary condition for a subset S of $\mathcal{X}$ to be a support set of a distribution contained in $\overline{\mathcal{E}_A}$ is condition (23) restricted to pairs from a subset $\mathcal{H} \subseteq \mathcal{C}(A)$. For example, one can take $\mathcal{H} := \operatorname{sgn}(B)$, where B is a finite subset of $\ker A$, such as a basis.

Example 20. We continue Example 13. If $\alpha > 0$ then we deduce the following implications from the circuits:

$$p(3) \neq 0 \iff p(4) \neq 0, \tag{24a}$$
$$p(1) \neq 0 \text{ and } p(2) \neq 0 \iff p(4) \neq 0, \tag{24b}$$
$$p(1) \neq 0 \text{ and } p(2) \neq 0 \iff p(3) \neq 0. \tag{24c}$$

Again, as above, the last two implications are equivalent given the first.

From this it follows easily that the possible support sets in this example are $\{1\}$, $\{2\}$ and $\{1, 2, 3, 4\}$. From the spanning set (13) we only obtain the implication

$$p(1) \neq 0 \text{ and } p(2) \neq 0 \iff p(3) \neq 0 \text{ and } p(4) \neq 0. \tag{25}$$

We conclude this section with two examples where a complete characterization of the face lattice of the convex support and thus of the possible supports is easily achievable.

Example 21 (*Supports in the binary no-n-way interaction model*)**.** Consider the binary hierarchical model [20] whose simplicial complex is the boundary of an n simplex. If $n = 3$, this model is called the no-3-way interaction model, and its Markov bases have been recognized to be arbitrarily complicated [21], so we do not expect to find an easy description of the signed circuits. However, if we restrict ourselves to binary variables $x = (x_i)_{i=1}^n \in \mathcal{X} := \{0, 1\}^n$, the structure is very simple. In this case the exponential family is of dimension $2^n - 2$, i.e. of codimension 1 in the simplex, so $\ker A$ is one dimensional. It is spanned by the "parity function"

$$e_{[n]}(x) := \begin{cases} -1 & \text{if } \sum_{i=1}^n x_i \text{ is odd,} \\ 1 & \text{otherwise.} \end{cases} \tag{26}$$

Using Theorem 18 we can easily describe the face lattice of the marginal polytope (i.e. convex support) $P^{(n-1)}$: A set $\mathcal{Y} \subsetneq \{0, 1\}^n$ is a support set if and only if it does not contain all configurations with even parity, or all configurations with odd parity. It follows that $P^{(n-1)}$ is *neighborly*, i.e. the convex hull of any $\left\lfloor \frac{\dim(P^{(n-1)})}{2} \right\rfloor = 2^{n-1} - 1$ vertices is a face of the polytope. To see this, note that no set of cardinality less than 2^{n-1} can contain all configurations with even or odd parity. We can easily count the support sets by counting the non-faces of the corresponding marginal polytope, i.e. all sets $\mathcal{Y}$ that contain either the configurations with even parity, or the configurations with odd parity. Let s_k be the number of support sets of cardinality k, i.e. the number of faces with k vertices. It is given by:

$$s_k = \binom{2^n}{k} - 2\binom{2^{n-1}}{k - 2^{n-1}}, \tag{27}$$

where $\binom{m}{l} = 0$ if $l < 0$. Since this polytope has only one affine dependency (26) which includes all the vertices, we see that it is *simplicial*, i.e. all its faces are simplices. It follows that f_k, the number of k-dimensional faces, is given by $f_k = s_{k+1}$.

Altogether we have determined the face lattice of the polytope, which means that we know the "combinatorial type" of the polytope. It turns out that the face lattice of $P^{(n-1)}$ is isomorphic to the face lattice of the $(2^n - 2)$-dimensional *cyclic polytope* with 2^n vertices.

Next, we take a closer look at cyclic polytopes. Define the *moment curve* in $\mathbb{R}^d$ by

$$\boldsymbol{x} : \mathbb{R} \to \mathbb{R}^d, \quad t \mapsto \boldsymbol{x}(t) := \left(t, t^2, \cdots, t^d\right)^T. \tag{28}$$

The *d-dimensional cyclic polytope with n vertices* is

$$C(d, n) := \operatorname{conv}\{\boldsymbol{x}(t_1), \ldots, \boldsymbol{x}(t_n)\}, \tag{29}$$

the convex hull of $n > d$ distinct points $(t_1 < t_2 < \cdots < t_n)$ on the moment curve. The face lattice of a cyclic polytope can easily be described using *Gale's evenness condition*, see [29]. The cyclic polytope is simplicial and neighborly, i.e. the convex hull of any $\lfloor \frac{d}{2} \rfloor$ vertices is a face of $C(n, d)$, but even better, one has

Theorem 22 (Upper Bound Theorem)**.** *If P is a d-dimensional polytope with $n = f_0$ vertices, then for every k it has at most as many k-dimensional faces as the cyclic polytope $C(d, n)$:*

$$f_k(P) \leq f_k(C(d, n)), \quad k = 0, \ldots, d. \tag{30}$$

If equality holds for some k with $\lfloor \frac{d}{2} \rfloor \leq k \leq d$ then P is neighborly.

Theorem 22 was conjectured by Motzkin in 1957 and its proof has a long and complicated history. The final result is due to McMullen [24].

The Upper Bound Theorem shows that the exponential families constructed above have the largest number of support sets among all exponential families with the same dimension and the same number of vertices. Finally, we consider a cyclic polytope of dimension two which gives an exponential family of smallest dimension containing all the vertices of the probability simplex. The construction is due to [22].

Example 23. Let $\mathcal{X} = \{1, \ldots, m\}$ and consider the matrix A, whose columns are the points on the 2-dimensional moment curve, augmented with row $(1, \ldots, 1)$:

$$A := \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 3 & \dots & m \\ 1 & 4 & 9 & \dots & m^2 \end{pmatrix}. \tag{31}$$

This matrix defines a two-dimensional exponential family. To approximate an arbitrary extreme point δ_x of the probability simplex, consider the parameter vector $\theta = (x^2, -2x, 1)^T$, giving rise to probability measures $p_{\beta\theta} = \frac{1}{Z}\exp(-\beta\theta^T A)$. Since $\theta^T a_y = (y-x)^2$, we get that $\lim_{\beta\to\infty} p_{\beta,\theta} = \delta_x$.

Summarizing we see that cyclic polytopes, owing to their extremal properties, have something to offer not only for convex geometry, but also for statistics.

2.3. Parametric description of $\overline{\mathcal{E}_{q,A}}$

We have seen how the implicit description of an exponential family can be used as a tool to investigate the possible support sets of an exponential family. It is also possible to find a parametrization of the exponential family which extends to the boundary. However, in general this parametrization will not be injective.

In the following we assume that $A \in \mathbb{R}_+^{d\times m}$ has only non-negative real entries. Since the rowspan of A contains the constant row $(1, \dots, 1)$ it is always possible to replace an arbitrary matrix by a non-negative matrix without changing the exponential family (see Remark 2). We now replace the exponential parametrization in (3) by a "monomial" parametrization

$$\mathbb{R}_+^d \ni \xi \mapsto \hat{p}_\xi(x) = \frac{q(x)}{Z_\xi}\prod_{j=1}^d \xi_j^{a_{j,x}} \tag{32}$$

Then $\hat{p}_\xi = p_\theta$ if $\xi_j = \exp(\theta_j)$. Formula (32) also makes sense when some of the parameters are zero, as long as $Z_\xi \neq 0$. In this case $\hat{p}_\xi$ will describe an element from the closure of the exponential family.

The first question to answer is which parts of the closure can be reached for a given matrix A. Let

$$\hat{\mathcal{E}}_{q,A} := \{\hat{p}_\xi : \xi_j \geq 0 \text{ for all } j, \text{ and } Z_\xi > 0\} \tag{33}$$

be the *image of the monomial parametrization*. We want to characterize the support sets of $\hat{\mathcal{E}}_{q,A}$.

Definition 24. A subset $F \subseteq \mathcal{X}$ is called *A-feasible* if for every $x \in \mathcal{X}\setminus F$ the set $\operatorname{supp}(a_x) = \{j \in \{1,\dots,d\} : a_{j,x} \neq 0\}$ is not contained in $\bigcup_{y\in F}\operatorname{supp}(a_y)$.

The following proposition is a mild generalization of [14, Theorem 3.1]; the proofs carry over without difficulty.

Proposition 25. *Assume that A has no negative entries. A probability measure $p \in \overline{\mathcal{E}_{q,A}}$ lies in $\hat{\mathcal{E}}_{q,A}$ if and only if* $\operatorname{supp}(p)$ *is A-feasible.*

Proof. If $\hat{p}_\xi(x) = 0$ then $\xi_j = 0$ for some $j \in \operatorname{supp}(a_x)$. This implies $a_{j,y} = 0$ for all $y \in \operatorname{supp}(\hat{p}_\xi)$, so $\operatorname{supp}(a_x)$ is not contained in $\bigcup_{y\in\operatorname{supp}(\hat{p}_\xi)}\operatorname{supp}(a_y)$, showing that $\operatorname{supp}(\hat{p}_\xi)$ is A-feasible.

For the other direction we may assume that $q(x) = 1$ for all $x \in \mathcal{X}$; the general case then follows readily. Let $p \in \overline{\mathcal{E}_{q,A}}$. We need to show that the system of equations

$$\prod_{i=1}^d \xi_i^{a_{i,x}} = p(x), \text{ for all } x \in \mathcal{X}, \tag{34}$$

has a solution. In the proof of Theorem 4 it was shown that the related system (9) has a solution b. Let $\hat{\xi}_i = \exp(b_i)$ for all i. Note that b_i (and also $\hat{\xi}_i$) is not restricted by Eq. (9) if i is not in $\bigcup_{x\in\operatorname{supp}(p)}\operatorname{supp}(a_x)$. Put

$$\xi_i = \begin{cases} \hat{\xi}_i, & \text{if } i \in \bigcup_{x\in\operatorname{supp}(p)}\operatorname{supp}(a_x), \\ 0, & \text{else.} \end{cases} \tag{35}$$

Then $\hat{p}_\xi(x) = 0$ if and only if $p(x) = 0$, by definition of A-feasibility, and $\hat{p}_\xi(x) = p(x)$ if $p(x) > 0$ by definition of $\hat{\xi}$. $\square$

In order to parametrize the closure of an exponential family $\mathcal{E}_{q,A}$ with the help of the monomial parametrization we need to find a matrix A' such that $\mathcal{E}_{q,A} = \mathcal{E}_{q,A'}$ and such that every possible support set is A'-feasible. In order to define A' we need the following notion from matroid theory:

Definition 26. A *cocircuit vector* of a matrix A is a vector v in the row span of A with inclusion minimal support. A *cocircuit* is the support set of a cocircuit vector. A cocircuit vector is positive if all of its components are non-negative. A positive cocircuit is the support set of a positive cocircuit vector.

Note the similarity to the definition of a circuit vector. In fact, if A^* is a matrix the rows of which span $\ker A$, then the cocircuit vectors of A equal the circuit vectors of A^*, and vice versa (see Remark 35). Consequently, Lemmas 11 and 12 remain valid if $\ker A$ is replaced by the row span of A and if the word "circuit" is replaced by "cocircuit".

Theorem 27. *Let $\mathcal{E}_{q,A}$ be an exponential family. Let A' be a matrix the rows of which contain one positive cocircuit vector for every positive cocircuit of A. Then $\mathcal{E}_{q,A} = \mathcal{E}_{q,A'}$, and the image of the monomial parametrization of $\mathcal{E}_{q,A'}$ consists of $\overline{\mathcal{E}_{q,A'}}$.*

Proof. As above we may assume that A has no negative entries.

For the first statement we need to show that A and A' have the same row space (cf. Remark 2). By definition every row of A' is a linear combination of rows from A. For the other direction we may use Lemma 12 by the remark before the theorem. By assumption all entries of A are non-negative, so every row of A is a linear combination of positive cocircuit vectors.

For the second statement it is enough to prove that every support set which occurs in $\overline{\mathcal{E}_{q,A}}$ is A'-feasible. Let $F = \operatorname{supp}(p)$ for some $p \in \overline{\mathcal{E}_{q,A}}$ and fix $x \in \mathcal{X} \setminus F$. By Proposition 8 F is facial, so there exists $c \in \mathbb{R}^d$ as in Definition 6. The row vector $v = c^T A$ is positive and lies in the row span of A. Furthermore $v(x) > 0$. By Lemma 12 and the remark before the theorem there is a positive cocircuit vector u such that $u(x) > 0$ and $\operatorname{supp}(u) \cap F = \emptyset$. It follows that $\operatorname{supp}(a'_x)$ is not a subset of $\bigcup_{y \in F} \operatorname{supp}(a'_y)$, where a'_y are the columns of A'. Therefore F is A'-feasible. □

It is easy to see that the matrix A' is also the smallest matrix satisfying the conclusions of the theorem: Let v be a positive cocircuit vector. Using the parametrization induced by A' it is easy to see that $\overline{\mathcal{E}_{q,A}} = \overline{\mathcal{E}_{q,A'}}$ contains a probability measure p with support $F := \operatorname{supp}(p) = \mathcal{X} \setminus \operatorname{supp}(v)$ (set the parameter corresponding to the row of v in A' to zero). Now suppose that $p \in \hat{\mathcal{E}}_{q,A}$. Then F is A-feasible. Fix $x \in \operatorname{supp}(v)$. Then there is an index i such that $a_{i,x} \neq 0$, but $a_{j,y} = 0$ for all $y \in F$. This means that the support of the ith row of A is contained in the support of v. Since v is a cocircuit vector, it follows that A contains a row which is proportional to v.

Example 28. We continue Examples 13 and 20. If $\alpha > 0$, then we have

$$A' = \begin{pmatrix} 0 & 1+\alpha & \alpha & \alpha \\ 1+\alpha & 0 & 1 & 1 \end{pmatrix}.$$

Quite generally, if the exponential family is onedimensional, then two parameters are enough to parametrize its closure. This fact now follows directly from oriented matroid theory (see Section 3): A one-dimensional polytope has only two facets, therefore the corresponding oriented matroid has exactly two positive cocircuits.

Remark 29. Theorem 27 is related to results of Katsabekis and Thoma [18,19], who study the image of the monomial parametrization of a toric variety over an arbitrary field. They show that there is a surjective monomial parametrization of any toric variety over an algebraically closed field. However, there are toric varieties over $\mathbb{R}$ which have no surjective monomial parametrization. For exponential families this problem disappears, since we are only concerned with non-negative real numbers.

Related results were proved in [26]. Under the assumption that A has only integer entries it is shown that the monomial parametrization can be improved by replacing A with the matrix A'' the rows of which consist of a *Hilbert basis* [2] of the row span of A. However, it is not verified that each possible support set is A''-feasible. Of course, this is evident by the previous theorem.

Using the Hilbert basis has two disadvantages compared to the positive cocircuit vectors proposed here: Hilbert bases are only defined for integer A, and furthermore it is much more difficult to compute a Hilbert basis than to compute the cocircuits of a matrix.

3. Relations to oriented matroids

In this section the results from the previous section are related to the theory of oriented matroids. The proofs in this section are only sketched, since the main results of this work have already been proved directly. We refer to Chapters 1–3 of [5] for a more detailed introduction to oriented matroids.

[2] See [26] for the definition of a Hilbert basis in the setting of linear programming.

Definition 30. Let E be a finite set and $\mathcal{C}$ a non-empty collection of signed subsets of E (see the previous section). For every signed set $X = (X^+, X^-)$ of E we let $\underline{X} := X^+ \cup X^-$ denote the *support* of X. Furthermore, the *opposite signed set* is $-X = (X^-, X^+)$. Then the pair $(E, \mathcal{C})$ is called an *oriented matroid* if the following conditions are satisfied:

(C1) $\mathcal{C} = -\mathcal{C}$, (*symmetry*)
(C2) for all $X, Y \in \mathcal{C}$, if $\underline{X} \subseteq \underline{Y}$, then $X = Y$ or $X = -Y$, (*incomparability*)
(C3) for all $X, Y \in \mathcal{C}, X \neq -Y$, and $e \in X^+ \cap Y^-$ there is a $Z \in \mathcal{C}$ such that $Z^+ \subseteq (X^+ \cup Y^+) \setminus \{e\}$ and $Z^- \subseteq (X^- \cup Y^-) \setminus \{e\}$. (*weak elimination*)

In this case each element of $\mathcal{C}$ is called a *signed circuit*.

Note that to every oriented matroid $(E, \mathcal{C})$ we have an associated unoriented matroid (E, C), called the *underlying matroid*, where

$$C = \left\{X^+ \cup X^- = \underline{X} : X \in \mathcal{C}\right\} \tag{36}$$

is the set of *circuits* of (E, C). In this way oriented matroids can be considered as ordinary matroids endowed with an additional structure, namely a *circuit orientation* which assigns two opposite signed circuits $\pm X \in \mathcal{C}$ to every circuit $\underline{X} \in C$.

The most important example of an oriented matroid here is the oriented matroid of a matrix $A \subseteq \mathbb{R}^{d\times m}$. In this case let $E = \mathcal{X} = \{1, \dots, m\}$, and let

$$\mathcal{C} = \left\{(\mathrm{supp}(n^+), \mathrm{supp}(n^-)) : n \in \ker A \text{ has inclusion minimal support}\right\}. \tag{37}$$

An oriented matroid is called *realizable* if it is induced by some matrix A.[3]

The only axiom which is not trivially fulfilled for this example is **(C3)**. However, if we drop the minimality condition and let $\mathcal{V} = \{(\mathrm{supp}(n^+), \mathrm{supp}(n^-) : n \in \ker A\}$, then it is easy to see that $\mathcal{V}$ satisfies **(C3)**. Thus $(E, \mathcal{C})$ satisfies **(C3)** by the following proposition:

Proposition 31. *Let $\mathcal{V}$ be a nonempty collection of signed subsets of E satisfying* **(C1)** *and* **(C3)**. *Write* $\mathrm{Min}(\mathcal{V})$ *for the minimal elements of $\mathcal{V}$ (with respect to inclusion of supports). Then*

(1) *for any $X \in \mathcal{V}$ there is $Y \in \mathrm{Min}(\mathcal{V})$ such that $Y^+ \subseteq X^+$ and $Y^- \subseteq X^-$.*
(2) $\mathrm{Min}(\mathcal{V})$ *is the set of circuits of an oriented matroid.*

Proof. [5], Proposition 3.2.4. □

This illustrates how **(C2)** corresponds to the minimality condition. It is possible to define oriented matroids without this minimality condition using the following construction:

Definition 32. The *composition* of two signed subsets X, Y of E is the signed subset $X \circ Y$ with

$$(X \circ Y)^+ := X^+ \cup (Y^+ \setminus X^-), \quad (X \circ Y)^- := X^- \cup (Y^- \setminus X^+). \tag{38}$$

A composition $X \circ Y$ is *conformal* if X and Y are *sign-consistent*, i.e. $X^+ \cap Y^- = \emptyset = X^- \cap Y^+$.

An *o.m. vector* of an oriented matroid is any composition of an arbitrary number of circuits.[4] The set of o.m. vectors shall be denoted by $\mathcal{V}$. If the oriented matroid comes from a matrix A, then $\mathcal{V}$ equals the set $\mathcal{V}$ from above.

Note that composition is associative but not commutative in general.

The above proposition shows that an oriented matroid can alternatively be defined as a pair $(E, \mathcal{V})$, where $\mathcal{V}$ is a collection of signed subsets satisfying **(C1)**, **(C3)** and

(V0) $\emptyset \in \mathcal{V}$,
(V2) for all $X, Y \in \mathcal{V}$ we have $X \circ Y \in \mathcal{V}$.

Note that in the realizable case linear combinations of vectors correspond to composition of their sign vectors in the following sense:

$$\lim_{\epsilon\to 0, \epsilon>0} \mathrm{sgn}(n + \epsilon n') = \mathrm{sgn}(n) \circ \mathrm{sgn}(n'). \tag{39}$$

[3] Note that this definition depends, in fact, only on the kernel of A, compare Remark 2.
[4] In [5], o.m. vectors are simply called vectors. The name "o.m. vector" has been proposed by F. Matúš to avoid confusion.

Now Lemmas 11 and 12 correspond to the following two lemmas:

Lemma 11'. *For every o.m. vector Y there exists a sign-consistent signed circuit X such that $\underline{X} \subseteq \underline{Y}$.*

Lemma 12'. *Any o.m. vector is a conformal composition of circuits.*

To every matrix A we can associate a polytope which was called convex support in the last section. Many properties of this polytope can be translated into the language of oriented matroids. This yields constructions which also make sense, if the oriented matroid is not realizable. In order to make this more precise, we need the notion of the dual oriented matroid. The general construction of the dual of an oriented matroid is beyond the scope of this work. Here, we only state the definition for realizable oriented matroids.

In the following we assume that the matrix A has the constant vector $(1, \ldots, 1)$ in its rowspace. This means that all the column vectors a_x lie in a hyperplane $l_1 = 1$. In the general case, this can always be achieved by adding another dimension. Technically we require that the face lattice of the polytope spanned by the columns of A is combinatorially equivalent to the face lattice of the cone over the columns. See also the remarks before Definition 6.

For every dual vector $l \in (\mathbb{R}^d)^*$ let $N_l^+ := \{x \in \mathcal{X} : l(a_x) > 0\}$ and $N_l^- := \{x \in \mathcal{X} : l(a_x) < 0\}$. This way we can associate a signed subset $\mathrm{sgn}^*(l) := (N_l^+, N_l^-)$ with l. The signed subset $\mathrm{sgn}^*(l)$ is called a *covector*. Let $\mathcal{L}$ be the set of all covectors. If the signed subset (N_l^+, N_l^-) has minimal support (i.e. "many" vectors a_x lie on the hyperplane $l = 0$), then l is called a *cocircuit vector*, and $\mathrm{sgn}^*(l)$ is called a *signed cocircuit*. The collection of all signed cocircuits shall be denoted by $\mathcal{C}^*$.

Lemma 33. *Let $(E, \mathcal{C})$ be an oriented matroid induced by a matrix A. Then $(E, \mathcal{C}^*)$ is an oriented matroid, called the dual oriented matroid.*

Proof. See Section 3.4 of [5]. □

Note that the faces of the polytope correspond to hyperplanes such that all vertices lie on one side of this hyperplane, compare Definition 6. Thus the faces of the polytope are in a one-to-one relation with the positive covectors, i.e. the covectors $X = (X^+, X^-)$ such that $X^- = \emptyset$. The face lattice of the polytope can be reconstructed by partially ordering the positive covectors by inclusion of their supports; however, the relation needs to be inverted: Covectors with small support correspond to faces which contain many vertices. The empty face (which is induced, for example, by the dual vector l_1 which defines the hyperplane containing all a_x) corresponds to the covector $T := (\mathcal{X}, \emptyset)$. This correspondence of faces and positive covectors shows that the parameterization of Theorem 27 is, in fact, related to the face structure of the convex support.

We can apply these remarks to all abstract oriented matroids such that $T = (\mathcal{X}, \emptyset)$ is a covector. Such an oriented matroid is usually called *acyclic*. Thus a face of an acyclic oriented matroid is any positive covector. A *vertex* is a maximal positive covector X in $\mathcal{L} \setminus \{T\}$, i.e. if $\underline{X} \subseteq \underline{Y}$ for some positive covector $Y \in \mathcal{L} \setminus \{X\}$, then $Y = T$.

In this setting we have the following result, which clearly corresponds to the second statement of Lemma 7:

Proposition 34 (Las Vergnas)**.** *Let $(E, \mathcal{C})$ be an acyclic oriented matroid. For any subset $F \subseteq E$ the following are equivalent:*

- *F is a face of the oriented matroid.*
- *For every signed circuit $X \in \mathcal{C}$, if $X^+ \subseteq F$ then $X^- \subseteq F$.*

Proof. The proof of Proposition 9.1.2 in [5] applies (note that the statement of Proposition 9.1.2 includes an additional assumption which is never used in the proof). □

With the help of the moment map defined in the previous section, this proposition can be used to easily derive Theorem 18: By the properties of the moment map, every face of the convex support corresponds to a possible support set of an exponential family, and the proposition links this to the signed circuits of the corresponding oriented matroid.

Finally, Corollary 19 can be rewritten as

Corollary 19'. *The possible support sets of two exponential families coincide if they have the same oriented matroids.*

Unfortunately, this correspondence is not one-to-one: Different oriented matroids can yield the same face lattice, i.e. combinatorially equivalent polytopes. A simple example is given by a regular and a non-regular octahedron as described in [29]. The special case has a name: an oriented matroid is *rigid*, if its positive covectors (i.e. its face lattice) determine all covectors (i.e. the whole oriented matroid). Still, Corollary 19' implies that the instruments of the theory of oriented matroids should suffice to describe the support sets of an exponential family.

Remark 35 (*Importance of Duality*)**.** There are mainly two reasons why the theory of oriented matroids (as well as the theory of ordinary matroids) is considered important. First, it yields an abstract framework which allows to describe a multitude of different combinatorial questions in a unified manner. This, of course, does not in itself lead to any new theorem. The second reason is that the theory provides the important tool of matroid duality.

It turns out that the dual of a realizable matroid is again realizable: If A is a matrix representing an oriented matroid $(E, \mathcal{C})$, then any matrix A^* such that the rows of A^* span the orthogonal complement of the row span of A represents the oriented matroid $(E, \mathcal{C}^*)$.

To motivate the importance of this construction we sketch its implications for the case that the oriented matroid comes from a polytope. In this case the duality is known under the name *Gale transform* [29, Chapter 6]. A d-dimensional polytope with N vertices can be represented by N vectors in $\mathbb{R}^{d+1}$ lying in a hyperplane. These vectors form a $(d+1) \times N$-matrix A. Now we can find an $(N-d-1) \times N$-matrix A^* as above, so the dual matroid is represented by a configuration of N vectors in $\mathbb{R}^{N-d-1}$. This means that this construction allows us to obtain a lowdimensional image of a highdimensional polytope, as long as the number of vertices is not much larger than the dimension. This method has been used for example in [28] in order to construct polytopes with quite unintuitive properties, leading to the rejection of some conjectures. Furthermore, oriented matroid duality makes it possible to classify polytopes with "few vertices" by classifying vector configurations.

The notion of dimension generalizes to arbitrary oriented matroids (and ordinary matroids). In the general setting one usually talks about the *rank* of a matroid, which is defined as the maximal cardinality of a subset $E \subseteq F$ such that E contains no support of a signed circuit. In this sense duality exchanges examples of high rank and low rank, where "high" and "low" is relative to $|E|$.

Remark 36 (*Computing with Matroids*). In Section 2 we already recommended 4ti2 as a tool to compute the circuit vectors of a matrix. Alternatively, TOPCOM [25] is a software package which allows to do many common computations with oriented matroids. Usually, the first step of a calculation is the extraction of the circuits from a matrix. Both programs can only work with integer matrices. Note that it is difficult to treat "arbitrary" (real) matrices on a computer. Rounding the matrix entries to a floating number produces essentially rational matrices, and rational matrices can be turned into integer matrices by multiplying the matrix with the least common multiple of the denominators of all its entries. However, even small changes to the matrix can change the oriented matroid (generically, d vectors in $\mathbb{R}^d$ will be independent after adding small rounding errors). In principle, the algorithms mentioned in Remark 14 can also be used with floating point entries, if a robust criterion is available for checking when a floating number appearing in this algorithm numerically vanishes.

Acknowledgement

The authors want to thank Fero Matúš and Bastian Steudel for many discussions. Nihat Ay has been supported by the Santa Fe Institute. Thomas Kahle has been working on this article during his time at the Max Planck Institute for Mathematics in Leipzig, supported by the Volkswagen Foundation.

This paper is an expanded version of a contribution to WUPES'09, which did not include the material of Section 2.3.

References

[1] 4ti2 team, 4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces. Available at <www.4ti2.de>.
[2] Satoshi Aoki, Akimichi Takemura, The list of indispensable moves of the unique minimal markov basis for $3 \times 4 \times K$ and $4 \times 4 \times 4$ contingency tables with fixed two-dimensional marginals, METR 2003-38, University of Tokyo, Japan, 2003.
[3] Tristram Bogart, Anders N. Jensen, Rekha R. Thomas, The circuit ideal of a vector configuration, Journal of Algebra 309 (2) (2007) 518–542.
[4] Ole E. Barndorff-Nielsen, Information and Exponential Families in Statistical Theory, Wiley, 1978.
[5] A. Björner, M. Las Vergnas, B. Sturmfels, N. White, G. Ziegler, Oriented matroids, Encyclopedia of Mathematics and its Applications, Cambridge, 1993.
[6] Nikolai N. Chentsov, Statistical Decision Rules and Optimal Inference, American Mathematical Society, Providence, RI, 1982. (Russian original: Nauka, Moscow, 1972).
[7] David A. Cox, John B. Little, Don O'Shea, Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra, Springer, 2008.
[8] Imre Csiszár, František Matúš, Closures of exponential families, The Annals of Probability 33 (2005) 582–600.
[9] Imre Csiszár, František Matúš, Generalized maximum likelihood extimates for exponential families, Probability Theory and Related Fields 141 (2008) 213–246.
[10] Michel M. Deza, Monique Laurent, Geometry of cuts and metrics, Algorithms and Combinatorics, Springer, 1997
[11] Persi Diaconis, Bernd Sturmfels, Algebraic algorithms for sampling from conditional distributions, Annals of Statistics 26 (1998) 363–397.
[12] György Dósa, István Szalkai, Claude Laflamme, The maximum and minimum number of circuits and bases of matroids, Pure Mathematics and Applications 15 (4) (2004) 383–392.
[13] David Eisenbud, Bernd Sturmfels, Binomial ideals, Duke Mathematical Journal 84 (1) (1996) 1–45.
[14] Dan Geiger, Christopher Meek, Bernd Sturmfels, On the toric algebra of graphical models, The Annals of Statistics 34 (5) (2006) 1463–1492.
[15] Raymond Hemmecke, On the positive sum property and the computation of Graver test sets, Mathematical Programming 96 (2003) 247–269., 10.1007/s10107-003-0385-7.
[16] Raymond Hemmecke, Peter Malkin, Computing generating sets of lattice ideals, Journal of Symbolic Computation 44 (2009) 1463–1476.
[17] Thomas Kahle, Neighborliness of marginal polytopes, Contributions to Algebra and Geometry 51 (1) (2010) 45–56.
[18] Anargyros Katsabekis, Apostolos Thoma, Toric sets and orbits on toric varieties, Journal of Pure and Applied Algebra 181 (1) (2003) 75–83.
[19] Anargyros Katsabekis, Apostolos Thoma, Parametrizations of toric varieties over any field, Journal of Algebra 308 (2) (2007) 751–763.
[20] Thomas Kahle, Walter Wenzel, Nihat Ay, Hierarchical models, marginal polytopes, and linear codes, Kybernetika 45 (2) (2009) 189–208.
[21] Jesús A. De Loera, Shmuel Onn, Markov bases of three-way tables are arbitrarily complicated, Journal of Symbolic Computation 41 (2006) 173–181.
[22] František Matúš, Nihat Ay, On maximization of the information divergence from an exponential family, in: Proceedings of the WUPES'03, University of Economics, Prague, 2003, pp. 199–204.
[23] Peter Malkin, Computing Markov bases, Gröbner bases, and extreme rays, Ph.D. Thesis, Université de Louvain, 2007.
[24] Peter McMullen, The maximum number of faces of a convex polytope, Mathematika (1970) 179–184.

[25] Jörg Rambau, TOPCOM: Triangulations Of Point Configurations and Oriented Matroids, Mathematical Software—ICMS 2002, Arjeh M. Cohen, Xiao-Shan Gao, Nobuki Takayama (Eds.), World Scientific, 2002, pp. 330–340.
[26] Fabio Rapallo, Toric statistical models: parametric and binomial representations, Annals of the Institute of Statistical Mathematics 4 (59) (2007) 727–740.
[27] Johannes Rauh, Finding the maximizers of the information divergence from an exponential family, IEEE Transactions on Information Theory, submitted for publication. Available from: arXiv:0912.4660.
[28] Bernd Sturmfels, Some applications of affine Gale diagrams to polytopes with few vertices, SIAM Journal of Discrete Mathematics 1 (1988) 121–133.
[29] Günter M. Ziegler, Lectures on polytopes, GTM, vol. 152, Springer Verlag, 1994