PRIMARY DECOMPOSITION FOR THE INTERSECTION AXIOM

ALEX FINK

1. Introduction and background

Consider the discrete conditional independence model $\mathcal{M}$ given by

$$\{X_1 \perp\!\!\!\perp X_2 \mid X_3, X_1 \perp\!\!\!\perp X_3 \mid X_2\}.$$

The intersection axiom for conditional independence can be applied with the statements of $\mathcal{M}$ as premises to derive the conclusion $X_1 \perp\!\!\!\perp (X_2, X_3)$. But the independence axiom only holds in general when X is in the interior of the probability simplex, and it's a natural question to ask what can be inferred about X when it may lie on the boundary, that is, what the primary decomposition of $I_{\mathcal{M}}$ is.

A conjecture of Dustin Cartwright and Alexander Engström recorded in [2, p. 152], our Theorem 2.1, characterises the minimal primary components of $\mathcal{M}$ for discrete distributions at the set-theoretic level, in terms of subgraphs of a complete bipartite graph. We state and prove this conjecture in Section 2. Then in Section 3 we discuss the ideal-theoretic question.

2. The set-theoretic conjecture

Let $K_{p,q}$ be the complete bipartite graph with bipartitioned vertex set $[p] \amalg [q]$. The following theorem was the conjecture of Cartwright and Engström, essentially as it appeared in the original source.

Theorem 2.1 (Cartwright-Engström)**.** *The minimal primes of the ideal $I_{\mathcal{M}}$ correspond to the subgraphs of K_{r_2,r_3} that have the same vertex set $[r_2] \amalg [r_3]$ and that have all connected components isomorphic to some complete bipartite graph $K_{p,q}$ with $p, q \geq 1$.*

Call the subgraphs of K_{r_2,r_3} of the form described in the theorem *admissible*. Then, restating,

$$V(I_{\mathcal{M}}) = \bigcup_G V(P_G) \tag{1}$$

as sets, where the union is over admissible graphs G. In particular, the value of r_1 is irrelevant to the combinatorial nature of the primary decomposition.

Let p_{ijk} be the unknown probability $P(X_1 = i, X_2 = j, X_3 = k)$ in a distribution from the model $\mathcal{M}$. Given a subgraph G with edge set $E(G)$, the prime to which it corresponds is $P_G = P_G^{(0)} + P_G^{(1)}$ where

$$P_G^{(0)} = (p_{ijk} : (j,k) \notin E(G)),$$
$$P_G^{(1)} = (p_{i_1j_1k_1}p_{i_1j_2k_2} - p_{i_1j_2k_2}p_{i_2j_1k_1} : (j_\alpha, k_\beta) \in E(G)).$$

For later we'll also want to refer to the individual summands P_G^C of $P_G^{(1)}$, including only the generators $\{p_{ijk} : (j,k) \in C\}$ arising from edges in the component C. If G is admissible, the ideal P_G is prime since it's the sum of a collection of ideals generated in disjoint subsets of the unknowns p_{ijk} each of which is prime: for each connected component $C \subseteq G$ and fixed i, the generators of P_G^C are the 2×2 determinantal ideal of $(p_{ijk})_{(j,k)\in C}$, and all the other variables are themselves generators, appearing in $P_G^{(0)}$.

To give some combinatorial intuition for this, suppose $(p_{ijk}) \in V(P_G)$. Look at the 3-tensor (p_{ijk}) "head-on" with respect to the (j,k) face: that is, think of it as a format $r_2 \times r_3$ table whose entries are vectors $(p_{\cdot jk})$ of format r_1.

The components of G determine subtables of this table. Suppose one of these subtables has format $s_2 \times s_3$. Collapse it into an $r_1 \times (s_2 s_3)$ matrix. Then the conditions $(p_{ij_1k_1}p_{ij_2k_2} - p_{ij_1k_2}p_{ij_2k_1}) \subseteq P_G^{(1)}$ says that the $2 \times r_1$ matrix obtained by setting the vectors at (j_1,k_1) and (j_2,k_2) side by side has rank ≤ 1. Therefore all nonzero vectors in our subtable are equal up to possible scalar multiplication. Entries outside of any subtable must be the zero vector, by the vanishing of $P_G^{(0)}$.

Thus we see that, for G and G' distinct admissible graphs and $r_1 \geq 2$, P_G does not contain $P_{G'}$. That is, the decomposition asserted in Theorem 2.1 is irredundant. Indeed, either G contains an edge (j,k) that G' doesn't, in which case the vector $(p_{\cdot jk})$ is zero on $V(G')$ but generically nonzero on $V(G)$, or $G \subseteq G'$ but two edges $(j,k), (j',k')$ in different components of G are in the same component of G', in which case the vectors $(p_{\cdot jk})$ and $(p_{\cdot j'k'})$ are linearly dependent on $V(G')$ but generically linearly independent on $V(G)$.

The ideas of the proof of Theorem 2.1 were anticipated in part 4 of the problem stated in [2, §6.6], which was framed for the prime corresponding to the subgraph G, the case where the conclusion of the intersection axiom is valid.

Proof of Theorem 2.1. The $\supseteq$ containment of (1) is direct from the definition of P_G: any format $r_1 \times r_2$ slice through (p_{ijk}), say fixing $r_3 = k$, is rank ≤ 1, since the submatrix on columns j for which $(j,k) \in E(G)$ is rank 1 and the other columns are zero.

So we prove the $\subseteq$ containment. Let $(p_{ijk}) \in V(I_{\mathcal{M}})$. Construct the bipartite graph G' on the bipartitioned vertex set $[r_2] \amalg [r_3]$ with edge set

$$\{(j,k) : p_{ijk} \neq 0 \text{ for some } i\}.$$

If $(p_{ijk}) \in \Delta_{r_1r_2r_3-1}$ lies in the closed probability simplex, then this is the set of (j,k) for which the marginal probability p_{+jk} is nonzero.

Consider two edges of G' which share a vertex, assume for now a vertex in the first partition, so that they may be written (j,k) and (j',k) for $j,j' \in [r_2], k \in [r_3]$. Now (p_{ijk}) satisfies the conditional independence statement $X_1 \perp\!\!\!\perp X_2 \mid X_3$, so that the matrix

$$\begin{pmatrix} p_{1jk} & \cdots & p_{r_1jk} \\ p_{1j'k} & \cdots & p_{r_1j'k} \end{pmatrix}$$

has rank ≤ 1. Since neither of its rows is 0, each row is a nonzero scalar multiple of the other. If our two edges had shared a vertex in the second partition, the same argument would go through using the statement $X_1 \perp\!\!\!\perp X_3 \mid X_2$. Now if (j,k) and (j',k') are two edges of a single connected component of G', iterating this argument along a path between them shows that the vectors $(p_{ijk})_i$ and $(p_{ij'k'})_i$ are nonzero scalar multiples of one another.

Let G be obtained from G' by first completing each connected component to a complete bipartite graph on the same set of vertices (giving a graph $\overline{G'}$), and then connecting each isolated vertex in turn to all vertices of an arbitrary block in the other bipartition. The resulting graph G is admissible. Observe also that edges in different connected components of G' remain in different connected components of G.

To establish that $(p_{ijk}) \in V(P_G)$, we must show that given any two edges (j,k), (j',k') in the same component of G, the matrix

$$\begin{pmatrix} p_{1jk} & \cdots & p_{r_1jk} \\ p_{1j'k'} & \cdots & p_{r_1j'k'} \end{pmatrix} \tag{2}$$

has rank ≤ 1. This is immediate from the fact that no connected components were merged. If either of the edges (j,k) or (j',k') did not occur in G', the corresponding row of (2) is 0; otherwise, both edges belong to the same connected component of G' and we just showed that rank (2) ≤ 1. □

In fact, we've done more than prove the last containment. In making G by adding edges to G', given that we weren't allowed to join two connected components that each contained edges, the only choice we had was where to anchor the isolated vertices.

Corollary 2.2. *The components $V(P_G)$ containing a point (p_{ijk}) are exactly those for which G can be obtained from $\overline{G'}$ by adding edges incident in $\overline{G'}$ to isolated vertices, where $\overline{G'}$ is as in the last proof.*

It is noted in [2, §6.6] that the number $\eta(p,q)$ of admissible graphs G on $[p] \amalg [q]$ is given by the generating function

$$\exp((e^x-1)(e^y-1)) = \sum_{p,q\geq 0} \eta(p,q)\frac{x^p y^q}{p!q!} \tag{3}$$

which in that reference is said to follow from manipulations of Stirling numbers. We can also see (3) as a direct consequence of a bivariate form of the exponential formula for exponential generating functions [3, §5.1], using that

$$(e^x-1)(e^y-1) = \sum_{p,q\geq 1} \frac{x^p y^q}{p!q!}$$

is the egf for complete bipartite graphs with $p,q \geq 1$, the possible connected components of admissible graphs.

3. Ideal-theoretic results

It turns out that $I_{\mathcal{M}}$ is exactly the intersection of the minimal primes found in the previous section.

Theorem 3.1. *The primary decomposition*

$$I_{\mathcal{M}} = \bigcap_G P_G \tag{4}$$

holds, where the union is over admissible graphs G on $[r_2] \amalg [r_3]$. In particular $I_{\mathcal{M}}$ is a radical ideal.

In view of the last section we must only prove radicality. This we do by showing that $I_{\mathcal{M}}$ has a radical initial ideal. That is, the next proposition proves Theorem 3.1.

Proposition 3.2. *Let $\prec$ be any term order on monomials in the p_{ijk}, and let P_G have the primary decomposition* $\mathrm{in}_\prec P_G = \bigcap_{\pi\in\Pi_G} Q_{G,\pi}$*. Then*

$$\mathrm{in}_\prec I_{\mathcal{M}} = \bigcap_{G,\pi} Q_{G,\pi} = \bigcap_{G} \mathrm{in}_\prec P_G,$$

G ranging over admissible graphs and π within $\Pi(G)$. Each $Q_{G,\pi}$ is squarefree, so $\mathrm{in}\, I_{\mathcal{M}}$ *is radical.*

Before embarking on this, we outline a bit of the standard treatment of binomial and toric ideals. [[warning! I don't know the best reference, and this presentation is probably awkward for those who know this stuff.]]. Let I be a binomial ideal in $\mathbb{C}[x_1,\dots,x_s]$. The exponents of the binomials generating I define a lattice in $\mathbb{Z}^s$, and the kernel of a $\mathbb{Z}$-linear map $\phi_I : \mathbb{Z}^m \to \mathbb{Z}^s$ whose image is this lattice provides a multigrading, in terms of the *minimal sufficient statistics*, with respect to which I is homogeneous. For any d, the monomials p^u for $u \in \phi_I^{-1}(d) \cap (\mathbb{Z}_{\geq 0})^s$ span the d-graded part of $\mathbb{C}[x_1,\dots,x_s]$. Define an undirected graph H whose vertices are this fiber and whose edges, the *moves*, are (u,u') whenever $x^u - x^{u'}$ is a monomial multiple of a binomial generator of I. Then

$$I_d = \left\{ \sum_{\phi_I(u)=d} c_u u : \textstyle\sum_{u\in C} c_u = 0 \text{ for each connected component } C \subseteq H \right\}.$$

As the P_G are toric ideals, their primary decompositions are understood, and are associated to regular triangulations of certain polytopes. The ideal of 2×2 minors of a matrix $Y = (y_{ij})$, of which $P_K := P_{K_{r_2 r_3}}$ is a particular case, was treated explicitly by Sturmfels [1]. This same treatment extends to arbitrary graphs G, since these are generated as the sum of such ideals P_G^C in several disjoint sets of variables, plus individual sums of other variables, generating $P_G^{(0)}$. So given a collection of primary decompositions $\mathrm{in}_\prec P_G^C = \bigcap_{\pi\in\Pi^C(G)} Q_{G,\pi}^C$,

$$\mathrm{in}_\prec P^G = \bigcap_{(\pi_i)} \left(\bigoplus_i Q_{G,\pi_i}^{C_i} \right) \oplus P_G^{(0)}.$$

We quote some useful results from that paper's treatment.

Theorem 3.3 (Sturmfels, [1])**.** *Let I be the ideal of 2×2 minors of an $r \times s$ matrix of indeterminates.*

(a) *For any term order $\prec$, $\mathrm{in}_\prec I$ is a squarefree monomial ideal.*
(b) *Any two initial ideals of I have the same Hilbert function.*

The distinct Gröbner bases of I are in one-to-one correspondence with the regular triangulations of the product of simplices $\Delta_{r-1} \otimes \Delta_{s-1}$. We repeat from [1] one especially describable example, corresponding to the so-called staircase triangulation. Suppose our term order is the revlex term order over the lexicographic variable order on subscripts. Call this term order $\prec_{\mathrm{dp}}$. Then the primary components of $\mathrm{in}\, P_G$

are parametrised by the paths π through the format $r \times s$ flattening of the matrix of indeterminates, starting at the upper-left corner, taking only steps right and down, and terminating at the lower left corner. The component $Q_{G,\pi}$ is generated by all $(r-1)(s-1)$ indeterminates not lying on the path π. Alternatively, these primes $Q_{G,\pi}$ are generated by exactly the minimal subsets of the indeterminates x_{ij} which include at least one of $x_{ij'}$ and $x_{i'j}$ whenever $i < i'$ and $j < j'$.

Proof of Proposition 3.2. Write $I = I_{\mathcal{M}}$. By Theorem 3 it's enough to show

$$\operatorname{in}_\prec I = \bigcap_G \operatorname{in}_\prec P_G \tag{5}$$

We will do this in two steps: first we'll show $I \subseteq P_G$ for each G, giving the $\subseteq$ containment of (5); then we'll show an equality of Hilbert functions $H(\operatorname{in} I) = H(\bigcap_G \operatorname{in} P_G)$.

Containment. Let $f \in I$. Immediately, terms of f containing a variable p_{ijk} for $(j,k) \notin E(G)$ are in $P_G^{(0)} \subseteq P_G$, so we may assume f has no such terms.

The minimal sufficient statistics of a monomial p^u are given by the row and column marginals of the format $r_1 \times r_2 r_3$ flattening of the table of exponents (u_{ijk}). This is the content of Proposition 1.2.9 of [2], our model $\mathcal{M}$ being the model of the simplicial complex [1][23].

Let $f \in I_d$, and suppose f has no variables corresponding to nonedges of G. Let C be a connected component of G. Then we have that, on each connected component of the graph of the fiber $\phi^{-1}(d)$, the sum $s_{C,i} = \sum_{j,k \in C} u_{ijk}$ is constant, since each individual move holds it constant: the only types of moves are

$$((i,j,k) + (i',j',k), \quad (i,j',k) + (i',j,k))$$
$$((i,j,k) + (i',j,k'), \quad (i,j,k') + (i',j,k))$$

and neither of these changes the value of $s_{C,i}$, because, by the assumption on f, j, j' respectively k, k' must lie in a single connected component of G. But the $s_{C,i}$ are exactly the minimal sufficient statistics for P_G^C which aren't already minimal sufficient statistics of I. That is, on each span of monomials on which the sufficient statistics of P_G are constant, the coefficients of f sum to zero. Therefore $f \in (P_G)_d$, and we've shown $I \subseteq P_G$.

Hilbert functions. We may compute $H(I)$ instead of $H(\operatorname{in} I)$. On the other side, by (b) of Theorem 3, we're free to choose the term order we use in computing $H(\bigcap_G \operatorname{in} P_G)$, and we will choose $\prec_{\mathrm{dp}}$.

The images of monomials in $\mathbb{C}[p_{ijk}]/I$ are a basis, so we want to count these. The multigrading of $\mathbb{C}[p_{ijk}]$ by minimal sufficient statistics passes to the quotient, and so given any multidegree d, the generators of $\mathbb{C}[p_{ijk}]/I$ are in bijection with the connected components of the graph on the fiber $\phi_I^{-1}(d)$.

Let $u \in (\mathbb{Z}_{\geq 0})^{r_1 r_2 r_3}$, and construct the bipartite graph G' on the bipartitioned vertex set $[r_2] \amalg [r_3]$ with edge set

$$\{(j,k) : u_{ijk} \neq 0 \text{ for some } i\}.$$

Note that G' depends only on $\phi(u)$. For any connected component C of G', our $s_{C,i}$ from above is constant for each i.

We next claim that any two exponent vectors u, u' supported on $i \times E(C)$ such that $s_{C,i}(u) = s_{C,i}(u')$ for all i and $u_{+jk} = u'_{+jk}$ for all j, k are in the same component. We show by induction that there are moves carrying u' to u. Let (j,k) be an edge of C whose removal leaves C connected. If there is no index i such that $u'_{ijk} < u_{ijk}$, then there can't be any index i' with $u'_{i'jk} > u_{i'jk}$ either, since the (j,k) marginals are equal. Otherwise pick such an i, and let j',k' be any indices such that $u'_{ij'k'}$ is positive. There is a path of edges $e_0 = (j',k'), e_1, \ldots, e_l = (j,k)$ such that e_i and e_{i+1} share a vertex for each i, and by performing a succession of moves replacing $(i,e_m) + (i_m, e_{m+1})$ by $(i_m, e_m) + (i, e_{m+1})$, we reach from u' a vector of exponents in which the (i,j,k) entry has increased. By repeating this process for each i we can can reach a vector u'' with $u''_{i,j,k} = u_{i,j,k}$ for each i, and then induction onto the graph $C \setminus (j,k)$ completes the argument.

Hence the components of $\phi^{-1}(d)$ are in bijection with the ways to assign a vector of nonnegative integers $s_{C,i}$ to each component C of G' such that $\sum_C s_{C,i} = s_{G',i}$, where the entries of $s_{G',i}$ are particular components of d. This determines $H(I) = H(\operatorname{in} I)$.

We turn to $H(\bigcap_G \operatorname{in} P_G)$. Choose a multidegree d and let G' be the graph defined above. By the discussion surrounding . a monomial of multidegree d is contained in none of the $\operatorname{in} P_G$ if and only if it's not a multiple of $p_{ij'k'}p_{i'jk}$ for any (j,k) and (j',k') in the same component of G' with $i < i'$ and $(j,k) < (j',k')$ lexicographically.

Consider a subtable of the $r_1 \times r_2 r_3$ flattening of the matrix of exponents u_{ijk}, retaining only the rows corresponding to edges of a single component C of G'. Then we claim that, given the row and column marginals of this flattening, there's a monomial whose flattened exponent table has these marginals and is not a multiple of $p_{ij'k'}p_{i'jk}$ for any $i < i'$ and $(j,k) < (j',k')$ lexicographically. Indeed, this minimum is the least monomial with its marginals with respect to $\prec_{\mathrm{dp}}$. By definition of $\prec_{\mathrm{dp}}$, a monomial with the factor $p_{ij'k'}p_{i'jk}$ is greater than the monomial that results by replacing this factor with $p_{ijk}p_{i'j'k'}$. So in the least monomial $\prod p_{ijk}^{u_{ijk}}$, no such replacements are possible; on the other hand, in any monomial $m' = \prod p_{ijk}^{u'_{ijk}}$ that is not the least, we must have $u'_{i'j'k'} < u_{i'j'k'}$ for the last indices at which these exponents differ; then, since the marginals are preserved, there are indices $i < i'$, $(j,k) < (j',k')$ with $u'_{i'jk} > u_{i'jk} \geq 0$ and $u'_{ij'k'} > u_{ij'k'} \geq 0$. In particular $m' p_{ijk} p_{i'j'k'} / p_{ij'k'} p_{i'jk}$ is a monomial less than m'.

Therefore, the monomials of multidegree d not in $\bigcap_G \operatorname{in} P_G$ are in bijection with the ways to choose the row marginals of each of these tables to achieve the sums dictated by d, that is, the ways to choose a vector of nonnegative integers $s_{C,i}$ for each component C of G' such that $\sum_C s_{C,i} = s_{G',i}$. Thus $H(\bigcap_G \operatorname{in} P_G) = H(\operatorname{in} I)$. □

As an appendix, I include `Singular` code to compute the primary decomposition of $I_{\mathcal{M}}$, which I used for my initial investigations. On my machine this ran relatively quickly up through around $r_1 r_2 r_3 = 24$.

```
LIB "primdec.lib";
int r1=2; int r2=3; int r3=4; // adjust as appropriate
int i;
ring R=0,(p(1..r1)(1..r2)(1..r3)),dp;
matrix M[r1][r2]; matrix N[r1][r3];
ideal I=0;
```

```
for(i=1; i<=r3; i{+}{+}) {
  M = p(1..r1)(1..r2)(i);
  I = I + minor(M,2);
}
for(i=1; i<=r2; i{+}{+}) {
  N = p(1..r1)(i)(1..r3);
  I = I + minor(N,2);
}
primdecGTZ(I);
```

References

[1] Bernd Sturmfels, Gröbner bases of toric varieties, *Tōhoku Math. J.*, **43** (1991), 249–261.

[2] Mathias Drton, Bernd Sturmfels and Seth Sullivant, *Lectures on Algebraic Statistics*, Oberwolfach Seminars vol. 39, Springer, 2009.

[3] R. P. Stanley, *Enumerative Combinatorics* vol. 2, Cambridge Studies in Advanced Mathematics, vol. 62, Cambridge University Press, Cambridge, 1997.