KYBERNETIKA — VOLUME *34* (1998), NUMBER 1, PAGES 57 – 68

ON FACTORIZATIONS OF PROBABILITY DISTRIBUTIONS OVER DIRECTED GRAPHS

František Matúš and Bernhard Strohmeier

Four notions of factorizability over arbitrary directed graphs are examined. For acyclic graphs they coincide and are identical with the usual factorization of probability distributions in Markov models. Relations between the factorizations over circuits are described in detail including nontrivial counterexamples. Restrictions on the cardinality of state spaces cause that a factorizability with respect to some special cyclic graphs implies the factorizability with respect to their, more simple, strict edge-subgraphs. This gives sometimes the possibility to break circuits and get back to the acyclic, well-understood case.

1. INTRODUCTION

During the last two decades graphs have been intensively employed to specify models for associations among random variables. Vertices of the graphs correspond to the variables and various types of edges give usually rise to assumptions on conditional independences or on the form of factorizations of probability distributions, see [4], [10] and [2]. Though the focus has been mainly on acyclic or modularly acyclic graphs, a progress has been reported also on models with feedback, see [7]. Even an elegant generalization of the close relation between the Markov properties and Gibbs factorizations, see [5] and [6], was achieved for a very general class of graphs.

For acyclic directed graphs, the widely accepted models are defined by the recursive factorization formula that consists of the product of Markov kernels depending, in the condition, on parental vertices. The formula does not necessarily provide a probability distribution when extended mechanically on arbitrary directed graphs. The starting point of this note was our endeavour to understand those cases when a probability distribution does come out and to describe the class of probability distributions obtained in this way; they are called *recursively factorizable* here. On the way, three other kinds of factorizations were found to be of some interest, namely a *marginal, consistent* and *projective* one.

By absence of cycles, the four kinds of factorizations coincide and bring nothing new. The marginal and consistent factorizations lead sometimes to trivial classes of models, see Lemma 2.3. The relation between the projective and recursive factorizations reminds the old classical question about the existence of a positive eigenvector of a stochastic matrix, cf. Example 3.3.

The main attention is focused here on circuits. Under binarity restrictions on cardinality of the state spaces of variables, factorizability over a circuit is proved to imply factorizability over a path in the circuit, see Theorem 4.3. This phenomenon, called here *arrow erasure,* occurs also for noncircuits, cf. Lemma 4.6.

2. BASIC OBSERVATIONS

Let V be a finite nonempty set of *vertices* and $E \subset \{(u,v) \in V \times V;\ u \neq v\}$ be a set of *arrows*. The pair $G = (V,E)$ is called here *graph;* usually one speaks about the "directed graph without loops and multiple edges". Two opposite arrows (u,v) and (v,u) are allowed at the same time. For every $v \in V$ the elements of the set $pa(v) = \{u \in V;\ (u,v) \in E\}$ are the *parents* of v and $cl(v) = v \cup pa(v)$. We make no difference between elements v and singletons $\{v\}$ of V.

With every vertex $v \in V$, a finite nonempty *state space* X_v is associated and X_A stands for the Cartesian product of X_v over $v \in A$ where $A \subset V$ is any vertex set. Elements of X_A are denoted by x_A; for $A = V$ the subindices are omitted and for $A = \emptyset$ the set $X_\emptyset$ is supposed to have only one element $x_\emptyset$. The coordinate projection of X on X_A works as $x \to x^A$. Marginals of a probability distribution P_A on X_A are denoted as P_A^B, $B \subset A$.

Definition 2.1. A probability distribution P on X factorizes w.r.t. $G = (V,E)$

recursively *if* $P(x) = \prod_{v \in V} \psi_v(x^v | x^{pa(v)})$, $x \in X$,
for some nonnegative functions ψ_v *on* $X_v \times X_{pa(v)}$, *henceforth kernels, such that* $\sum_{y_v \in X_v} \psi_v(y_v|x_{pa(v)}) = 1$, $x_{pa(v)} \in X_{pa(v)}$,

projectively *if* $P(x) = \prod_{v \in V} \left[Q_v(x^{cl(v)}) / Q_v^{pa(v)}(x^{pa(v)}) \right]$, $x \in X$,
for some probability distributions Q_v *on* $X_{cl(v)}$, $v \in V$, *such that the projectivity conditions* $Q_u^{pa(u) \cap cl(v)} = Q_v^{pa(u) \cap cl(v)}$ *take place for any* $u, v \in V$,

consistently *if* $P(x) = \prod_{v \in V} \left[Q^{cl(v)}(x^{cl(v)}) / Q^{pa(v)}(x^{pa(v)}) \right]$, $x \in X$,
for some probability distribution Q *on* X,

marginally *if* $P(x) = \prod_{v \in V} \left[P^{cl(v)}(x^{cl(v)}) / P^{pa(v)}(x^{pa(v)}) \right]$, $x \in X$,

where 0 in a denominator occurs only with 0 in the corresponding numerator and the ratio is then taken as equal to 0.

If a probability distribution P is marginally factorizable w.r.t. a graph then it has obviously a consistent factorization w.r.t. the same graph via $Q = P$. If P is consistently factorizable via Q then it must be also projectively factorizable via $Q_v = Q^{cl(v)}$, $v \in V$. In symbols, $\mathsf{MF} \Rightarrow \mathsf{CF} \Rightarrow \mathsf{PF}$. For (everywhere) positive probability distributions the implication $\mathsf{PF} \Rightarrow \mathsf{RF}$ holds, defining the kernels ψ_v obviously as $(x_v | x_{pa(v)}) \to Q_v(x_v, x_{pa(v)}) / Q_v^{pa(v)}(x_{pa(v)})$ for all $v \in V$.

The following lemma describes factorizations w.r.t. *acyclic* graphs; $G = (V,E)$ is called acyclic if every its *path* $v_1, \ldots, v_{n+1}$, $n \geq 1$, has $v_1 \neq v_{n+1}$. Here the path is a sequence of vertices such that $(v_i, v_{i+1}) \in E$, $1 \leq i \leq n$.

Lemma 2.2. The four kinds of factorizations w.r.t. any acyclic graph coincide.

P r o o f. It suffices to show PF $\Rightarrow$ RF $\Rightarrow$ MF and this will be done by induction on the cardinality n of the vertex set. For $n = 1$ every probability distribution on X is factorizable in any of the four ways. Let us assume that the implications are valid for all acyclic graphs with $n \geq 1$ vertices and let $G = (V, E)$ be an acyclic graph with $n + 1$ vertices. A *terminal* vertex $u \in V$ ($(u, v) \notin E$ for all $v \in V$) of G exists and is fixed arbitrarily.

Let a probability distribution P be PF w.r.t. G via some distributions Q_v, $v \in V$. Then $P^{V-u}(x_{V-u})$ for $x_{V-u} \in X_{V-u}$ equals 0 or

$$\widehat{P}_{V-u}(x_{V-u}) = \prod_{v \in V-u} Q_v(x_{V-u}^{cl(v)}) / Q_v^{pa(v)}(x_{V-u}^{pa(v)})$$

according to whether $Q_u^{pa(u)}(x_{V-u}^{pa(u)})$ equals zero or not, respectively. We are going to show that $Q_u^{pa(u)}(x_{V-v}^{pa(u)}) = 0$ implies $\widehat{P}_{V-u}(x_{V-u}) = 0$. In fact, in the opposite case we would have $\sum_{x \in X} P(x) < \sum_{x_{V-u} \in X_{V-u}} \widehat{P}(x_{V-u})$ and, continuing with $\widehat{P}_{V-u}$ over $(V - u, E \cap (V - u)^2)$, a repetition of this reasoning would lead to $1 < 1$. We conclude $P^{V-u} = \widehat{P}_{V-u}$ whence the marginal P^{V-u} is PF via Q_v, $v \in V - u$. By induction, P^{V-u} is RF via some ψ_v, $v \in V - u$. Adding the kernel ψ_u defined by $\psi_u(x_u|x_{pa(u)}) = Q_u(x_u, x_{pa(u)})/Q_u^{pa(u)}(x_{pa(u)})$ if the denominator is positive and by $\psi_u(x_u|x_{pa(u))} = |X_u|^{-1}$ otherwise, the probability distribution P is RF w.r.t. G via ψ_v, $v \in V$.

If P is RF w.r.t. G via some kernels ψ_v, $v \in V$, then the marginal distribution P^{V-u} factorizes recursively with respect to $(V - u, E \cap (V - u)^2)$ through the kernels ψ_v, $v \in V - u$. Obviously, $P(x)$ equals $P^{V-u}(x^{V-u})\psi_u(x^u|x^{pa(u)})$ and then $P^{cl(u)}(x^{cl(u)}) = P^{pa(u)}(x^{pa(u)})\psi_u(x^u|x^{pa(u)})$, $x \in X$, by marginalization. The induction assumption implies that P^{V-u} is MF and this factorization combined with the previous two equalities yield the MF of P w.r.t. G. Note that $P^{pa(u)}(x^{pa(u)}) = 0$ entails $P(x) = P^{V-u}(x^{V-u}) = 0$. $\square$

Due to Lemma 2.2 all factorizations from Definition 2.1 are generalizations of the usual recursive factorization in the Markov models over acyclic graphs, see [4]. Whereas the way to the definitions of MF and RF was straightforward, the definitions of CF and PF emerged later as alternatives to the MF behaving sometimes "pathologically". An example of this behaviour follows.

Lemma 2.3. Let $K_n = (V, E)$ where $V = \{1, 2, \ldots, n\}$, $n \geq 1$, and let E contain all arrows (u, v) with different endpoints $u, v \in V$. A probability distribution P on X is MF w.r.t. K_n if and only if P is CF w.r.t. K_n and this is equivalent to the marginal factorization $P(x) = \prod_{v \in V} P^v(x^v)$, $x \in X$, of P w.r.t. $(V, \emptyset)$.

P r o o f. The only nontrivial claim is that CF implies the product formula. If P is CF w.r.t. the graph then $P = \prod_{v \in V} Q/Q^{V-v}$ for some distribution Q on X. Since Q is absolutely continuous w.r.t. P the I-divergence $I(Q\|P)$, see [9], [10], is a nonnegative real number. This yields $(n-1)h(V) \geq \sum_{v \in V} h(V - v)$ where

$h(A)$ is the Shannon entropy of Q^A, $A \subset V$. The set function h is submodular, i.e. $h(A)+h(B) \geq h(A \cup B)+h(A \cap B)$, for any A, $B \subset V$, whence $h(V) \leq \sum_{v \in V} h(v)$. We will prove in a moment that the latter inequality cannot be strict what means that P equals the desired product.

Let c_k denote the sum of $h(A)$ over all $A \subset V$ of cardinality k, $1 \leq k \leq n$. We know that $(n-1)c_n \geq c_{n-1}$ and want to show that $c_n \geq c_1$. The number $2k(n-k)c_k$, $1 < k < n$, can be casted into

$$\sum \big\{h(u \cup A) + h(v \cup A);\ A \subset V,\ |A| = k-1,\ u,v \in V - A,\ u \neq v\big\} \\ \geq (n-k+1)(n-k)\, c_{k-1} + k(k+1)\, c_{k+1}$$

owing to the submodularity of h. Thus we see that the inequality

$$(\ell+1)\, c_{k+\ell} \;\geq\; \binom{n}{k+\ell}\binom{n}{k}^{-1} c_k \;+\; \ell\, \frac{k+\ell+1}{n-k-\ell}\, c_{k+\ell+1}\,, \quad 1 \leq k < k+\ell < n\,,$$

is valid for $\ell = 1$. If it is valid for some $1 \leq \ell < n-k-1$ then we combine it with the previous one for $k \to k+\ell+1$, exclude $c_{k+\ell}$ and obtain it also for $\ell+1$. By induction, the inequality holds for $\ell = n-2$ and $k = 1$, i.e. $(n-1)c_{k-1} \geq c_1 + (n-2)c_n$ and we arrive at the desired inequality $c_n \geq c_1$. □

Informally rephrased, $\mathsf{MF} = \mathsf{CF}$ over K_n, $n \geq 1$, and this amounts the mutual stochastic independence. On the other hand, every probability distribution P is RF w.r.t. K_n; even w.r.t. any graph $G = (V,E)$ with $V = \{1,\dots,n\}$ and E containing $E^< = \{(i,j);\ 1 \leq i < j \leq n\}$, $n \geq 1$. It is namely always possible to factorize P w.r.t. $(V, E^<)$ via some kernels $\psi_v^<$, $v \in V$, and then define the new kernels $\psi_v(y_v|x_{pa(v)}) = \psi_v^<(y_v|y_{v^<})$, $v \in V$, where $v^< = \{u \in V;\ u < v\}$ and $y_{v^<}$ is the coordinate projection of $x_{pa(v)}$ on $X_{v^<}$. The new kernels ψ_v, $v \in V$, factorize projectively P w.r.t. the graph G.

In the case of K_2 every probability distribution is PF, too. In fact, if P is a probability distribution on $X_1 \times X_2$ we set $Q_1 = P$ and $Q_2 = P^1 P^2$. The probability distributions Q_1 and Q_2 satisfy the projectivity conditions $Q_1^2 = Q_2^2$, $Q_2^1 = Q_1^1$ and P factorizes projectively via Q_1 and Q_2. So that $\mathsf{MF} = \mathsf{CF} \Rightarrow \mathsf{PF} = \mathsf{RF}$ over K_2 and the implication cannot be reversed. We conjecture that over K_3 there exists a probability distribution that is not PF (and is RF as we saw above); its construction might be similar to the construction of Example 3.3 below.

Note that if the intriguing projectivity conditions in Definition 2.1 had been stated, maybe more naturally, as $Q_u^{cl(u)\cap cl(v)} = Q_v^{cl(u)\cap cl(v)}$, $u, v \in V$, we would have had even the pathology by the "projective factorization" over K_n, $n \geq 2$.

3. EXAMPLES OF FACTORIZATIONS

Let $V = \{1,2,\dots,n\}$, $n \geq 2$, and $E = \{(v^-, v) \in V^2; v \in V\}$ where $v^- = v-1$ for $1 < v \leq n$ and $v^- = n$ for $v = 1$. The graph $C_n = (V,E)$ is called a *circuit*. In this section all factorizations are w.r.t. C_3. The situation is more interesting than over $C_2 = K_2$. Namely, in the chain of implications $\mathsf{MF} \Rightarrow \mathsf{CF} \Rightarrow \mathsf{PF} \Rightarrow \mathsf{RF}$, the last one owing to Lemma 4.1 to be proved in the next section, no reversion occurs. The following three examples and figures demonstrate it, respectively.

Example 3.1. (CF $\not\Rightarrow$ MF) Let $X_1 = X_2 = X_3 = \{0,1\}$ and the probability distribution P on $X = \{0,1\}^3$ be given by $P(x_1x_20) = 1/8$ for $x_1, x_2 \in \{0,1\}$ and by $P(001) = 1/2$, see Figure 1 left. Obviously, this distribution P is not MF because $P(001)P^1(0)P^2(0)P^3(1)$ equals $1/2 \cdot 3/4 \cdot 3/4 \cdot 1/2$ whereas $P^{12}(00)P^{13}(01)P^{23}(01)$ equals $5/8 \cdot 1/2 \cdot 1/2$. However, P factorizes consistently through the probability distribution Q given by $Q(000) = Q(110) = 1/16$, $Q(100) = Q(010) = 3/16$, and $Q(001) = 1/2$, see Figure 1 right. Indeed, owing to $Q^{12} = Q^1 \cdot Q^2$, $Q^{13} = P^{13}$, and $Q^{23} = P^{23}$ the equality

$$P = Q^{12}/Q^1 \cdot Q^{23}/Q^2 \cdot Q^{13}/Q^3$$

is equivalent to $P \cdot P^3 = P^{13}P^{23}$ what is the case.

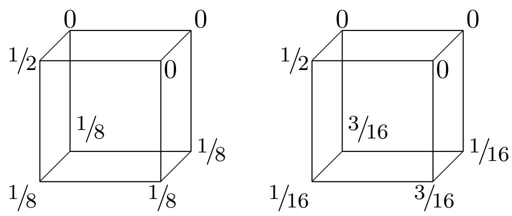


Fig. 1. The left distribution is CF via the right one, but not MF.

Example 3.2. (PF $\not\Rightarrow$ CF) Let the state spaces be exactly as in Example 3.1 and let us take $P(x) = 1/4$ for x equal to 000, 100, 101 and 111, see Figure 2. The probability distribution P is PF via $Q_1 = P^{13}$, $Q_2 = P^1P^2$ and $Q_3 = P^{23}$ by the same argument as in the previous example. We claim that P is not CF; if it were through some probability distribution Q on X then $Q^{23}(10) = 0$ and $Q^{13}(01) = 0$. Further,

$$1 = \frac{P(100)}{P(101)} = \frac{Q^{23}(00)Q^{13}(10)Q^3(1)}{Q^{23}(01)Q^{13}(11)Q^3(0)} = \frac{Q^{13}(10)}{Q^{23}(01)} = \frac{Q(100)}{Q(101)}$$

whence $Q(100) = Q(101) = b$. From

$$P(000) = \frac{Q(000)}{Q^2(0)} = \frac{Q(000)}{Q(000) + 2b} = P(111) = \frac{Q(111)}{Q^1(1)} = \frac{Q(111)}{Q(111) + 2b}$$

we deduce $Q(000) = Q(111) = a$. Then $2a + 2b = 1$ and $P(000) = 1/4 = a/(a+2b)$ imply $a = 1/5$ and $b = 3/10$. But, $1/4 = P(100) \neq \left(3/5 \cdot 1/2 \cdot 3/10\right)/\left(4/5 \cdot 4/5 \cdot 1/2\right)$, a contradiction.

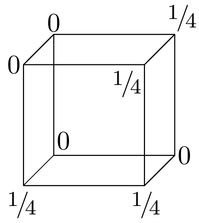


Fig. 2. This probability distribution is PF but not CF.

Example 3.3. * (RF $\not\Rightarrow$ PF) Let $X_1 = X_2 = X_3 = \{0,1,2\}$ and let a distribution P on X be given by $P(x) = 1/8$ for $x \in \{000, 222\}$ and $P(x) = 1/4$ for $x \in \{100, 010, 001\}$, see Figure 3 left. Let us suppose that three kernels ψ_1, ψ_2, ψ_3 factorize recursively P as $\psi_1\psi_2\psi_3$. Then

$$1/64 = P(100)P(010)P(001) = \prod_{v \in V} \psi_v(1|0)\psi_v(0|0)\psi_v(0|1) \leq \prod_{v \in V} 1/4\, \psi_v(0|1)$$

and we obtain $\psi_v(0|1) = 1$ and $\psi_v(1|0) = \psi_v(0|0) = 1/2$, $v \in V = \{1,2,3\}$. On the other hand, if a triple of kernels satisfies the previous nine equalities and $\prod_{v \in V} \psi_v(2|2) = 1/8$ then $P = \psi_1\psi_2\psi_3$, as in Figure 3 right. Hence, P is RF and we found even all triples of kernels providing this kind of factorization.

If the probability distribution P were PF via some Q_1, Q_2, and Q_3 then the marginals $Q_1^3 = Q_3^3$, $Q_2^1 = Q_1^1$, and $Q_3^2 = Q_2^2$ must be positive probability distributions and $\psi_1 = Q_1/Q_1^3$, $\psi_2 = Q_2/Q_2^1$, and $\psi_3 = Q_3/Q_3^2$ are three well-defined kernels factorizing P recursively into $\psi_1\psi_2\psi_3$. But then we can compute

$$\begin{aligned} Q_3^3(x_3) &= \sum_{x_2 \in X_2} \psi_3(x_3|x_2) Q_2^2(x_2) = \sum_{x_2 \in X_2} \psi_3(x_3|x_2) \sum_{x_1 \in X_1} \psi_2(x_2|x_1) Q_1^1(x_1) \\ &= \sum_{x_2 \in X_2} \psi_3(x_3|x_2) \sum_{x_1 \in X_1} \psi_2(x_2|x_1) \sum_{y_3 \in X_3} \psi_1(x_1|y_3) Q_3^3(y_3) = \sum_{y_3 \in X_3} \phi(x_3|y_3) Q_3^3(y_3) \end{aligned}$$

where

$$\phi(x_3|y_3) = \sum_{x_1 \in X_1} \sum_{x_2 \in X_2} \psi_1(x_1|y_3)\psi_2(x_2|x_1)\psi_3(x_3|x_2)$$

is a kernel on $X_3 \times X_3$. If ϕ is considered for a 3×3 matrix with its rows indexed by $y_3 = 0,1,2$ and columns by $x_3 = 0,1,2$, then

$$\phi = \begin{pmatrix} 5/8 & 3/8 & 0 \\ 3/4 & 1/4 & 0 \\ a & b & 1/8 \end{pmatrix}$$

where $a \geq 0$ and $b \geq 0$ depend on ψ_1, ψ_2, and ψ_3 and $a+b = 7/8$. Hence, we know that $\big(Q_3^3(0), Q_3^3(1), Q_3^3(2)\big)$ is a positive left eigenvector of the stochastic matrix ϕ. This matrix has, however, no positive left eigenvector what contradicts the assumption

*Cf. Theorem 8 in [8], p. 33, where the assumption of positivity is missing.

P be PF. In fact, a or b is positive whence $0, 1$ are its transient states, see [1], Theorem 3.10, p. 40.

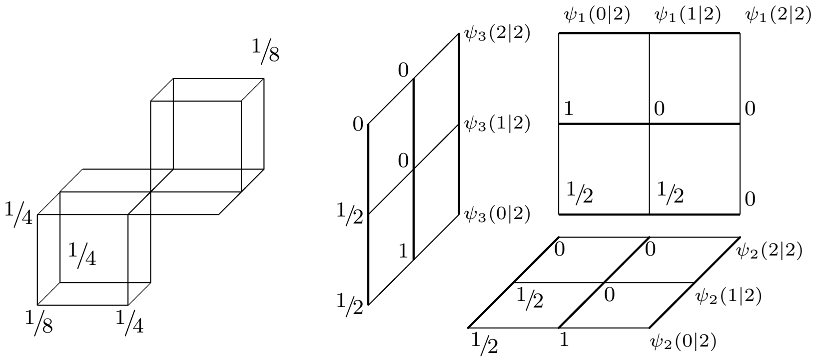


Fig. 3. The left probability distribution is RF but not PF.

4. FACTORIZATIONS OVER CIRCUITS

In this section the implication PF ⇒ RF w.r.t. C_n announced earlier will be demonstrated. Under restrictions on cardinalities of the state spaces also the reversed implication appears to be true. Stronger restrictions of this kind cause the arrow-erasure-phenomenon as exhibited in Theorem 4.3 and in Lemma 4.6.

Lemma 4.1. The projective factorization w.r.t. a circuit implies the recursive factorization w.r.t. the same circuit.

Proof. If P is PF w.r.t. C_n, $n \geq 2$, then $P = \prod_{v \in V} Q_v / Q_v^{v^-}$ for some distributions Q_v on $X_{\{v,v^-\}}$, $v \in V$, satisfying the projectivity conditions $Q_v^{v^-} = Q_{v^-}^{v^-}$, $v \in V$. Let us denote by $Y_v = \{x_v \in X_v;\ Q_v^v(x_v) > 0\}$. One defines

$$\psi_v(x_v|x_{v^-}) = \begin{cases} Q_v(x_v, x_{v^-})/Q_v^{v^-}(x_{v^-}), & x_{v^-} \in Y_{v^-}, \\ |Y_v|^{-1}, & x_{v^-} \notin Y_{v^-},\ x_v \in Y_v, \\ 0, & x_{v^-} \notin Y_{v^-},\ x_v \notin Y_v\,. \end{cases}$$

We claim that P factorizes projectively via $\psi_1, \ldots, \psi_n$. For $x \in Y_1 \times \cdots \times Y_n$, this is obvious. If $x \in X$ has its coordinate x_v in $X_v - Y_v$ for some $v \in V$ then $P(x) = 0$. In this case also $\psi_v(x_v|x_{v^-})$ is equal to zero when $x_{v^-} \in X_{v^-} - Y_{v^-}$ (by the definition of ψ_v). Otherwise $\psi_v(x_v|x_{v^-})$ does not exceed $Q_v^v(x_v)/Q_{v^-}^{v^-}(x_{v^-}) = 0$. □

Proposition 4.2. If a probability distribution P on X is RF w.r.t. C_n, $n \geq 2$, and $|X_v| \leq 2$ for at least one $v \in V$ then P is also PF w.r.t. C_n.

Proof. Without any loss of generality we can assume $|X_n| \leq 2$. Let P have a RF over $C_n = (V, E)$ via $\psi_1, \ldots, \psi_n$. If $\psi_1(x_1|x_n) = \psi_1(x_1|y_n)$, $x_1 \in X_1$, $x_n, y_n \in X_n$, i.e. ψ_1 does not depend on its condition, then P has a RF over $(V, E - \{(n,1)\})$ and then by Lemma 2.2 P is MF over the latter acyclic graph. Hence, P is PF over C_n via $Q_v = P^{v \cup v^-}$, $1 < v \le n$, and $Q_1 = P^1 \cdot P^n$. From now on we can suppose $X_n = \{0,1\}$.

Let ϕ be the Markov kernel on $X_n \times X_n$ given by

$$\phi(x_n|y_n) = \sum_{x_1 \in X_1} \cdots \sum_{x_{n-1} \in X_{n-1}} \psi_1(x_1|y_n)\psi_2(x_2|x_1)\cdots\psi_n(x_n|x_{n-1})$$

and let us assume first $\phi(1|0) = 0$. We set $\psi_1'(x_1|x_n) = \psi_1(x_1|0)$ for $x_1 \in X_1$, $x_n \in X_n$, and claim that the kernels $\psi_1', \psi_2, \dots, \psi_n$ provide a RF of P over C_n. This is obvious for $x \in X$ such that $x^n = 0$. In the opposite case, $x^n = 1$, we observe that the product $\psi_1 \cdots \psi_n$ sums to 1 whence $0 = \phi(0|0) + \phi(1|1) - 1 = \phi(1|1) = P^n(1)$ and thus $P(x) = 0$. At the same time, $0 = \phi(1|0) \ge \psi_1'(x^1|x^n)\cdots\psi_n(x^n|x^{n-1})$. We conclude that P has also the RF $\psi_1'\psi_2\cdots\psi_n$ with the factor ψ_1' not depending on its condition and that P is therefore PF. In the case $\phi(1|0) = 0$, or symmetrically $\phi(0|1) = 0$, we are done.

Let both $a = \phi(1|0)$ and $b = \phi(0|1)$ be positive. The positive distribution R_n on X_n given by $R_n(0) = {}^a\!/\!{}_{(a+b)}$ and $R_n(1) = {}^b\!/\!{}_{(a+b)}$ is stable for the kernel ϕ. That means $R_n(x_n) = \sum_{y_n \in X_n} \phi(x_n|y_n)R_n(y_n)$, $x_n \in X_n$. We define recursively $Q_v = \psi_v R_{v^-}$ and $R_v = Q_v^v$ for $1 \le v < n$ and $Q_n = \psi_n R_{n-1}$. These Q_v, $v \in V$, are claimed to provide a PF of P. Namely, the projectivity conditions are satisfied by the definition and by the stability, $Q_n^n = R_n = Q_1^n$. Further, $P = \psi_1 \cdots \psi_n$ equals $\big(Q_1/R_n\big)\big(Q_2/R_1\big)\cdots\big(Q_n/R_{n-1}\big)$ obviously for $x \in X$ such that the product $R_1(x^1)\cdots R_n(x^n)$ is positive. In the opposite case, we take the smallest $1 \le v < n$ such that $R_v(x^v) = 0$ and deduce the inequality $0 = Q_v^v(x^v) \ge \psi_v(x^v|x^{v^-})R_{v^-}(x^{v^-})$, i.e. $\psi_v(x^v|x^{v^-}) = 0$. Note that the positivity of R_n was crucial. □

As a consequence we see that Example 3.3, witnessing RF $\not\Rightarrow$ PF over C_3, is minimal not only because the underlying graph has the smallest possible number of nodes and arcs but also because all its state spaces have minimal possible cardinalities.

Let us remark that the assumption on the binarity of a state space in Proposition 4.2. can be replaced by the assumption on positivity of P. The proof then works similarly (the Markov kernel ϕ is now positive and does have therefore a positive stable distribution R_n).

Theorem 4.3. If a probability distribution P on X is RF w.r.t. C_n, $n \ge 2$, through some kernels ψ_v, $v \in V$, and $|X_v| \le 2$ for all $v \in V$ then $\psi_v(x_v|x_{v^-}) = \psi_v(x_v|y_{v^-})$, $x_v \in X_v$, $x_{v^-}, y_{v^-} \in X_{v^-}$ for at least one $v \in V$.

P r o o f. When some state space X_v is a singleton, the assertion is valid trivially. We can take therefore $X_v = \{0,1\}$, $v \in V$. If $P = \psi_1\psi_2$ over C_2 then

$$\begin{aligned} 0 &= 1 - \sum_{x_2 \in X_2} \psi_1(0|x_2)\psi_2(x_2|0) + \big[1 - \psi_1(0|x_2)\big]\psi_2(x_2|1) \\ &= \big[\psi_1(0|0) - \psi_1(0|1)\big]\big[\psi_2(0|0) - \psi_2(0|1)\big] \end{aligned}$$

so that either ψ_1 or ψ_2 does not depend on its condition.

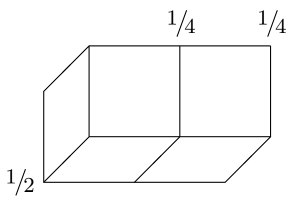
If $P = \psi_1 \cdots \psi_n$ over C_n, $n \geq 3$, the kernel ψ_n° on $X_n \times X_1$ defined by

$$\psi_n^\circ(x_n|x_1) = \sum_{x_2 \in X_2} \cdots \sum_{x_{n-1} \in X_{n-1}} \psi_2(x_2|x_1) \cdots \psi_n(x_n|x_{n-1})$$

suits to the factorization $P^{\{1,n\}} = \psi_1 \psi_n^\circ$ w.r.t. $(\{1,n\}, \{(1,n)(n,1)\})$. Hence, being over C_2, either $\psi_1(x_1|0) = \psi_1(x_1|1)$, $x_1 \in X_1$, and we are ready or necessarily $\psi_n^\circ(x_n|0) = \psi_n^\circ(x_n|1)$, $x_n \in X_n$. In the second case, $\psi_n^\circ(0|0) + \psi_n^\circ(1|1) = 1$ and the product $Q = \delta_1 \psi_2 \cdots \psi_n$ sums to one over X. Here $\delta_1(x_1|x_n) = 1$ or 0 according to $x_1 = x_n$ or not. This argumentation is repeated cyclically doing the next step with the probability distribution Q. If every kernel ψ_v, $v \in V$, depended on its condition then the product $R = \delta_1 \cdots \delta_n$ would sum to one, a contradiction to the obvious $\sum_{x \in X} R(x) = 2$. □

No single restriction on the cardinality can be relaxed in Theorem 4.3.

Example 4.4. Let $X_1 = \{0,1,2\}$, $X_2 = X_3 = \{0,1\}$ and the probability distribution P be given by $P(000) = 1/2$ and $P(111) = P(211) = 1/4$, see Figure 4 left. This probability distribution is RF w.r.t. C_3 via the kernels given in Figure 4 right. Each of the kernels depends on its condition.



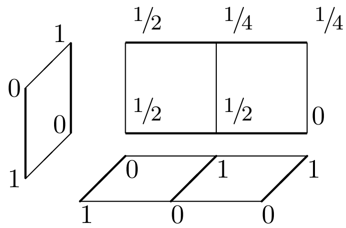


Fig. 4. A projective factorization w.r.t. C_3 without 'trivial' kernels.

Theorem 4.3 can also be verbally reformulated as follows: under the binarity restrictions on all state spaces any of the RF, PF, CF and MF over a circuit implies any of the RF, PF, CF and MF over a path contained in the circuit, cf. Lemma 4.1 and Lemma 2.2. In other words, at least one arrow of the circuit can be erased gaining the acyclicity. It seems also worthwhile to comment the reverse direction under the binarity: it is not difficult to see that RF (= PF = CF = MF) over a path in C_n implies only RF (= PF) over C_n and not MF and CF over C_n, $n \geq 2$, cf. Example 3.1 and Example 3.2, respectively.

Example 4.5. Let $X_1 = X_2 = \{0,1,2\}$, $X_3 = \{0,1\}$ and P be given by $P(x) = 1/4$ for x equal to 000, 101, 211 and 220, see Figure 5. This distribution is MF w.r.t. C_3, but since no conditional independence is present, P is not factorizable w.r.t. a path in C_3. We do not know whether such an example with two two-element state spaces exists.

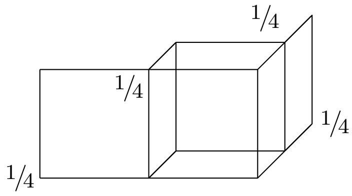


Fig. 5. A distribution that is MF w.r.t. C_3 but not w.r.t. any path.

If instead of RF even MF is assumed in Theorem 4.3 then the assertion can be reformulated equivalently as $P^{\{v,v^-\}} = P^v \cdot P^{v^-}$ for at least one $v \in V$. Here, a special attention deserves the graph C_3 because this pairwise independence takes place even for at least two $v \in V$. In fact, if for example the kernel $\psi_3 = P^{23}/P^2$ is trivial then $P^{23} = P^2 \cdot P^3$ and $P = P^{13}P^{12}/P^1$ and this entails, by Theorem 8.3 of [3] on p. 615, $P = P^{12}P^3$ or $P = P^2P^{13}$. This observation is employed in the last lemma to show another instance of the arrow-erasure.

Lemma 4.6. If a probability distribution P is MF w.r.t. the graph $G = (V, E)$ of the Figure 6 and $|X_v| \le 2$ for all $v \in V$ then there exists $v \in \{1, 2, 3\}$ such that P is MF w.r.t. the acyclic graph $G = \big(V, E - \{(v^-, v)\}\big)$ (at least one arrow of the outer circuit C_3 can be erased).

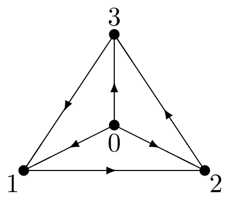


Fig. 6. A graph admitting the arrow erasure.

P r o o f . We can suppose that the marginal P^0 is positive on $X_0 = \{0, 1\}$ otherwise $P = P^0P^{123}$ and one can immediately apply the arrow erasure over C_3. Knowing that

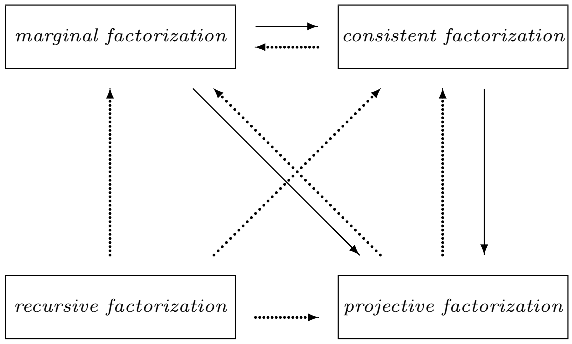
$$P = P^0 \cdot \frac{P^{013}}{P^{03}} \cdot \frac{P^{012}}{P^{01}} \cdot \frac{P^{023}}{P^{02}}$$

we fix $x_0 = 0$ from X_0 and set $Q_{(0)}(x_1x_2x_3) = P(x_0x_1x_2x_3)/P^0(x_0)$ for all x_1, x_2 and x_3. The probability distribution $Q_{(0)}$ on $X_1 \times X_2 \times X_3$ is MF w.r.t. C_3 and thus, for at least two $v \in \{1, 2, 3\}$, $Q_{(0)}^{\{v,v^-\}} = Q_{(0)}^v \cdot Q_{(0)}^{v^-}$. This consideration is repeated with $x_0 = 1$ and its corresponding conditional distribution $Q_{(1)}$ getting again another two vertices $v \in \{1, 2, 3\}$ such that $Q_{(1)}^{\{v,v^-\}} = Q_{(1)}^v \cdot Q_{(1)}^{v^-}$. Hence, we arrive at $P^{\{0,v,v^-\}} = P^{\{0,v\}}P^{\{0,v^-\}}/P^0$ for at least one $v \in \{1, 2, 3\}$. This equality, substituted into the starting MF of P, gives the MF of P w.r.t. the acyclic graph $G = (V, E - \{(v^-, v)\})$.

□

5. CONCLUSION

The following diagram summarizes our knowledge about relations among the four kinds of factorizations.



A full arrow means the implication and a dotted arrow existence of a counterexample to the implication. We do not know whether MF $\Rightarrow$ RF, CF $\Rightarrow$ RF and PF $\Rightarrow$ RF. Note that over C_n or K_n by $n \geq 2$ or over a graph with at most three vertices the three missing arrows were full. To clarify the three open implications, cyclic graphs with at least four vertices must be examined.

Under the binarity restrictions and a fixed kind of factorization, another open and difficult question is to find all graphs that admit the arrow erasure for this kind of factorization.

ACKNOWLEDGEMENT

This work refines and completes results of the dissertation of the second author, cf. [8]. It was done while the first author was Research Fellow of A. von Humboldt Foundation at the University of Bielefeld. The authors offer thanks to M. Studený who helped them in the beginnigs of this work to realize Theorem 4.3 for the case $n = 3$.

(Received November 7, 1997.)

REFERENCES

[1] A. Berman and R. J. Plemmons: Nonnegative Matrices in the Mathematical Sciences. Academic Press, New York – San Francisco – London 1979.
[2] D. R. Cox and N. Wermuth: Multivariate Dependencies. (Monographs on Statistics and Applied Probability 67.) Chapman & Hall, London 1996.
[3] A. P. Dawid: Conditional independence for statistical operations. Ann. Statist. *8* (1980), 598–617.
[4] S. L. Lauritzen: Graphical Models. Clarendon Press, Oxford 1996.
[5] J. T. A. Koster: Gibbs and Markov properties of graphs. Ann. Math. and Artificial Inteligence *21* (1997), 13–26.
[6] J. T. A. Koster: Markov properties of non–recursive causal models. Ann. Statist. *24* (1996), 2148–2177.

[7] P. Spirtes: Directed cyclic graphical representation of feedback models. In: Proceedings of the Eleventh Conference on Uncertainty and Artificial Inteligence (P. Besnard and S. Hanks, eds.), Morgan Kaufman Publ. Inc., San Mateo 1995.

[8] B. Strohmeier: Cyclical Causal Networks and Knowledge Integration. Ph.D. Dissertation. Fakultät für Wirtschaftswissenschaften, Universität Bielefeld 1996.

[9] I. Vajda: Theory of Statistical Inference and Information. Kluwer Academic Publishers, Dordrecht – Boston – London 1989.

[10] J. Whittaker: Graphical Models in Applied Multivariate Statistics. J. Wiley, New York 1990.

František Matúš and Bernhard Strohmeier, Statistik und Informatik, Universität Bielefeld, Postfach 100131, 33 501 Bielefeld. Germany.
e-mail: matus@utia.cas.cz, bstrohmeier@wiwi.uni-bielefeld.de