Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Support Sets in Exponential Families and Oriented Matroid Theory

by

Johannes Rauh, Thomas Kahle, and Nihat Ay

Preprint no.: 28 2009

Support Sets in Exponential Families and Oriented Matroid Theory

Johannes Rauh
Max Planck Institute
for Mathematics in the Sciences
rauh@mis.mpg.de

Thomas Kahle
Max Planck Institute
for Mathematics in the Sciences
kahle@mis.mpg.de

Nihat Ay
Max Planck Institute
for Mathematics in the Sciences
nay@mis.mpg.de

Abstract

We discuss how to obtain an implicit description of the closure of a discrete exponential family with a finite set of equations derived from 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 framework allows us to study the possible support sets of an exponential families with the help of oriented matroid theory. In particular, if two exponential families induce the same oriented matroid, then they have the same support sets.

1 Introduction

In this paper we study exponential families, which are well known statistical models with many nice properties. Let $\mathcal{E}$ be an exponential family on a finite set $\mathcal{X}$, and $\overline{\mathcal{E}}$ its closure. We want to describe the set

$$\mathcal{S} := \left\{\operatorname{supp}(P) \subseteq \mathcal{X} : P \in \overline{\mathcal{E}}\right\}. \tag{1}$$

of all possible support sets occurring in $\overline{\mathcal{E}}$.

The problem of determining the possible support sets in an exponential family is a classical problem in statistics. It amounts to describing the boundary of the most basic statistical models. This problem is related to characterizing the marginal polytope, which can be used, for example, to study the existence or non-existence of the MLE [EFRS06]. One can show that computing the support sets of any exponential family is of the same complexity class as NP hard combinatorial problems such as the problem of finding maximal cuts in graphs, since it is known that the class of marginal polytopes contains the so-called cut polytopes (see [KWA09]). This means that there is no corresponding fast algorithm, unless NP = co-NP [DL97]. Nevertheless, considering only certain subclasses of exponential families, the situation may simplify so that explicit statements about support sets become possible. For instance, one of the authors discusses support sets of small cardinality in hierarchical models, a particular kind of exponential families [Kah10]. In this paper we find a concise characterization of the support sets in general exponential families with the help of oriented matroids. We hope that this will allow for further theoretical results in this direction.

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 for discrete random variables inspired by so called Markov bases [GMS06]. It is described in Theorem 4. 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.

This paper is organized as follows. In Section 2 we develop a theory of implicit representations of exponential families which is analogue to and inspired by algebraic statistics [GMS06]. 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 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 $\mathcal{X} := \{1, \dots, m\}$ 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 $n = n^+ - n^-$ 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).

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 to 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 $(1,\dots,1)$ 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 $(1,\dots,1)$ to A without changing the exponential family.

Remark 2. The exponential family depends on A only through its row span $\mathcal{L}$. 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 $\mathcal{L}$.

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

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

In the context of hierarchical models, the convex support is also called *marginal polytope*.

We will see later that the faces of $\mathrm{cs}(\mathcal{E}_{q,A})$ are in a one-to-one correspondence with the different support sets occurring in $\overline{\mathcal{E}}_{q,A}$. Even more is true: The mapping A, restricted to $\overline{\mathcal{E}}_{q,A}$, defines a homeomorphism $\overline{\mathcal{E}}_{q,A} \cong \mathrm{cs}(\mathcal{E}_{q,A})$ which maps every probability measure $p \in \overline{\mathcal{E}}_{q,A}$ into the face corresponding to its support, see for example [BN78]. This homeomorphism is called the *moment map*. One can use the properties of the moment map to prove Theorem 15 using arguments from the theory of oriented matroids. This will be discussed in the next section.

Note that the parametrization in (3) does not extend to the boundary. This is one of the motivations to move on to an implicit description of the exponential family. The next theorem shows how to obtain an implicit description from $\mathcal{E}_{q,A}$ from the kernel of A. This gives a nice "duality" as the parametrization itself is derived from the image of A.

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

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

hold for p.

Remark 5. This theorem is a direct generalization of Theorem 3.2 in [GMS06]. 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 was considered. However, the proof of the theorem generalizes without any major problem. Actually, the proof of our theorem here needs one step less, since we don't need to show the reduction to the polynomial equations. The different flavor of the results will be made more precise in Remark 13 later.

Our proof closely follows [GMS06]. In our presentation of the proof we want to explicitly point out how matroid-type arguments are used, the first example being Lemma 7.

Before giving the proof of Theorem 4 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 face of a polytope P 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 a compact set defined by finitely many inequalities (i.e. facets), see [Zie94].

In particular we are interested in the face structure of $\mathrm{cs}(\mathcal{E}_{q,A})$. 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). This motivates the following

Definition 6. Let $\{a_x : x \in \mathcal{X}\}$ be the vertex set of a polytope. A set $F \subseteq \mathcal{X}$ is called *facial* if there exists a vector $c \in \mathbb{R}^d$ such that

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

Lemma 7. *Fix a matrix $A = (a_x)_{x\in\mathcal{X}} \in \mathbb{R}^{d\times m}$ and a nonempty subset $F \subseteq \mathcal{X}$. Then we have:*

- *If F is facial then no non-zero non-negative linear combination of the a_x, $x \notin F$, can be written as linear combination of the a_x, $x \in F$.*
- *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. For the first statement, assume to the contrary that we can find $\alpha(x) \geq 0$ and $\beta(x)$ not all zero such that $u = \sum_{x\notin F} \alpha(x) a_x = \sum_{x\in F} \beta(x) a_x$, and let c be as in (6). We have

$$0 \leq \sum_{i\notin F} \alpha_i \leq \sum_{i\notin F} \alpha_i c^T a_i = c^T \left(\sum_{i\notin F} \alpha_i a_i \right) = c^T \left(\sum_{i\in F} \beta_i a_i \right) = 0,$$

whence $\alpha_i = 0$ for all $i \notin F$. This also proves the first direction of the second statement.

The opposite direction is a bit more complicated and uses Farkas' Lemma (see for example [Zie94]): *Let $B \in \mathbb{R}^{l\times d}$, and $z \in \mathbb{R}^l$. Either there exists a point in the polyhedron $\{x : Bx \le z\}$, or there exists a non-negative vector $y \in \mathbb{R}^l_{\ge}$ with $y^T B = 0$ and $y^T z < 0$, but not both.* Assume that $F \subsetneq \mathcal{X}$ is nonempty and satisfies (7) for all $u \in \ker A$. Let B be the $(|F| + m) \times d$ matrix with rows $\left\{a_x^T : x \in F\right\}$, $\left\{-a_x^T : x \in F\right\}$, $\left\{-a_x^T, x \notin F\right\}$, and z be the vector which has entries zero in the first $2\,|F|$ components and entries -1 in the last $m - |F|$. Then a solution to $Bx \le z$ provides a facial vector. Thus it remains to show that each non-negative $y = (y^{(1)}, y^{(2)}, y^{(3)})^T$, decomposed according to the rows of B, with $y^T B = 0$ satisfies $y^T z \ge 0$. Assume that the columns of A are ordered such that the columns with indices $x \in F$ come first. Then $y^{(3)}$ must be zero as otherwise $(y^{(2)} - y^{(1)}, y^{(3)})^T \in \ker A$ would violate (7) by non-negativity of y. But then $y^T z = 0$ trivially.

The last statement follows immediately from the second statement. □

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 equations (5) since they are homogeneous.

Denote Z_A the set of solutions of (5). We first show that $\mathcal{E}_A$ satisfies the equations defining Z_A. We plug in the parametrization to find

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

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

Next, let $p \in Z_A \setminus \mathcal{E}_A$. We construct a sequence p_μ in $\mathcal{E}_A$ that converges to p as $\mu \to -\infty$.

Consider the following system of equations in variables $d = (d_1, \dots, d_n)$:

$$d^T a_x = \log p(x) \qquad \text{for all } x \in \operatorname{supp}(p). \tag{9}$$

We claim that this linear system has a solution. Otherwise we can find numbers $v(x)$, $x \in F$, such that $\sum_x v(x) \log p(x) \ne 0$ and $\sum_x v(x) a_x = 0$. This leads to the contradiction $p^{v^+} \ne p^{v^-}$.

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

$$p_{(\mu)} := p_{\mu c+d} = \left(\mathrm{e}^{\mu c^T a_1}\mathrm{e}^{d^T a_1}, \dots, \mathrm{e}^{\mu c^T a_m}\mathrm{e}^{d^T a_m}\right) \in \mathcal{E}_A.$$

By (6) it is clear 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 [GMS06, Lemma A.2]:

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

1. *F is facial.*
2. *The uniform distribution $\frac{1}{|F|}\mathbb{1}_{\{F\}}$ of F lies in $\overline{\mathcal{E}}_A$.*
3. *There is a vector with support F in $\overline{\mathcal{E}}_A$.*

According to Theorem 4, in order test whether p is an element of the closure of $\mathcal{E}_{q,A}$, we have to test all the equations (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: A *circuit vector* of a matrix A is a nonzero vector $n \in \mathbb{R}^m$ corresponding to a linear dependency $\sum_x n(x)a_x$ with inclusion minimal support, i.e if $n' \in \mathbb{R}^m$ satisfies $\operatorname{supp}(n') \subseteq \operatorname{supp}(n)$, then $n' = \lambda n$ for some $\lambda \in \mathbb{R}$. Equivalently, n is an element of $\ker A$ with inclusion minimal support.

A *circuit* is the support set of a circuit vector. The minimality condition implies that the circuit determines its corresponding circuit vectors up to a multiple. A *circuit basis* C contains one circuit vector for every circuit.[1]

If we replace n by a nonzero multiple of n then equation (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 9. *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.

The proof is based on the following two lemmas:

Lemma 10. *For every vector $n \in \ker A$ there exists a sign-consistent circuit vector $c \in \ker A$, i.e. if $c(x) \neq 0$ 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 which is sign-consistent with n and satisfies $\operatorname{supp}(c) \subseteq \operatorname{supp}(n)$. If c is not a circuit, then there exists a circuit c' with $\operatorname{supp}(c') \subseteq \operatorname{supp}(c)$. Using a suitable linear combination $c + \alpha c'$, $\alpha \in \mathbb{R}$, we can obtain a contradiction to the minimality of c. □

Lemma 11. *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, as in the last lemma, to reduce the support. □

Proof of Theorem 9. 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

[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.

circuits 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

$$p^{n^+} - p^{n^-} = p^{\sum_{i=1}^{r-1} c_i^+} \left(p^{c_1^+} - p^{c_1^-} \right) + \left(p^{\sum_{i=1}^{r-1} c_i^+} - p^{\sum_{i=1}^{r-1} c_i^-} \right) p^{c_1^-}, \tag{11}$$

so the theorem follows easily by induction. □

The theorem implies that a finite number of equations is sufficient to describe $\overline{\mathcal{E}}_{q,A}$. The number of equations that are necessary is bounded from above by the number of different support sets occurring in C.

Example 12. 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

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

These two generators correspond to the two relations

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

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 equations (14): Possible solutions to these equations are given by

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

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:

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

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

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

Remark 13 (Relation to algebraic statistics). In the particular case where the vector space $\ker A$ has a basis with integer components (for example, if A itself has only integer entries), every circuit is proportional to a circuit with integer components. In this case the corresponding equations (5) are polynomial, and the theorem implies that $\overline{\mathcal{E}}_A$ is the non-negative real part of a *projective variety*, i.e. the solution set of homogeneous polynomials. If we want to use the tools of commutative algebra and algebraic geometry, then it turns out that circuits are not the right object to consider: For example, proportional circuits only yield equivalent equations if we consider them over the non-negative reals, but we may obtain a different solution set if we allow negative real solutions or complex solutions, which may greatly increase the running time of many algorithms of computational commutative algebra. Hence, if we want to use algebraic tools, it is best to work with a *Markov basis*, which can be defined as a finite set of kernel vectors such that the solution set over $\mathbb{C}$ of the corresponding equations equals the *Zariski closure* of $\overline{\mathcal{E}}$, i.e. the smallest variety containing $\overline{\mathcal{E}}$.[2] In this algebraic setting, Theorem 4 remains valid if we replace "closure" by "Zariski closure" and $\ker A$ by the integer kernel $\ker_{\mathbb{Z}} A$. This fact was first noted in [DS98].

In the algebraic case one can also look at the *ideal* (see [CLO08]) generated by all polynomial equations induced by integer valued circuit vectors. This ideal is called the *circuit ideal*. By what was said above this ideal is in general smaller than the associated *toric ideal*, which contains the polynomial equations induced by all integer valued kernel vectors. Circuit ideals have been studied already in the seminal paper [ES96]. For further results illuminating their nice relations to polyhedral geometry we refer to [BJT07].

Finding a Markov basis is in general a non-trivial task, see [HM09]. It seems to be much easier to compute the circuits of a matrix. However, a minimal Markov basis is usually much smaller than a circuit basis, and thus it is easier to handle (but cf. the next remark). For experiments in this direction we recommend the open source software package `4ti2` [4ti2] which can compute circuits as well as Markov bases.

Remark 14. Using arguments from matroid theory the number of circuits can be shown to be less or equal than $\binom{m}{r+2}$, where $m = |\mathcal{X}|$ is the size of the state space and r is the dimension of the exponential family $\mathcal{E}_{q,A}$, see [DSL04]. This gives us an upper bound on the number of implicit equations which is necessary to describe $\overline{\mathcal{E}}_{q,A}$. 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 to this, 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 column span of A. This amounts to checking whether $\log(p/q)$ is orthogonal to $\ker A$, which is equivalent to $m-r-1$ equations, once we have chosen 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 12, the equations 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. Of course, 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 circuits.

[2] It turns out that it is not so easy to find an example of a Markov basis which does not consists of circuits. In [AT03], S. Aoki and A. Takemura give a model and a Markov basis element which is not a circuit. Interestingly, the full Markov basis of this model is not known.

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

$$\mathcal{S}(q, A) := \{\operatorname{supp}(p) : p \in \overline{\mathcal{E}}_{q,A}\} \subseteq 2^{\mathcal{X}}. \tag{18}$$

Proposition 8 gives 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. In order to formalize this observation, consider the map

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

which associates to each vector a pair of disjoint subsets of $\mathcal{X}$. Such a pair of disjoint subsets shall be called a *signed subset* of $\mathcal{X}$ in the following. Alternatively, signed subsets (A, B) can also be represented as sign vectors $X \in \{-1, 0, +1\}^{\mathcal{X}}$, where

$$X(x) = \begin{cases} +1, & \text{if } x \in A, \\ -1, & \text{if } x \in B, \\ 0, & \text{else.} \end{cases} \tag{19}$$

In this representation, sgn corresponds to the usual signum mapping extended to vectors. As a slight abuse of notation, we don't make a difference between these two representations in the following.

The signed subset $\operatorname{sgn}(c)$ corresponding to a circuit $c \in \ker A$ shall be called an *oriented circuit*. The set of all oriented circuits is denoted by

$$\mathcal{C}(A) := \pm \operatorname{sgn}(C) = \{\operatorname{sgn}(c) : c \in C \text{ or } c \in -C\}, \tag{20}$$

where C is a circuit basis of A.

We immediately have the following

Theorem 15. *Let S be a nonempty subset of $\mathcal{X}$. Then $S \in \mathcal{S}$ if and only if the following holds for all signed circuits $(A, B) \in \mathcal{C}(A)$:*

$$A \subseteq S \quad \Leftrightarrow \quad B \subseteq S. \tag{21}$$

Corollary 16. *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 14, Theorem 15 gives us 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 (21) 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 17. Let's continue Example 12. From the circuits we deduce the following implications:

$$p(3) \neq 0 \iff p(4) \neq 0, \tag{22a}$$

$$p(1) \neq 0 \text{ and } p(2) \neq 0 \iff p(4) \neq 0, \tag{22b}$$

$$p(1) \neq 0 \text{ and } p(2) \neq 0 \iff p(3) \neq 0. \tag{22c}$$

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

$$p(1) \neq 0 \text{ and } p(2) \neq 0 \iff p(3) \neq 0 \text{ and } p(4) \neq 0. \tag{23}$$

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 18 (Supports in the binary no-n-way interaction model). Consider the binary hierarchical model [KWA09] 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 [LO06], so we cannot hope to find an easy description of the oriented 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":

$$e_{[n]}(x) := \begin{cases} -1 & \text{if } \sum_{i=1}^n x_i \text{ is odd,} \\ 1 & \text{otherwise.} \end{cases} \tag{24}$$

Using Theorem 15 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 $\lfloor \frac{\dim(P^{(n-1)})}{2} \rfloor = 2^{n-1} - 1$ 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 of k, i.e. the number of faces with k vertices. It is given by:

$$s_k = \binom{2^n}{k} - 2\binom{2^{n-1}}{k - 2^{n-1}}, \tag{25}$$

where $\binom{m}{l} = 0$ if $l < 0$. Since this polytope has only one affine dependency (24) 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

$$\boldsymbol{x} : \mathbb{R} \to \mathbb{R}^d, \qquad t \mapsto \boldsymbol{x}(t) := \left(t, t^2, \cdots, t^d\right)^T. \tag{26}$$

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

$$C(d,n) := \operatorname{conv}\left\{\boldsymbol{x}(t_1), \ldots, \boldsymbol{x}(t_n)\right\}, \tag{27}$$

the convex hull of $n > d$ distinct points ($t_1 < t_2 < \ldots < t_n$) on the moment curve. The face lattice of a cyclic polytope can easily be described using *Gale's evenness condition*, see [Zie94]. 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 19 (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 $C(d,n)$:*

$$f_k(P) \leq f_k(C(d,n)), \quad k = 0, \ldots, d. \tag{28}$$

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

Theorem 19 was conjectured by Motzkin in 1957 and its proof has a long and complicated history. The final result is due to McMullen [McM70].

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 also gives an interesting exponential family, answering the question for the exponential family of smallest dimension containing all the vertices of the probability simplex. The construction is due to [MA04].

Example 20. 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 $(1, \ldots, 1)$:

$$A := \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & 2 & 3 & \ldots & m \\ 1 & 4 & 9 & \ldots & m^2 \end{pmatrix}. \tag{29}$$

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

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

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 to 3 of [BVS+93] for a more detailed introduction to oriented matroids.

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) $\mathcal{C} = -\mathcal{C}$,

(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

$$C = \{X^+ \cup X^- = \operatorname{supp}(X) : X \in \mathcal{C}\} \tag{30}$$

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 $X \in C$.

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

$$\mathcal{C} = \big\{(\operatorname{supp}(n^+), \operatorname{supp}(n^-) : n \in \ker A \text{ has inclusion minimal support.}\big\}. \tag{31}$$

This example is so important that oriented matroids which arise in this way are given a name: 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 $\mathcal{V} = \{(\operatorname{supp}(n^+), \operatorname{supp}(n^-) : n \in \ker A\}$, then it is easy to see that $\mathcal{V}$ satisfies **(C3)**. Thus $(E, \mathcal{C})$ satisfies **(C3)** by the following proposition:

Proposition 21. *Let $\mathcal{V}$ be a nonempty collection of signed subsets of E satisfying* **(C1)** *and* **(C3)**. *Write $\operatorname{Min}(\mathcal{V})$ for the minimal elements of $\mathcal{V}$ (with respect to inclusion of supports). Then*

1. *for any $X \in \mathcal{V}$ there is $Y \in \operatorname{Min}(\mathcal{V})$ such that $Y^+ \subseteq X^+$ and $Y^- \subseteq X^-$.*

[3] Note that this definition depends, in fact, only on the kernel of A, compare Remark 2

2. $\mathrm{Min}(\mathcal{V})$ *is the set of circuits of an oriented matroid.*

Proof. [BVS+93], 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:

For two signed subsets X, Y of E define the *composition* of X and Y as

$$(X \circ Y)^+ := X^+ \cup (Y^+ \setminus X^-), \qquad (X \circ Y)^- := X^- \cup (Y^- \setminus X^+). \tag{32}$$

Note that this operation is associative but not commutative in general. 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.

The above proposition implies easily that an oriented matroid can 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:

$$\mathrm{sgn}(n + \epsilon n') = \mathrm{sgn}(n) \circ \mathrm{sgn}(n'), \quad \text{for } \epsilon > 0 \text{ small enough.} \tag{33}$$

Now Lemmas 10 and 11 correspond to the following two lemmas

Lemma 10'. *For every o.m. vector Y there exists a sign-consistent signed circuit X such that $\underline{X} \subseteq \underline{Y}$.*

Lemma 11'. *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

[4] In [BVS+93], o.m. vectors are simply called vectors. The name "o.m. vector" has been proposed by F. Matúš to avoid confusion.

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 $\operatorname{sgn}^*(l) := (N_l^+, N_l^-)$ with l. The signed subset $\operatorname{sgn}^*(l)$ 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 $\operatorname{sgn}^*(l)$ is called a *signed cocircuit*. The collection of all signed cocircuits shall be denoted by $\mathcal{C}^*$.

Lemma 22. *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 [BVS+93]. □

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 $T := (\mathcal{X}, \emptyset)$.

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 7:

Proposition 23 (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 [BVS+93] 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 15: 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 16 can be rewritten as

Corollary 16'. *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 [Zie94]. 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 16' implies that the instruments of the theory of oriented matroids should suffice to describe the support sets of an exponential family.

Remark 24 (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* [Zie94, 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 [Stu88] 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 $|E|$.

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 supported by the Volkswagen Foundation.

References

[4ti2] 4ti2 team, *4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces*, available at www.4ti2.de.

[AT03] Satoshi Aoki and Akimichi Takemura, *The list of indispensable moves of the unique minimal markov basis for 3x4xk and 4x4x4 contingency tables with fixed two-dimensional marginals*, METR 2003-38, University of Tokyo, Japan, 2003.

[BJT07] Tristram Bogarta, Anders N. Jensen, and Rekha R. Thomas, *The circuit ideal of a vector configuration*, Journal of Algebra **309** (2007), no. 2, 518–542.

[BN78] Ole E. Barndorff-Nielsen, *Information and exponential families in statistical theory*, Wiley, 1978.

[BVS+93] A. Björner, M. Las Vergnas, B. Sturmfels, N. White, and G. Ziegler, *Oriented matroids*, Encyclopedia of Mathematics and its Applications, Cambridge, 1993.

[CLO08] David A. Cox, John B. Little, and Don O'Shea, *Ideals, varieties, and algorithms: An introduction to computational algebraic geometry and commutative algebra*, Springer, 2008.

[DL97] Michel M. Deza and Monique Laurent, *Geometry of cuts and metrics*, Algorithms and Combinatorics, Springer, 1997.

[DS98] Persi Diaconis and Bernd Sturmfels, *Algebraic algorithms for sampling from conditional distributions*, Annals of Statistics **26** (1998), 363–397.

[DSL04] György Dósa, István Szalkai, and Claude Laflamme, *The maximum and minimum number of circuits and bases of matroids*, Pure Mathematics and Applications **15** (2004), no. 4, 383–392.

[EFRS06] Nicholas Eriksson, Stephen E. Fienberg, Alessandro Rinaldo, and Seth Sullivant, *Polyhedral conditions for the nonexistence of the mle for hierarchical log-linear models.*, Journal of Symbolic Computation (2006), no. 41, 222–233, Special issue on Computational Algebraic Statistics.

[ES96] David Eisenbud and Bernd Sturmfels, *Binomial ideals*, Duke Mathematical Journal **84** (1996), no. 1, 1–45.

[GMS06] Dan Geiger, Christopher Meek, and Bernd Sturmfels, *On the toric algebra of graphical models*, The Annals of Statistics **34** (2006), no. 5, 1463–1492.

[HM09] Raymond Hemmecke and Peter Malkin, *Computing generating sets of lattice ideals*, Journal of Symbolic Computation (2009), accepted.

[Kah10] Thomas Kahle, *Neighborliness of marginal polytopes*, Contributions to Algebra and Geometry (2010), accepted, arXiv:0809.0786.

[KWA09] Thomas Kahle, Walter Wenzel, and Nihat Ay, *Hierarchical models, marginal polytopes, and linear codes*, Kybernetika **45** (2009), no. 2, 189–208.

[LO06] Jesús A. De Loera and Shmuel Onn, *Markov bases of three-way tables are arbitrarily complicated*, Journal of Symbolic Computation **41** (2006), 173–181.

[MA04] František Matúš and Nihat Ay, *On maximization of the information divergence from an exponential family*, Kybernetika **41** (2004), 731–746.

[McM70] Peter McMullen, *The maximum number of faces of a convex polytope*, Mathematika (1970), 179–184.

[Stu88] Bernd Sturmfels, *Some applications of affine gale diagrams to polytopes with few vertices*, SIAM J. Discrete Mathematics **1** (1988), 121–133.

[Zie94] Günter M. Ziegler, *Lectures on polytopes*, GTM, vol. 152, Springer Verlag, 1994.