KYBERNETIKA — VOLUME *48* (2012), NUMBER 5, PAGES 845–864

ON SOLUTION SETS OF INFORMATION INEQUALITIES

Nihat Ay and Walter Wenzel

We investigate solution sets of a special kind of linear inequality systems. In particular, we derive characterizations of these sets in terms of minimal solution sets. The studied inequalities emerge as information inequalities in the context of Bayesian networks. This allows to deduce structural properties of Bayesian networks, which is important within causal inference.

Keywords: linear inequalities, polyhedral sets, Bayesian networks, information, entropy

Classification: 94A17, 15A39, 52Bxx

1. INTRODUCTION

This paper studies solution sets of linear inequalities

$$c_i \;\leq\; \sum_{j=1}^{m} \alpha_{ij} \cdot f_j, \qquad 1 \leq i \leq n\,, \tag{1}$$

where the numbers $c_1, \dots, c_n$ and α_{ij} for $1 \leq i \leq n$, $1 \leq j \leq m$, satisfy the following conditions:

(I) $c_i > 0$ for $1 \leq i \leq n$,

(II) $\alpha_{ij} \geq 0$ for $1 \leq i \leq n$, $1 \leq j \leq m$,

(III) for all i with $1 \leq i \leq n$ there exists j with $1 \leq j \leq m$ and $\alpha_{ij} > 0$,

(IV) for all j with $1 \leq j \leq m$ there exists i with $1 \leq i \leq n$ and $\alpha_{ij} > 0$.

The examination of solution sets of arbitrary finite systems of linear inequalities, which are always polyhedral sets, is well established, see for instance [8], Section 3.2, and [9], Chapter 1. However, for our special class of linear inequalities (1), given in terms of the conditions (I) – (IV), we can derive results on the characterization of solution sets that do not hold in general for arbitrary polyhedral sets. We particularly study minimal solutions with respect to the product order as well as certain projections from the full solution set into the set of minimal solutions with a variety of instructive properties (see, for instance, Theorem 2.1, Corollary 2.2, and Theorem 2.4).

The motivation of our special inequality systems comes from the study of Bayesian networks as formalism for a causality theory that has been proposed by Pearl [5]. In order to be more precise, in Section 1.1 we present two inequalities derived in the work of one of the authors [1, 6]. Although these examples serve as motivation of the present work, the direct applications of this paper to causal inference are not explored here and are subject of future research.

1.1. Information-theoretic inequalities

The two examples below refer to distributions that are factorizable with respect to a *directed acyclic graph* $G = (V, E)$, $E \subseteq V \times V$, where V is a finite set. To simplify notation, in this section we put $V = \{1, \ldots, N\}$. The acyclicity property of G simply means that there are no directed cycles in the graph, see as illustration Figure 1. With

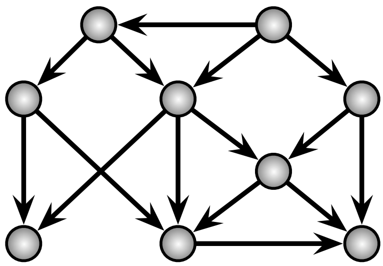

Fig. 1.

each node v we associate a random variable X_v and assume that the joint distribution of these variables satisfies

$$p(x_1, \ldots, x_N) \;=\; \prod_{v=1}^{N} p(x_v | x_{\mathsf{pa}(v)}) \,. \tag{2}$$

Here, $\mathsf{pa}(v)$ denotes the set of *parents* of node v. The graph, together with the conditional distributions $p(x_v | x_{\mathsf{pa}(v)})$ is called a *Bayesian network*. The required technical definitions related to Bayesian networks are given in the appendix.

Given a Bayesian network $\mathscr{B}$ and a subsystem $S \subseteq V$, we denote the joint distribution of the variables X_v, $v \in S$, by $p_S(\mathscr{B})$ (*marginal distribution*). To simplify notation, in this section we set $S := \{1, \ldots, n\}$, where $n \leq N$ is fixed. In [1, 6], general inequalities of the following type have been derived, which hold for any Bayesian network $\mathscr{B}$:

$$\sum_j \alpha_{ij} \cdot f_j(\mathscr{B}) \;\geq\; c_i\big(p_S(\mathscr{B})\big) \,. \tag{3}$$

Here, the f_j on the left hand side as well as the right hand side depend on the underlying Bayesian network $\mathscr{B}$. However, what makes the inequalities (3) special is the fact that the dependence of the right hand side is only through $p_S(\mathscr{B})$. This can be used for the inference of particular aspects of the underlying Bayesian network $\mathscr{B}$: Assume that the marginal distribution $p_S(\mathscr{B})$ is available to an observer who only observes the variables X_v, $v \in S$. Then for any Bayesian network $\mathscr{B}$ that is consistent with this observation, the right hand side of (3) is constant, and the values $f_j(\mathscr{B})$ have to satisfy the resulting linear inequalities which are of the form (1). Those Bayesian networks $\mathscr{B}$ for which the values $f_j(\mathscr{B})$ do not satisfy these constraints are not possible as underlying Bayesian networks. This kind of exclusion is of particular interest if it allows to deduce structural properties of the underlying network.

In the examples below, both the f_j's and the c_i's are given in terms of information-theoretic quantities. In this context, particularly important building blocks of these quantities are the entropy and the mutual information. Given two random variables X and Y and corresponding distributions $p(x)$, $p(y)$, and $p(x,y)$, they are defined as follows:

$$H(X) \;=\; -\sum_{x} p(x)\ln p(x) \qquad \text{(entropy)}\;,$$

$$I(X:Y) \;=\; \sum_{x,y} p(x,y)\ln\left(\frac{p(x,y)}{p(x)\,p(y)}\right) \qquad \text{(mutual information)}\;.$$

1.1.1. Local information flows

We consider the information inequalities (see [1], Theorem 3)

$$\sum_{v\in A} I(X_v : X_{\mathsf{pa}(v)}) \;\ge\; \sum_{v\in A} H(X_v) - H(X_A), \qquad A\subseteq S\;. \tag{4}$$

Here, each mutual information term $I_v := I(X_v : X_{\mathsf{pa}(v)})$ measures the local information flow into the node v. Therefore, the sum on the left hand side quantifies the total information flow into the observed subsystem S. Obviously, these inequalities have the form (3). That is, the right hand side, known as *multi-information*, only depends on the marginal p_S, whereas each term of the left hand side also depends on further information contained in $\mathscr{B}$. Note that in the case $|A| = 2$, the multi-information reduces to the mutual information. We use the abbreviation c_A for the right hand side of (4) and consider those inequalities for which $c_A > 0$ holds:

$$\sum_{v\in A} I_v \;\ge\; c_A, \qquad A\subseteq S\;,\;\; c_A > 0\;. \tag{5}$$

We now want to address the following question: What is the maximal number of vanishing I_v's? To put this question in more formal terms, we define

$$\mathfrak{M}(\mathscr{B}) \;:=\; \{A\subseteq S \;:\; I_v = 0 \text{ for all } v\in A\}$$

and have to determine

$$\nu \;:=\; \sup_{\mathscr{B}} \max_{A\,\in\,\mathfrak{M}(\mathscr{B})} |A|\;.$$

Here, the supremum is taken with respect to all Bayesian networks $\mathscr{B}$ that are consistent with the observed distribution, that is $p_S(\mathscr{B}) = p_S$. Note that we allow for a variation of the number N, $N \ge n$, of nodes in such a network $\mathscr{B}$. In order to determine ν, consider the set

$$\mathfrak{N} \;:=\; \{A\subseteq S \;:\; c_A = 0\}\;.$$

Obviously, if a set $A \subseteq S$ satisfies $I_v = 0$ for all $v \in A$, that is $A \in \mathfrak{M}(\mathscr{B})$, then $A \in \mathfrak{N}$. This implies

$$\nu \;\leq\; \max_{A \,\in\, \mathfrak{N}} \, |A| \,.$$

It is easy to see that even equality holds by finding a Bayesian network $\mathscr{B}$ for which

$$\max_{A \,\in\, \mathfrak{M}(\mathscr{B})} \, |A| \;\geq\; \max_{A \,\in\, \mathfrak{N}} \, |A| \tag{6}$$

holds. We define the Bayesian network as follows: As node set V we choose the observed subset $S = \{1, \ldots, n\}$, that is $N = n$, and select a set $A \in \mathfrak{N}$ with maximal cardinality which we denote by m. Without loss of generality we assume $A = \{1, \ldots, m\} \subseteq S$. From the definition of $\mathfrak{N}$ it follows that the A-marginal of p factorizes. This implies

$$\begin{aligned} p(x_1, \ldots, x_n) \quad &= \quad p(x_1, \ldots, x_m) \prod_{i=m+1}^{n} p(x_i \,|\, x_1, \ldots, x_{i-1}) \\ &= \quad p(x_1)\, p(x_2) \cdots p(x_m) \prod_{i=m+1}^{n} p(x_i \,|\, x_1, \ldots, x_{i-1}). \end{aligned}$$

This product structure suggests to choose the edge set

$$\{(i,j) \in S \times S \;:\; i < j, \;\; j \geq m+1\}$$

between the nodes of S. This implies $\mathsf{pa}(i) = \emptyset$ for $1 \leq i \leq m$, and $\mathsf{pa}(i) = \{1, \ldots, i-1\}$ for $m+1 \leq i \leq n = N$, and the equality (2) holds. Clearly, for this Bayesian network we have $I_i = 0$ for all $1 \leq i \leq m$, that is $A \in \mathfrak{M}(\mathscr{B})$. This implies inequality (6).

From our considerations it immediately follows that the minimal number ν^* of positive information flows is given by $|S| - \max_{A \,\in\, \mathfrak{N}} \, |A|$: with

$$\overline{\mathfrak{M}}(\mathscr{B}) \;:=\; \{A \subseteq S \;:\; I_v > 0 \text{ for all } v \in A\}$$

one has

$$\begin{aligned} \nu^* \quad &= \quad \inf_{\mathscr{B}} \; \min_{A \in \overline{\mathfrak{M}}(\mathscr{B})} \, |A| \\ &= \quad \inf_{\mathscr{B}} \; \min_{A \in \mathfrak{M}(\mathscr{B})} \, (|S| - |A|) \\ &= \quad |S| - \sup_{\mathscr{B}} \; \max_{A \,\in\, \mathfrak{M}(\mathscr{B})} \, |A| \\ &= \quad |S| - \max_{A \,\in\, \mathfrak{N}} \, |A| \\ &= \quad |S| - \nu \,. \end{aligned}$$

These results can be compared with the general results on solution sets of linear inequality systems given in Section 2 (see Example 2.18 (a)).

1.1.2. Entropy of common ancestors

In this example, we consider the partition generated by the *ancestral sets* $\mathsf{an}(v)$, $v \in S$:

$$\pi_A \;:=\; \left(\bigcap_{v\in A} \mathsf{an}(v)\right) \cap \left(\bigcap_{v\in V\setminus A} \mathsf{an}(v)\right), \qquad A \subseteq S\,.$$

Given A, π_A consists of the nodes $w \in V$ that satisfy $w \rightsquigarrow v$ for all $v \in A$ and $w \not\rightsquigarrow v$ for all $v \in S \setminus A$. Note that this set can be empty. As a convention, in that case the configuration set $\mathbb{X}_{\pi_A}$ consists of the empty configuration ϵ, and therefore $H(X_{\pi_A}) = 0$. This of course implies that $\pi_A \neq \emptyset$, if $H(X_{\pi_A}) > 0$. We define

$$\pi^{(g)} \;:=\; \{v \in V \;:\; v \rightsquigarrow a \text{ for at least } g \text{ nodes } a \text{ in } S\} \;=\; \biguplus_{\substack{A\subseteq S\\ |A|\geq g}} \pi_A\,. \tag{7}$$

In [6], the following inequality has been derived:

$$H(X_{\pi^{(g)}}) \;\geq\; \frac{1}{|S|-g+1}\left(\sum_{v\in S} H(X_v) - (g-1)\cdot H(X_S)\right), \qquad 2 \leq g \leq |S|\,. \tag{8}$$

On the left hand side of this inequality we have the entropy of the common ancestors of at least g observed nodes in S. The expression on the right hand side only depends on the marginal distribution on S and can be positive or negative. If it is positive, then this inequality already implies the existence of common ancestors of at least g nodes in any Bayesian network that is consistent with the observation. Thus, we have a structural implication on the underlying Bayesian network based on the observed marginal distribution.

We abbreviate the right hand side of the inequality (8) by c_g and use the decomposition (7) of $\pi^{(g)}$ in order to obtain inequality constraints for the entropies of the atoms π_A:

$$\sum_{\substack{A\subseteq S\\ |A|\geq g}} H(X_{\pi_A}) \;\geq\; c_g, \qquad 2 \leq g \leq |S|,\;\; c_g > 0\,. \tag{9}$$

In contrast to the first example of local information flows, here only one positive entropy term is already sufficient for satisfying these inequalities.

2. SOLUTIONS AND MINIMAL SOLUTIONS

After having motivated the general problem, we now return to the inequalities (1) and study the sets

$$\mathbb{L} \;:=\; \{(f_1,\ldots,f_m) \in \mathbb{R}^m \;:\; f_1,\ldots,f_m \geq 0, \text{ and (1) is satisfied}\}$$

and

$$\mathbb{L}_0 \;:=\; \mathbb{L}_{\min}\text{- with respect to the coordinatewise order “}\leq\text{” in } \mathbb{R}^m.$$

More precisely: $f = (f_1,\ldots,f_m) \in \mathbb{L}_0$, $g = (g_1,\ldots,g_m) \in \mathbb{L}$, and $g_i \leq f_i$ for all i always implies $g = f$.

The set $\mathbb{L}_0$ is interesting, because one knows all solutions in $\mathbb{L}$ as soon as one knows all solutions in $\mathbb{L}_0$.

It follows directly from the assumptions that

$$(T, \ldots, T) \in \mathbb{L}, \text{ if } T \in \mathbb{R}^+ \text{ is sufficiently large.} \tag{10}$$

Theorem 2.1. There is a mapping $p : \mathbb{L} \to \mathbb{L}_0$ that satisfies the following conditions:

(a) $p(f) \leq f$ for all $f = (f_1, \ldots, f_m) \in \mathbb{L}$.

(b) $p(f) = f$ if $f \in \mathbb{L}_0$.

(c) There exists an $L \in \mathbb{R}^+$ such that for all $f, g \in \mathbb{L}$:

$$\|p(f) - p(g)\|_{\sup} \leq L \cdot \|f - g\|_{\sup}.$$

Proof. For $1 \leq j \leq m$ define

$$P_j := \{i : 1 \leq i \leq n, \ \alpha_{ij} > 0\}.$$

For given $f \in \mathbb{L}$ and $1 \leq j \leq m$ we then define $p_j(f) = (f'_1, \ldots, f'_m) \in \mathbb{L}$ as follows:

$$f'_k := \begin{cases} f_k & \text{for } k \neq j \\ \max\left(\{0\} \cup \left\{ \dfrac{c_i}{\alpha_{ij}} - \displaystyle\sum_{\substack{\nu=1 \\ \nu \neq j}}^{m} \dfrac{\alpha_{i\nu}}{\alpha_{ij}} \cdot f_\nu \ : \ i \in P_j \right\}\right) & \text{for } k = j. \end{cases}$$

From these definitions it follows that

$$p_j(f) \leq f; \qquad p_j(f) = f \text{ if } f \in \mathbb{L}_0; \qquad p_j(f) \in \mathbb{L}.$$

Furthermore, for $f = (f_1, \ldots, f_m) \in \mathbb{L}$, $g = (g_1, \ldots, g_m) \in \mathbb{L}$, and with

$$L_j := \max_{i \in P_j} \left(\sum_{\substack{\nu=1 \\ \nu \neq j}}^{m} \left| \frac{\alpha_{i\nu}}{\alpha_{ij}} \right| \right)$$

we obtain

$$\begin{aligned} f'_j \ &\leq \ \max\left(\{0\} \cup \left\{ \frac{c_i}{\alpha_{ij}} - \sum_{\substack{\nu=1 \\ \nu \neq j}}^{m} \frac{\alpha_{i\nu}}{\alpha_{ij}} \cdot g_\nu \ : \ i \in P_j \right\}\right) \\ &\qquad + \max\left(\{0\} \cup \left\{ \sum_{\substack{\nu=1 \\ \nu \neq j}}^{m} \frac{\alpha_{i\nu}}{\alpha_{ij}} \cdot (g_\nu - f_\nu) \ : \ i \in P_j \right\}\right) \\ &\leq \ g'_j + L_j \cdot \|g - f\|_{\sup}. \end{aligned}$$

Analogously we have

$$g'_j \;\leq\; f'_j + L_j \cdot \|g - f\|_{\sup}.$$

This means the following:

$$\|p_j(g) - p_j(f)\|_{\sup} \;\leq\; (L_j + 1) \cdot \|g - f\|_{\sup}.$$

Now we define $p : \mathbb{L} \to \mathbb{L}_0$ as

$$p(f) \;:=\; (p_m \circ \cdots \circ p_1)(f).$$

Then the three properties stated in the theorem follow with

$$L \;:=\; \prod_{j=1}^{m} (L_j + 1).$$

□

Corollary 2.2. The mapping $p : \mathbb{L} \to \mathbb{L}_0$ in the above theorem satisfies the Lipschitz-condition and is therefore continuous. In particular, $\mathbb{L}_0 = p(\mathbb{L})$ is, as image of the convex set $\mathbb{L}$, connected.

Remark 2.3. We have the following chain of implications:

$$x_0 \text{ is an extreme point of } \mathbb{L} \;\;\Rightarrow\;\; x_0 \in \mathbb{L}_0 \;\;\Rightarrow\;\; x_0 \text{ is a boundary point of } \mathbb{L}.$$

We introduce the following conventions: Let $p : \mathbb{L} \to \mathbb{L}_0$ be as in Theorem 2.1. Furthermore, let S denote the set of extreme points of $\mathbb{L}$, which is non-empty and finite, because $\mathbb{L}$ is a closed polyhedron (see also Theorem 1.2 in [9]). Moreover, put

$$A \;:=\; \mathsf{conv}(S).$$

For $y_1, \ldots, y_k \in \mathbb{R}^m \setminus \{0\}$ put

$$\mathsf{cone}(\{y_1, \ldots, y_k\}) \;:=\; \left\{ \sum_{j=1}^{k} \lambda_j \cdot y_j \;:\; \lambda_1, \ldots, \lambda_k \geq 0 \right\}.$$

Finally, let $e_1, \ldots, e_m \in \mathbb{R}^m$ denote the canonical unit vectors, and put

$$C_0 \;:=\; \mathsf{cone}(\{e_1, \ldots, e_m\}).$$

Theorem 2.4. The following holds:

- **(a)** $\mathbb{L} = \mathbb{L}_0 + C_0$.
- **(b)** $\mathbb{L} = A + C_0$.
- **(c)** $\mathbb{L}_0 \subseteq A$.
- **(d)** $\mathbb{L}_0 = p(A) \subseteq A$, and $\mathbb{L}_0$ is compact.

P r o o f. **(a)** This clearly follows from the definition of $\mathbb{L}_0$ and the fact that $x \leq y$ and $x \in \mathbb{L}$ always implies $y \in \mathbb{L}$.

(b) The set $\mathbb{L}$ is non-empty and does not contain any line. Therefore, there are points $y_1, \ldots, y_k \in \mathbb{R}^m \setminus \{0\}$ satisfying

$$\mathbb{L} \; = \; A + \mathsf{cone}(\{y_1, \ldots, y_k\}).$$

(See for example [8], Theorem 4.1.3, or [9], Theorem 1.2.) From the fact that $\mathbb{L}$ contains only points with non-negative entries it follows immediately that all vectors $y_1, \ldots, y_k$ have only non-negative entries. Therefore, with $A \subseteq \mathbb{L}$ we also have

$$\mathbb{L} \; = \; A + \mathsf{cone}(\{y_1, \ldots, y_k\}) \; \subseteq \; A + C_0 \; \subseteq \; \mathbb{L}.$$

Therefore, we have $A + C_0 = \mathbb{L}$.

(c) Let $f \in \mathbb{L}_0$. Then, according to (b) there exist $x \in A$ and $y \in C_0$ with $f = x + y$. Then, $y \geq 0$, $x \in A \subseteq \mathbb{L}$, and $f \in \mathbb{L}_0$ yield:

$$y \; = \; 0 \quad \text{and therefore} \quad f \; = \; x \; \in \; A.$$

(d) According to (c) we have $\mathbb{L}_0 \subseteq A \subseteq \mathbb{L}$ and therefore

$$\mathbb{L}_0 \; = \; p(\mathbb{L}_0) \; \subseteq \; p(A) \; \subseteq \; p(\mathbb{L}) \; = \; \mathbb{L}_0.$$

This implies $p(A) = \mathbb{L}_0 \subseteq A$. With the compactness of A and the continuity of p we obtain the compactness of $\mathbb{L}_0 = p(A)$. □

Remark 2.5. Clearly, $\mathbb{L}$ is an m-dimensional subset of $\mathbb{R}^m$. In many examples, also the polytope A has dimension m; see for instance, Example 2.12. However, the polytope A can also have a smaller dimension and can even coincide with $\mathbb{L}_0$.

Example 2.6. For $m = 3$ we consider the following system of $n = 4$ linear inequalities for variables $x_1, x_2, x_3 \geq 0$:

$$x_1 + x_2 \; \geq \; 1, \quad x_1 + x_3 \; \geq \; 1, \quad x_2 + x_3 \; \geq \; 1, \quad x_1 + x_2 + x_3 \; \geq \; 2.$$

Here we have $S = \{v_1, v_2, v_3\}$ with

$$v_1 \; = \; (0,1,1), \quad v_2 \; = \; (1,0,1), \quad v_3 \; = \; (1,1,0).$$

Therefore we have

$$A \; = \; \mathsf{conv}(S) \; = \; \{(x,y,z) \in \mathbb{R}^3 \; : \; 0 \leq x,y,z \leq 1, \;\; x+y+z = 2\}.$$

The equality $A = \mathbb{L}_0$ immediately follows from the fact that each two distinct points in A are not comparable (with respect to the coordinatewise order).
Note that none of the four inequalities of the above system is redundant: consider the points

$$f_1 \; = \; (0,0,2), \quad f_2 \; = \; (0,2,0), \quad f_3 \; = \; (2,0,0), \quad f_4 \; = \; (\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}).$$

Each point f_i, $1 \leq i \leq 4$, satisfies all but the ith inequality.

In order to further study the structure of $\mathbb{L}_0$ we first show the following proposition.

Proposition 2.7. **(a)** Let $x, y \in \mathbb{L}$ with $x \neq y$, and let $\lambda, \nu > 0$ with $\lambda + \nu = 1$ and $z := \lambda \cdot x + \nu \cdot y \in \mathbb{L}_0$. Then we have: $x, y \in \mathbb{L}_0$.

(b) If K is a convex subset of $\mathbb{L}$ then $K \setminus \mathbb{L}_0$ is also convex.

(c) $\mathbb{L} \setminus \mathbb{L}_0$ and $A \setminus \mathbb{L}_0$ are convex sets.

Proof. **(a)** We prove this statement by contradiction. Assume $y \notin \mathbb{L}_0$. Then there exists $\delta > 0$ and i, $1 \leq i \leq m$, such that for the unit vector e_i we get:

$$y - \delta \cdot e_i \ \in \ \mathbb{L}.$$

This implies that also

$$z - \nu \cdot \delta \cdot e_i \ = \ \lambda \cdot x + \nu \cdot (y - \delta \cdot e_i) \ \in \ \mathbb{L}.$$

This contradicts the assumption $z \in \mathbb{L}_0$ because $\nu \cdot \delta > 0$.
Similarly we obtain $x \in \mathbb{L}_0$.

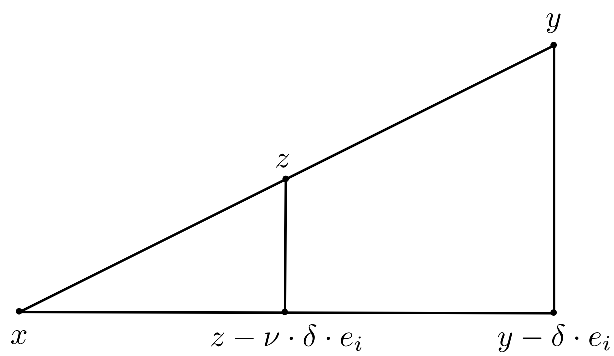

Fig. 2.

(b) If $x, y \in K \setminus \mathbb{L}_0$ then (a) implies that the line segment $\overline{xy} = \{(1-t)x + ty \ : \ 0 \leq t \leq 1\}$ does not intersect $\mathbb{L}_0$. The convexity of K implies that $\overline{xy} \subseteq K \setminus \mathbb{L}_0$.

(c) This follows from (b) by specialization. □

Corollary 2.8. For each line $g \subseteq \mathbb{R}^m$ that contains at least two points of $\mathbb{L}_0$ we have $g \cap \mathbb{L} \subseteq \mathbb{L}_0$.

In addition, the first part of Proposition 2.7 implies the following.

Theorem 2.9. $\mathbb{L}_0$ is the union of faces of A and also the union of faces of $\mathbb{L}$.

The following structural result implies an even stronger connection between the faces of A, the faces of $\mathbb{L}$ and the set $\mathbb{L}_0$.

Theorem 2.10. Let B be a non-empty face of A with $\dim B < m$. Then the following statements are equivalent.

(i) $B \subseteq \mathbb{L}_0$.

(ii) B is a face of $\mathbb{L}$.

(iii) B is contained in a supporting hyperplane H of $\mathbb{L}$ that has a normal vector, pointing into $\mathbb{L}$, with only positive coordinates.

Proof. **(i)** $\Rightarrow$ **(ii)**: Let $x, y \in \mathbb{L}$ with $x \neq y$ and let $\lambda, \nu > 0$ with $\lambda + \nu = 1$ and $\lambda \cdot x + \nu \cdot y \in B$. We have to show that $x, y \in B$.
From the first part of Proposition 2.7 and the assumption $B \subseteq \mathbb{L}_0$ we get $x, y \in \mathbb{L}_0 \subseteq A$. The fact that B is a face of A then implies $x, y \in B$.

(ii) $\Rightarrow$ **(iii)**: Let H be a supporting hyperplane of $\mathbb{L}$ with $\mathbb{L} \cap H = B$. It is sufficient to deduce a contradiction from the assumption that H has a normal vector $z = (z_1, \ldots, z_m)$ with $z_i > 0$ and $z_j \leq 0$ for some i, j.
The vector $x := z_i \cdot e_j + |z_j| \cdot e_i$ is perpendicular to z, and $x \geq 0$. Therefore, given an arbitrary $b \in B = \mathbb{L} \cap H$ we have

$$b + \lambda \cdot x \;\in\; \mathbb{L} \cap H = B \quad \text{for all } \lambda > 0.$$

However, this is not possible because $x \neq 0$ and B is bounded as a face of the polytope A.

(iii) $\Rightarrow$ **(i)**: Let $b \in B$. If $b \notin \mathbb{L}_0$ then there is an i and some $\lambda > 0$ with $b - \lambda \cdot e_i \in \mathbb{L}$. On the other hand, $b + \lambda \cdot e_i \in \mathbb{L}$, and therefore

$$\{b - \lambda \cdot e_i, b, b + \lambda \cdot e_i\} \;\subseteq\; H.$$

This would imply that H has a normal vector that is perpendicular to e_i . According to (iii) this is not possible. ☐

Remark 2.11. In [7], for the first time *visibility problems* have been studied; see also [3] and [4]. Given a convex subset K of $\mathbb{R}^m$, $p \in \mathbb{R}^m \setminus K$, and an element q of the boundary ∂K of K, we say that q is *visible* by p, if the line segment $\overline{pq}$ does not contain any relative interior point belonging to K, that is

$$\overline{pq} \cap K \;=\; \{q\}\,.$$

In the special case $K = \mathbb{L}$ the above theorems imply that each point $q \in \mathbb{L}_0$ is visible by the origin 0 because we have $\overline{0q} \cap \mathbb{L} = \{q\}$. This observation might be methodologically interesting and establishes connections between visibility problems and linear inequality systems. As the following example shows, not all points of $\mathbb{L}$ that are visible by 0 are in fact contained in $\mathbb{L}_0$ (see Figure 3):

$$x_1 \;\geq\; 1, \qquad x_2 \;\geq\; 1, \qquad x_1 + x_2 \;\geq\; 3\,.$$

In this example, all points of the unbounded set $\partial\mathbb{L}$ are visible by 0. On the other hand, with the two points $p = (2, 1)$ and $q = (1, 2)$ we have $\mathbb{L}_0 = \overline{pq}$.

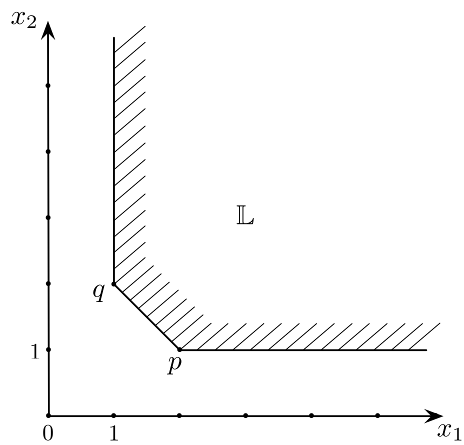

Fig. 3.

Finally, we study the following example.

Example 2.12. For $m = 3$, consider the following linear inequality system with variables $x_1, x_2, x_3 \geq 0$:

$$x_1 + 2\,x_2 + x_3 \;\geq\; 3, \quad x_1 + x_2 + 2\,x_3 \;\geq\; 3.$$

The corresponding set of extreme points is given by $S = \{v_1, v_2, v_3, v_4\}$ where

$$v_1 = (3,0,0), \quad v_2 = (0,3,0), \quad v_3 = (0,0,3), \quad v_4 = (0,1,1).$$

The set $A = \mathsf{conv}(S)$ is a 3-dimensional simplex. With

$$B_i \;:=\; \mathsf{conv}(S \setminus \{v_i\}), \qquad 1 \leq i \leq 4,$$

we have

$$\mathbb{L}_0 \;=\; B_2 \cup B_3.$$

B_2 and B_3 are those faces of A that are also faces of $\mathbb{L}$.

The face $B_1 = \mathsf{conv}(\{v_2, v_3, v_4\})$ is contained in the unbounded face $B_1 + \mathsf{cone}(\{v_2, v_3, v_4\})$ of $\mathbb{L}$. However, the face $B_4 = \mathsf{conv}(\{v_1, v_2, v_3\})$ is not contained in $\partial\mathbb{L}$ at all, but $B_4 \cap \partial\mathbb{L}$ coincides with the relative boundary of B_4.

Finally, we consider the projection $p = p_3 \circ p_2 \circ p_1$. The restriction of p to $\partial A \setminus \mathbb{L}_0$ is not injective:

for an element f of the relative interior of the face B_1, there exists a point $\widetilde{f}$ in the relative interior of the face B_4 satisfying $p_1(\widetilde{f}) = f$: In order to see this, consider $\lambda_1, \lambda_2, \lambda_3 > 0$ with $\lambda_1 + \lambda_2 + \lambda_3 = 1$ and

$$f \;=\; \lambda_1 \cdot (0,3,0) + \lambda_2 \cdot (0,0,3) + \lambda_3 \cdot (0,1,1) \;=\; (0, 3\,\lambda_1 + \lambda_3, 3\,\lambda_2 + \lambda_3).$$

Then the statement follows for

$$\begin{aligned}\widetilde{f} &= \frac{1}{3} \cdot \lambda_3 \cdot (3,0,0) + (\lambda_1 + \tfrac{1}{3} \cdot \lambda_3) \cdot (0,3,0) + (\lambda_2 + \tfrac{1}{3} \cdot \lambda_3) \cdot (0,0,3) \\ &= (\lambda_3, 3\,\lambda_1 + \lambda_3, 3\,\lambda_2 + \lambda_3).\end{aligned}$$

Since $\lambda_3 > 0$ we have $\widetilde{f} \neq f$. Furthermore,

$$p(\widetilde{f}) \;=\; (p_3 \circ p_2 \circ p_1)(\widetilde{f}) \;=\; (p_3 \circ p_2)(f) \;=\; (p_3 \circ p_2 \circ p_1)(f) \;=\; p(f).$$

More precisely, $(p_3 \circ p_2)(f) = p(f)$ is contained in the union of the line segments $\overline{v_2 v_4}$ and $\overline{v_3 v_4}$. From $\lambda_1 > 0$ and $\lambda_2 > 0$ it follows that $p(f)$ is distinct from f.

We point out that in this example the following holds:

$$\mathbb{L}_0 \;\subsetneq\; \partial\mathbb{L} \cap A.$$

Each point f of the relative interior of B_1 is not only contained in A but also in $\partial\mathbb{L}$. However, it is not contained in $\mathbb{L}_0$.

Question 2.13. Given j_0, we are now interested in the number

$$\begin{aligned}\underline{f_{j_0}} \quad := \quad \min\big\{ f_{j_0} \in \mathbb{R} \;:\; & (f_1, \ldots, f_{j_0}, \ldots, f_m) \in \mathbb{L} \\ & \text{for some } f_1, \ldots, f_{j_0-1}, f_{j_0+1}, \ldots, f_m \big\}.\end{aligned}$$

Here, we have to distinguish between the following two cases:

Case 1: There exists i such that in (1) there is an inequality of the form $c_i \leq \alpha_{ij_0} \cdot f_{j_0}$. Then one can assume that there is only one such inequality. In that case, we have $\underline{f_{j_0}} = \frac{c_i}{\alpha_{ij_0}}$.

Case 2: If such an i does not exist then $\underline{f_{j_0}} = 0$. This follows from (10).

Theorem 2.14. Assume $1 \leq j_1 < \cdots < j_k \leq m$. Then the following statements are equivalent:

- **(i)** There is $(f_1, \ldots, f_m) \in \mathbb{L}$ with $f_{j_\nu} = 0$ for $1 \leq \nu \leq k$.
- **(ii)** There is $(f_1, \ldots, f_m) \in \mathbb{L}_0$ with $f_{j_\nu} = 0$ for $1 \leq \nu \leq k$.
- **(iii)** For every i with $1 \leq i \leq n$ there exists $j \in \{1, \ldots, m\} \setminus \{j_1 \ldots, j_k\}$ with $\alpha_{ij} > 0$.

Proof. **(ii) ⇒ (i)**: This implication is trivial.

(i) ⇒ (ii): This follows immediately from the fact that the map $p : \mathbb{L} \to \mathbb{L}_0$ constructed in Theorem 2.1 satisfies $p(f) \leq f$ for all $f \in \mathbb{L}$.

(i) ⇒ (iii): Assume (iii) is false. Then the i'th inequality in (1) impies $c_i \leq 0$, which is impossible.

(iii) ⇒ (i): After removing all products $\alpha_{ij_\nu} \cdot f_{j_\nu}$ in (1) we get a new system of inequalities which is solvable according to (10). □

Specialization of this theorem implies:

Corollary 2.15. For $1 \le j \le m$ the following statements are equivalent:

(i) There is $(f_1, \dots, f_m) \in \mathbb{L}$ with $f_j = 0$.

(ii) There is $(f_1, \dots, f_m) \in \mathbb{L}_0$ with $f_j = 0$.

(iii) No inequality of the system (1) has the form $c_i \le \alpha_{ij} \cdot f_j$.

Definition 2.16. The system (1) is called *reduced*, if for all j with $1 \le j \le m$ the equivalent conditions of the above corollary are satisfied.

Remark 2.17. Every linear inequality system (1) can be transformed into a reduced one:

If (1) is not reduced then at least one of the inequalities has the form

$$c_{i_0} \le \alpha_{i_0 j_0} \cdot f_{j_0}.$$

From now on, we assume that there is no further such inequality with the same index j_0.

With

$$(f'_1, \dots, f'_m) \; := \; \left(f_1, \dots, f_{j_0-1}, f_{j_0} - \frac{c_{i_0}}{\alpha_{i_0 j_0}}, f_{j_0+1}, \dots, f_m\right),$$

the inequality system (1) is equivalent to the system

$$c_i - c_{i_0} \cdot \frac{\alpha_{i j_0}}{\alpha_{i_0 j_0}} \;\le\; \sum_{j=1}^{m} \alpha_{ij} \cdot f'_j \qquad \text{for } 1 \le i \le n. \tag{11}$$

Here, the inequalities with non-positive left-hand side, in particular for $i = i_0$, can be ignored. Repeating this procedure, after at most m steps we get a reduced system.

We now consider the following problem: What is the largest number k such that there exist $j_1, \dots, j_k$ with $1 \le j_1 < \dots < j_k \le m$ and also $(f_1, \dots, f_m) \in \mathbb{L}$ with

$$f_{j_\nu} = 0 \qquad \text{for } 1 \le \nu \le k?$$

This is the largest number k with the following property: k columns of the matrix $(\alpha_{ij})_{1 \le i \le n,\, 1 \le j \le m}$ can be cancelled in such a way that the remaining $n \times (m-k)$-matrix does not have any row with only zeros.

We can reinterpret this problem in terms of the bipartite graph $G = (Z \cup S, E)$ where $Z = \{z_1, \dots, z_n\}$ denotes the set of rows, $S = \{s_1, \dots, s_m\}$ denotes the set of columns, and

$$E \; := \; \{\{z_i, s_j\} \; : \; \alpha_{ij} > 0\}.$$

Then k is the largest number with the following property: There exist $m - k$ rows $s_{\nu_1}, \dots, s_{\nu_{m-k}}$ with

$$N(\{s_{\nu_1}, \dots, s_{\nu_{m-k}}\}) \; = \; Z.$$

Here, for $W \subseteq Z \cup S$, $N(W)$ denotes the set of neighbors of W.

Example 2.18. In this example, we revisit the Sections 1.1.1 and 1.1.2 of the introduction and use the notation given there.

(a) The inequalities (5) can be written as

$$\sum_{v \in S} \alpha_{A,v} \cdot I_v \;\geq\; c_A, \qquad A \subseteq 2^S \setminus \mathfrak{N}\,,$$

where $\alpha_{A,v} = 1$ if $v \in A$, and $\alpha_{A,v} = 0$ otherwise. The general results above refer to solution vectors $(I_v)_{v\in S}$ that are not necessarily induced by a Bayesian network. According to Theorem 2.14, the maximal number k of zeros the solution vector $(I_v)_{v\in S}$ might have is the maximal number of columns, indexed by v, which can be removed without having a vanishing row vector in the remaining matrix. It is easy to see that this maximal number k coincides with $\max_{A\,\in\,\mathfrak{N}}\ |A|$. This directly implies that the maximal number ν of vanishing I_v's that are induced by a Bayesian network has to be smaller than or equal to $\max_{A\,\in\,\mathfrak{N}}\ |A|$. According to the specific considerations of Section 1.1.1 we even have equality, which is a stronger statement that does not follow from our general results.

(b) We first rewrite the inequalities (9). Obviously there is a maximal g for which c_g is positive, which we denote by g^*. The number n of inequalities of type (1) coincides with the number $g^* - 1$. The number m of parameters is given by $2^{|S|} - |S| - 1$. We obtain

$$\sum_{\substack{A \subseteq S \\ |A| \geq 2}} \alpha_{g,A} \cdot H(X_{\pi_A}) \;\geq\; c_g, \qquad 2 \leq g \leq g^*\,,$$

with $\alpha_{g,A} = 1$ if $|A| \geq g$, and $\alpha_{g,A} = 0$ otherwise. According to the general results above, the minimal number of positive entropy terms is one.

3. EXTREME POINTS OF $\mathbb{L}_0$

In this section, we mainly study the following

Problem: Find recursively a point $(f_1, \ldots, f_m) \in \mathbb{L}_0$ with the following properties:

$(E.1)$	f_1 is minimal
$(E.j)$ for $2 \leq j \leq m$:	f_j is minimal with respect to the conditions $(E.1)$, …, $(E.j\!-\!1)$.

We proceed as follows.

Algorithm: Step 1:
If there exists one, and hence by our assumption (see Remark 2.17.), only one inequality of the system (1) that has the form $c_i \leq \alpha_{i1} \cdot f_1$ then we put

$$f_1 \;=\; c_i \cdot \alpha_{i1}^{-1}.$$

Otherwise, we put $f_1 = 0$.

Step j, $2 \le j \le m$:
Let $f_1, \dots, f_{j-1}$ be already determined. With these fixed values in (1) we obtain a new system of inequalities:

$$c_{ij} \; := \; c_i - \sum_{\nu=1}^{j-1} \alpha_{i\nu} \cdot f_\nu \; \le \; \sum_{\nu=j}^{m} \alpha_{i\nu} \cdot f_\nu, \qquad 1 \le i \le n. \tag{12}$$

Then those inequalities where the left hand side is non-positive are ignored. If there exists at least one inequality in (12) of the form $c_{ij} \le \alpha_{ij} \cdot f_j$, then consider the most restrictive of these inequalities and put

$$f_j \; = \; c_{ij} \cdot \alpha_{ij}^{-1}.$$

Otherwise put $f_j = 0$.

Before we analyze this algorithm, we consider the following

Special Case: For each two indices j_1, j_2 with $1 \le j_1 < j_2 \le m$ there exists an inequality in (1) of the form

$$c_i \; \le \; \alpha_{ij_1} \cdot f_{j_1} + \alpha_{ij_2} \cdot f_{j_2}. \tag{13}$$

In this case there is no $(f_1, \dots, f_m) \in \mathbb{L}$ that has at least two zeros. If we assume $f_{j_1} = f_{j_2} = 0$ then (13) would imply $c_i = 0$, which is impossible according to the assumption.

Otherwise, according to the above algorithm one can find a point $(f_1, \dots, f_m) \in \mathbb{L}_0$.

Here, each component is different from zero if and only if for all j with $1 \le j \le m$ in (1) there is one inequality of the form

$$c_i \; \le \; \alpha_{ij} \cdot f_j$$

where i depends on j.

Example 3.1. We consider the following system with $m = 3$:

$$\begin{aligned} 1 \; &\le \; f_1 + f_2 \;, \\ 2 \; &\le \; f_1 + f_3 \;, \\ 4 \; &\le \; f_2 + f_3 \;, \\ 3 \; &\le \; f_1 + f_2 + f_3 \;. \end{aligned} \tag{14}$$

Note that the last inequality follows from the first three inequalities. With the above algorithm we obtain $f_1 = 0$ and the remaining inequality system

$$1 \; \le \; f_2, \qquad 2 \; \le \; f_3, \qquad 4 \; \le \; f_2 + f_3 \;. \tag{15}$$

This yields the following solution:

$$(f_1, f_2, f_3) \; = \; (0, 1, 3) \;.$$

If we consider the modified order (f_3, f_2, f_1) then we obtain $f_3 = 0$ and

$$1 \leq f_1 + f_2, \qquad 2 \leq f_1, \qquad 4 \leq f_2 \,. \tag{16}$$

This yields the solution

$$(f_3, f_2, f_1) = (0, 4, 2) \,; \quad \text{this means} \quad (f_1, f_2, f_3) = (2, 4, 0) \,.$$

Theorem 3.2. The solution $(f_1, \ldots, f_m) \in \mathbb{L}_0$ described by the above algorithm is an extreme point of $\mathbb{L}$.

P r o o f. We prove the statement by contradiction and therefore assume that $f = (f_1, \ldots, f_m)$ is not an extreme point of $\mathbb{L}$. Then there exists $v = (v_1, \ldots, v_m) \in \mathbb{R}^m \setminus \{0\}$ with $f - v \in \mathbb{L}$ and $f + v \in \mathbb{L}$. Let j be minimal with $1 \leq j \leq m$ and $v_j \neq 0$. Without loss of generality we assume $v_j > 0$. Then step j of the algorithm, according to $(E.j)$, yields f'_j with $f'_j \leq f_j - v_j < f_j$. This is a contradiction, which completes the proof. □

Remark 3.3. In the above theorem, the converse implication is not true in general. Depending on the order of the coordinates, the described algorithm yields at most $m!$ distinct extreme points. However, for a given m it is possible to have an arbitrary number of extreme points.

Example 3.4. For $m = 2$ and $n \geq 1$, consider the following system of inequalities:

$$c_i := 2^{i-1} \cdot (n + 2 - i) - 1 \leq 2^{i-1} \cdot f_1 + f_2 \qquad \text{for } 1 \leq i \leq n.$$

The extreme points here are

$$p_i = (n - i, 2^i - 1) \qquad \text{for } 0 \leq i \leq n \,,$$

see Figure 4.

Remark 3.5. If for $m = 2$ the system (1) is reduced then there exist positive real numbers a and b, which are unique, such that $Q_1 = (0, a)$ and $Q_2 = (b, 0)$ are extreme points of $\mathbb{L}$. Each point (x, y) with $x \geq 0$, $y \geq 0$ and $a \cdot x + b \cdot y \geq a \cdot b$ lies above the line segment $\overline{Q_1 Q_2}$ and therefore also in $\mathbb{L}$. This means the following: Each extreme point of $\mathbb{L}$ lies in the closed triangle given by the points $(0, 0)$, Q_1, and Q_2.
This leads to the question whether a similar situation is also given for $m \geq 3$. More precisely, is it true that each extreme point of $\mathbb{L}$ lies in the convex hull of the origin and the lexicographically minimal solutions of $\mathbb{L}$ with respect to all $m!$ possible orderings of the coordinates? The next example proves that this is actually not always the case.

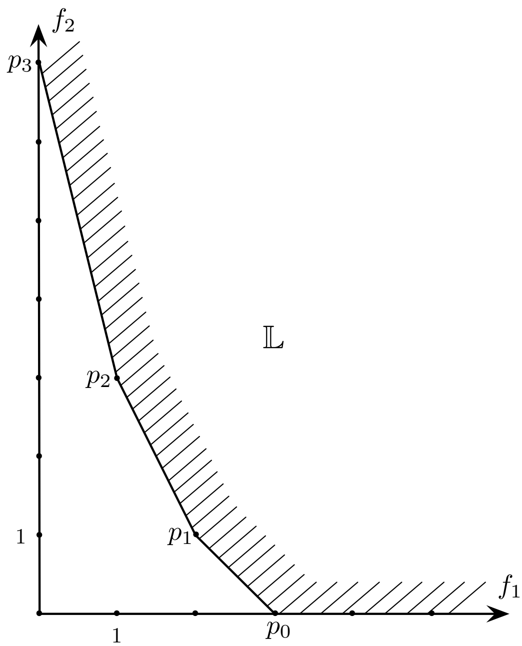

Fig. 4. Illustration of Example 3.4 in the case $n = 3$.

Example 3.6. For $m = 4$ we consider the following system of linear inequalities in which all non-vanishing coefficients have the value 1:

$$\begin{aligned} 1 &\leq f_i + f_j \qquad \text{for } 1 \leq i < j \leq 4\,, \\ \tfrac{3}{2} &\leq f_1 + f_2 + f_3\,, \\ 2 &\leq f_i + f_j + f_4 \qquad \text{for } 1 \leq i < j \leq 3\,, \\ 3 &\leq f_1 + f_2 + f_3 + f_4\,. \end{aligned}$$

Note that the inequality given in the second line of this system is redundant. It follows by addition of the three inequalities of the form

$$1 \leq f_i + f_j \qquad \text{for } 1 \leq i < j \leq 3\,.$$

We obtain the following lexicographically minimal solutions in $\mathbb{L}_0$ depending on the orderings of the coordinates:

$$Q_1 = (0,1,1,1), \quad Q_2 = (1,0,1,1), \quad Q_3 = (1,1,0,1), \quad Q_4 = (1,1,1,0)\,.$$

However,

$$\overline{Q} = (\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{3}{2})$$

is also an extreme point of $\mathbb{L}$. This is the unique intersection point of the following four affine hyperplanes:

$$\begin{aligned} H_1 &= \{(x_1,x_2,x_3,x_4) \in \mathbb{R}^4 \ : \ 1 = x_1 + x_2\} , \\ H_2 &= \{(x_1,x_2,x_3,x_4) \in \mathbb{R}^4 \ : \ 1 = x_1 + x_3\} , \\ H_3 &= \{(x_1,x_2,x_3,x_4) \in \mathbb{R}^4 \ : \ 1 = x_2 + x_3\} , \\ H_4 &= \{(x_1,x_2,x_3,x_4) \in \mathbb{R}^4 \ : \ 3 = x_1 + x_2 + x_3 + x_4\} . \end{aligned}$$

These hyperplanes are supporting hyperplanes of $\mathbb{L}$. The point $\overline{Q}$ has a coordinate with value $\frac{3}{2}$ and is therefore not contained in

$$\mathsf{conv}(\{0, Q_1, Q_2, Q_3, Q_4\}) \ \subseteq \ [0,1]^4.$$

With the same argument it follows that there is no $Q \in \mathsf{conv}(\{0, Q_1, Q_2, Q_3, Q_4\})$ with $p(Q) = \overline{Q}$: For $Q \in \mathbb{L} \cap [0,1]^4$ one also has $p(Q) \in [0,1]^4$.

APPENDIX

In this appendix we provide the technical definitions of directed acyclic graphs and Bayesian networks informally used in the introduction.

3.1. Directed acyclic graphs

We consider a *directed graph* $G := (V, E)$ where $V \neq \emptyset$ is a finite set of *nodes* and $E \subseteq V \times V$ is a set of *edges* between the nodes. An ordered sequence $(v_0, \ldots, v_k)$, $k \geq 0$, of distinct nodes is called a (*directed*) *path* from v_0 to v_k with *length* k if it satisfies $(v_i, v_{i+1}) \in E$ for all $i = 0, \ldots, k-1$. Given two subsets A and B of V, and a path $\gamma = (v_0, \ldots, v_k)$ with $v_0 \in A$ and $v_k \in B$, we write $A \overset{\gamma}{\rightsquigarrow} B$. If there exists a path γ such that $A \overset{\gamma}{\rightsquigarrow} B$ we write $A \rightsquigarrow B$, and $A \not\rightsquigarrow B$ if this is not the case. Note that $v \rightsquigarrow v$ for all $v \in V$ (path of length 0). A *directed acyclic graph* (*DAG*) is a graph that does not contain two distinct nodes v_0 and v_k with $v_0 \rightsquigarrow v_k$ and $v_k \rightsquigarrow v_0$.

Given a DAG, we define the *parents* of a node v as $\mathsf{pa}(v) := \{u \in V : (u, v) \in E\}$ and its *children* as $\mathsf{ch}(v) := \{w \in V : (v, w) \in E\}$. A set $C \subseteq V$ is called *ancestral* if for all $v \in C$ the parents $\mathsf{pa}(v)$ are also contained in C. The smallest ancestral set that contains a set A is denoted by $\mathsf{an}(A)$, and one has

$$\mathsf{an}(A) = \{v \in V \ : \ v \rightsquigarrow A\} . \tag{17}$$

3.2. Bayesian networks

For every node $v \in V$ we consider a finite and non-empty set $\mathbb{X}_v$ of states. Given a subset $A \subseteq V$, we write $\mathbb{X}_A$ instead of $\prod_{v \in A} \mathbb{X}_v$ (*configuration set on* A), and we have the natural projection

$$X_A : \mathbb{X}_V \to \mathbb{X}_A, \qquad (x_v)_{v \in V} \mapsto x_A := (x_v)_{v \in A}.$$

Note that in case $A = \emptyset$, the configuration set consists of exactly one element, namely the empty configuration which we denote by ϵ.

A *distribution* on $\mathbb{X}_V$ is a vector $p = (p(x))_x \in \mathbb{R}^{\mathbb{X}_V}$ with $p(x) \geq 0$ for all $x \in \mathbb{X}_V$ and $\sum_x p(x) = 1$. Given a distribution p on $\mathbb{X}_V$, the X_A's become random variables, and we write

$$p(x_A) \ := \sum_{x_{V\setminus A} \in \mathbb{X}_{V\setminus A}} p(x_A, x_{V\setminus A})$$

and, if $p(x_A) > 0$,

$$p(x_B|x_A) \ := \ \frac{p(x_A, x_B)}{p(x_A)} \,. \tag{18}$$

In particular, we have $p(x_B|\epsilon) = p(x_B)$ if $A = \emptyset$.
Given a DAG, we consider a family of conditional distributions $\kappa^v(x_{\mathsf{pa}(v)}; x_v)$, $v \in V$, that is

$$\kappa^v(x_{\mathsf{pa}(v)}; x_v) \geq 0 \quad \text{and} \quad \sum_{x_v} \kappa^v(x_{\mathsf{pa}(v)}; x_v) = 1.$$

If $\mathsf{pa}(v) = \emptyset$ we write $\kappa^v(x_v)$ instead of $\kappa^v(\epsilon; x_v)$. A triple $\mathscr{B} = (V, E, \kappa)$ consisting of a directed acyclic graph $G = (V, E)$ and such a family $\kappa = (\kappa^v)_{v \in V}$ of kernels is called a *Bayesian network*.

Given a Bayesian network $\mathscr{B}$, the corresponding joint distribution on $\mathbb{X}_V$ is defined as follows:

$$p(x) \ = \ p(\mathscr{B}; x) \ := \prod_{v \in V} \kappa^v(x_{\mathsf{pa}(v)}; x_v). \tag{19}$$

If a given distribution p on $\mathbb{X}_V$ can be decomposed in this way, we say that it *admits a recursive factorization with respect to* G. In that case one has $\kappa^v(x_{\mathsf{pa}(v)}; x_v) = p(x_v|x_{\mathsf{pa}(v)})$ if $p(x_{\mathsf{pa}(v)}) > 0$.

ACKNOWLEDGEMENTS

Both authors thank Bastian Steudel for helpful discussions. Walter Wenzel has been supported by the Max Planck Institute for Mathematics in the Sciences.

(Received April 21, 2011)

REFERENCES

[1] N. Ay: A refinement of the common cause principle. Discrete Appl. Math. *157* (2009), 2439–2457.

[2] N. Ay and D. Polani: Information flows in causal networks. Adv. Complex Systems *11* (2008), 1, 17–41.

[3] H. Martini and V. Soltan: Combinatorial problems on the illumination of convex bodies. Aequationes Math. *57* (1999), 121–152.

[4] H. Martini and W. Wenzel: Illumination and visibility problems in terms of closure operators. Beiträge zur Algebra und Geometrie *45* (2004), 2, 607–614.

[5] J. Pearl: Causality: Models, Reasoning and Inference. Cambridge University Press 2000.

[6] B. Steudel and N. Ay: Information-theoretic inference of common ancestors. Submitted. ArXiv preprint (2010) arXiv:1010.5720.

[7] F. A. Valentine: Visible shorelines. Amer. Math. Monthly *77* (1970), 146–152.

[8] R. Webster: Convexity. Oxford University Press 1994.

[9] G. Ziegler: Lectures on Polytopes. Springer Verlag Berlin 1997.

Nihat Ay, Max Planck Institute for Mathematics in the Sciences. Inselstrasse 22, 04103 Leipzig, Germany and Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501. U. S. A.
e-mail: nay@mis.mpg.de

Walter Wenzel, Max Planck Institute for Mathematics in the Sciences. Inselstrasse 22, 04103 Leipzig, Germany and Universität Kassel, Institut für Mathematik, 34109 Kassel. Germany.
e-mail: wenzel@mis.mpg.de