The EM Fixed Point Ideal and the Nonnegative Tensor Rank

Elizabeth Allman, Hector Baños Cervantes, Robin Evans, Serkan Hoșten, Kaie Kubjas, Daniel Lemke, John Rhodes, Piotr Zwiernik

Massachusetts Institute of Technology
Sorbonne Université
Aalto University

July 12, 2018

Nonnegative tensor rank

Rank-one tensor: an outer product of vectors

$$\left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\left(\begin{array}{cc|cc} 0 & 1 & 0 & 1 \\ 0 & 2 & 0 & 2 \end{array}\right) = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

A tensor has nonnegative rank (rank_+) at most r if it can be written as a sum of r nonnegative tensors of rank one:

$$P = P_1 + P_2 + \ldots + P_r,$$

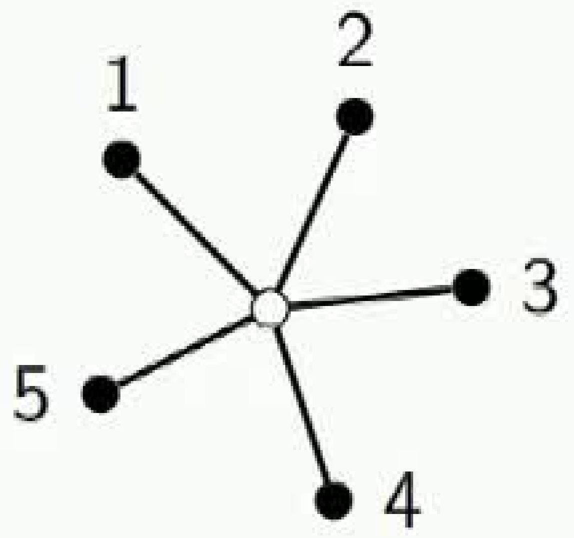
where $P_1, P_2, \ldots, P_r$ are nonnegative tensors of rank 1.

A tensor of nonnegative rank 2:

$$\left(\begin{array}{cc|cc} 1 & 1 & 0 & 1 \\ 0 & 2 & 0 & 2 \end{array}\right) = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) + \left(\begin{array}{cc|cc} 0 & 1 & 0 & 1 \\ 0 & 2 & 0 & 2 \end{array}\right)$$

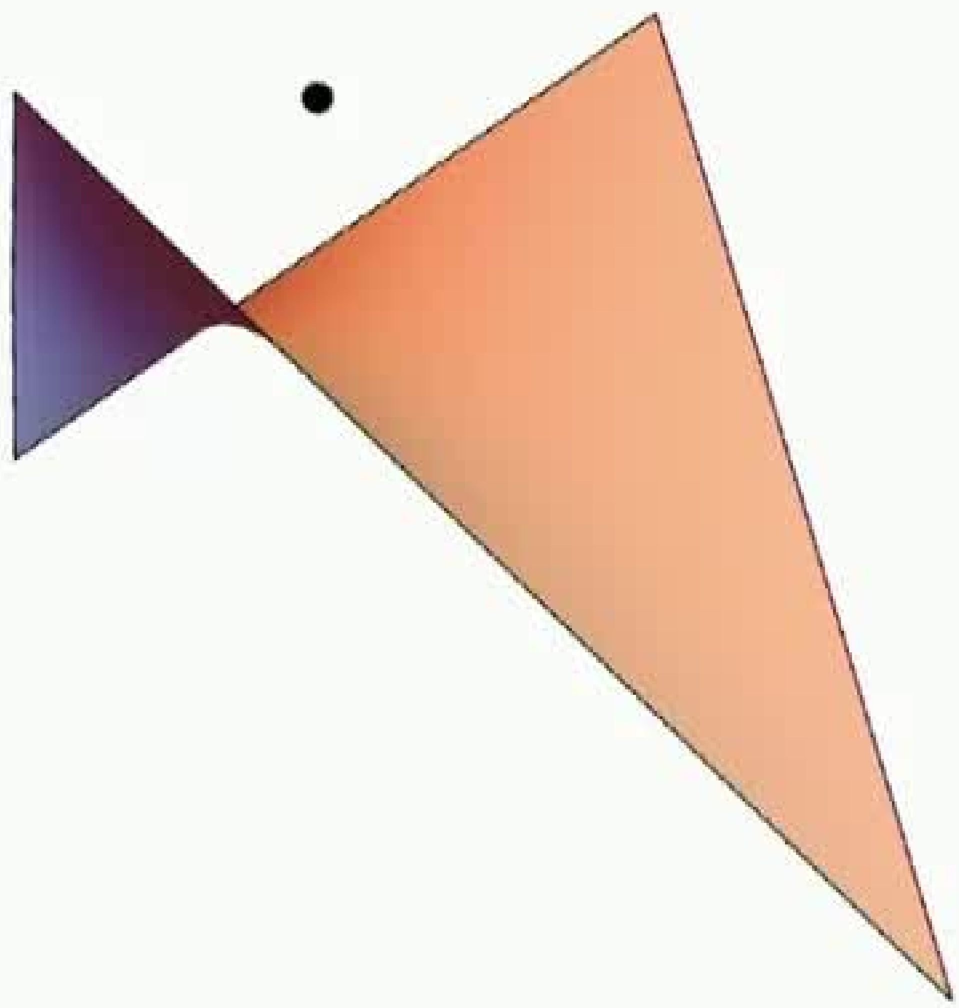
Latent class models

In statistics, the model of normalized tensors of nonnegative rank at most r is the latent class model.



Binary latent class model $\mathcal{M}_n$: $2 \times 2 \times 2 \cdots \times 2$ tensors of nonnegative rank at most two and entries summing to one (of order n)

Maximum likelihood estimation



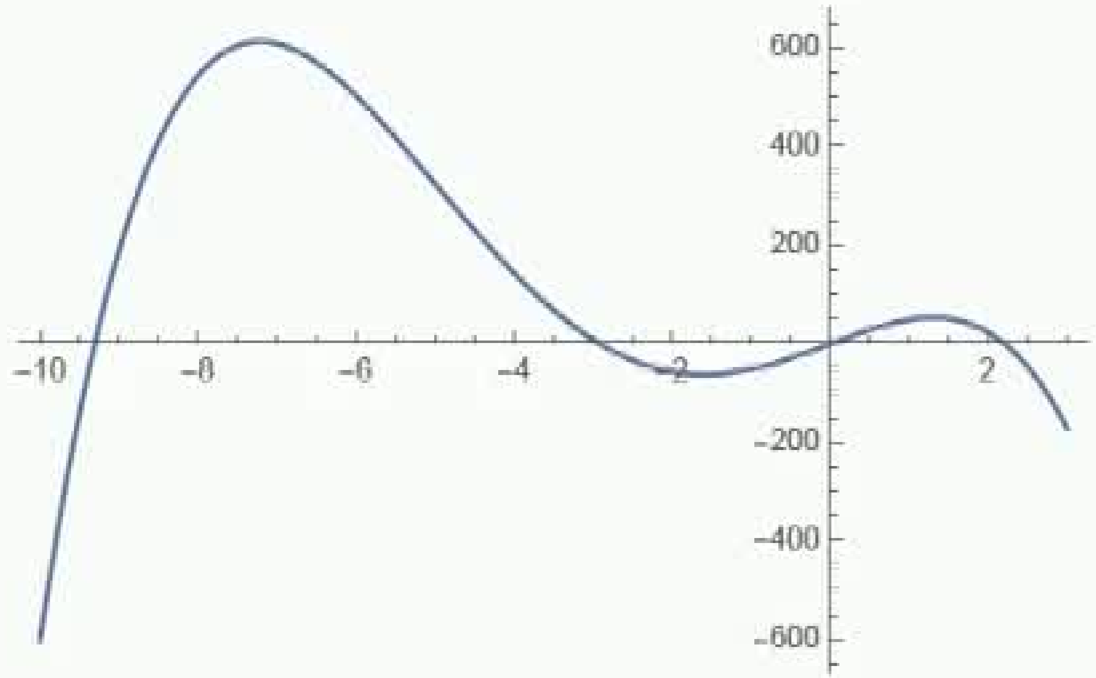
statistical model + data

$\Downarrow$

find the point of the statistical model that best explains the data

Challenges of MLE

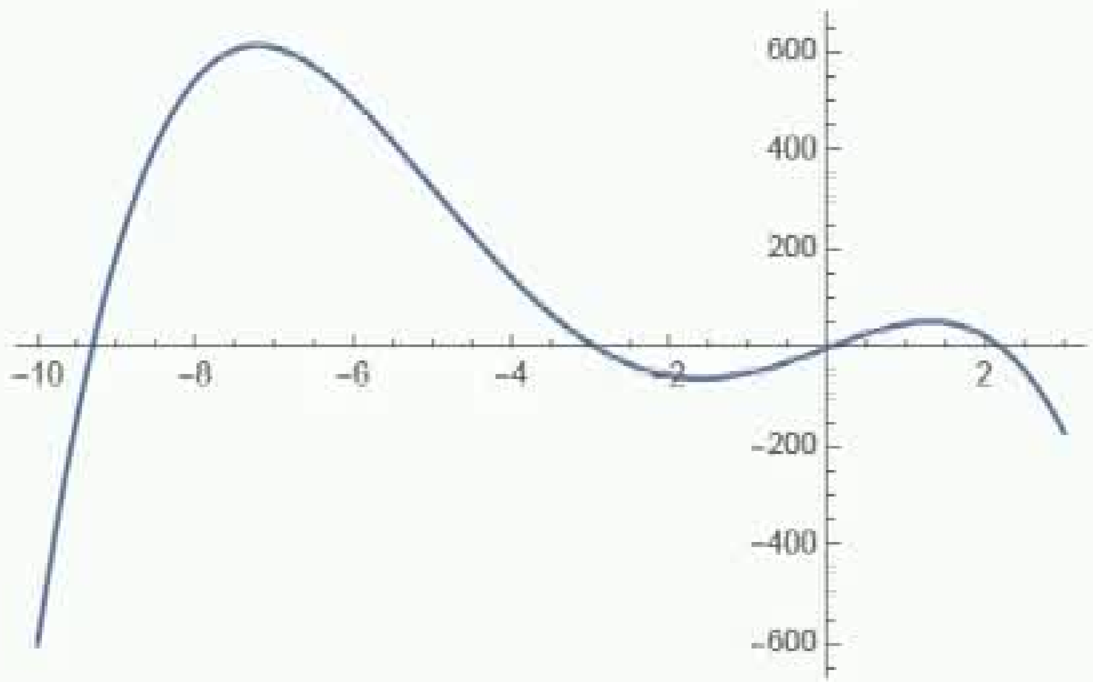
- In general, MLE is a non-convex optimization problem.



- The MLE can lie in the interior or on the boundary
- Sometimes: MLE is an explicit function of the observed data
- Often: MLE is found numerically using optimization methods
- Standard numerical methods find a local maximum

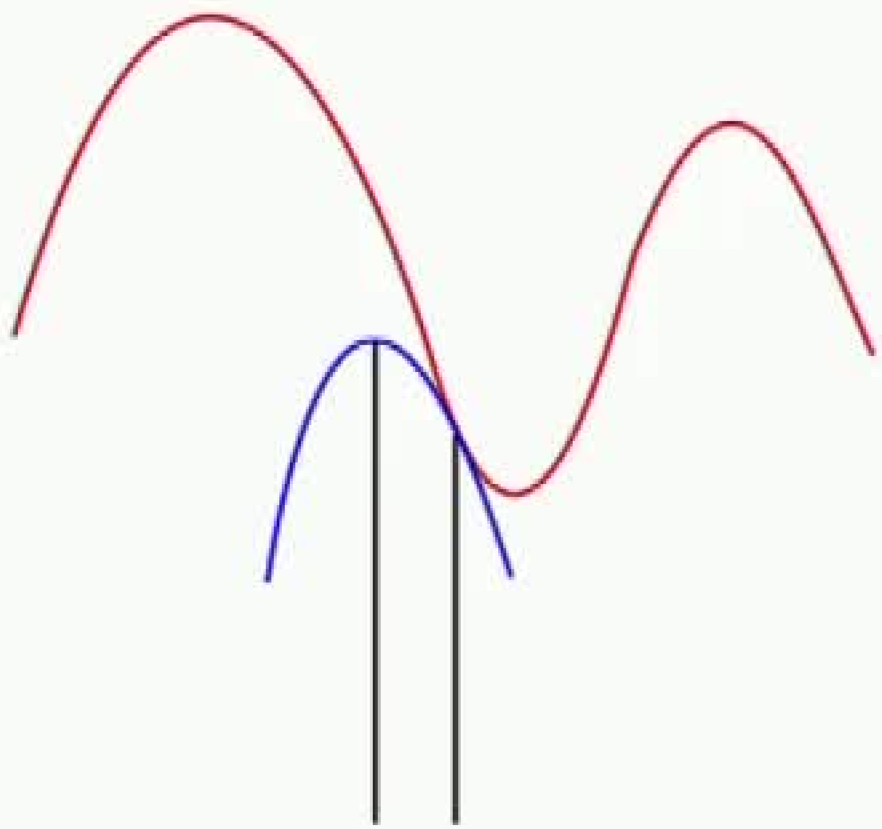
Challenges of MLE

- In general, MLE is a non-convex optimization problem.



- The MLE can lie in the interior or on the boundary
- Sometimes: MLE is an explicit function of the observed data
- Often: MLE is found numerically using optimization methods
- Standard numerical methods find a local maximum

EM algorithm



Initialize parameters.
Run until parameters converge.
E-step: Bound the log-likelihood function below by a simple concave function.
M-step: Maximize the concave function.

Parametrization of the binary latent class model

Let $P = (p_{ijk})$ be a $2 \times 2 \times 2$ tensor. It has nonnegative rank at most 2 if there exist three nonnegative matrices

$$A = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}, B = \begin{pmatrix} b_{00} & b_{01} \\ b_{10} & b_{11} \end{pmatrix}, C = \begin{pmatrix} c_{00} & c_{01} \\ c_{10} & c_{11} \end{pmatrix}$$

such that

$$p_{ijk} = a_{0i} b_{0j} c_{0k} + a_{1i} b_{1j} c_{1k}.$$

The parameter space of the $2 \times 2 \times 2$ tensors of nonnegative rank at most two is $\Theta := \mathbb{R}^{12}_{\geq 0}$ with elements

$$(a_{00}, \ldots, a_{11}, b_{00}, \ldots, b_{11}, c_{00}, \ldots, c_{11}).$$

EM algorithm

Input: Observed data tensor $U \in \mathbb{Z}^{d_1 \times d_2 \times d_3}$.
Output: A proposed maximum $\hat{P} \in \Delta_{d_1 d_2 d_3 - 1}$ of the log-likelihood function ℓ on the model $\mathcal{M}_{d_1 \times d_2 \times d_3, r}$.

Step 0: Select random $(\lambda_1, \ldots, \lambda_r) \in \Delta_{r-1}$, $(a_{i1}, \ldots, a_{id_1}) \in \Delta_{d_1 - 1}$, $(b_{i1}, \ldots, b_{id_2}) \in \Delta_{d_2 - 1}$, and $(c_{i1}, \ldots, c_{id_3}) \in \Delta_{d_3 - 1}$ for $i = 1, \ldots, r$.
Run the following steps until the entries of the $d_1 \times d_2 \times d_3$-tensor $P = (p_{ijk})$ converge.
E-Step: *Estimate the $r \times d_1 \times d_2 \times d_3$-table that represents the expected hidden data:*
Set $v_{lijk} := \frac{\lambda_l a_{li} b_{lj} c_{lk}}{\sum_{l=1}^{r} \lambda_l a_{li} b_{lj} c_{lk}} u_{ijk}$ for $l = 1, \ldots, r$, $i = 1, \ldots, d_1$, $j = 1, \ldots, d_2$, and $k = 1, \ldots, d_3$.
M-Step: *Maximize the likelihood function of the model for the hidden data:*
Set $\lambda_l := \sum_{i=1}^{d_1} \sum_{i=1}^{d_2} \sum_{i=1}^{d_3} v_{ijkl}/u_{+++}$ for $l = 1, ..., r$.
Set $a_{li} := \sum_{j=1}^{d_2} \sum_{k=1}^{d_3} v_{ijkl}/(u_{+++}\lambda_l)$ for $l = 1, \ldots, r$, $i = 1, \ldots, d_1$.
Set $b_{lj} := \sum_{i=1}^{d_1} \sum_{k=1}^{d_3} v_{ijkl}/(u_{+++}\lambda_l)$ for $l = 1, \ldots, r$, $j = 1, \ldots, d_2$.
Set $c_{lk} := \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} v_{ijkl}/(u_{+++}\lambda_l)$ for $l = 1, \ldots, r$, $k = 1, \ldots, d_3$.
Update *the estimate of the joint distribution for our mixture model:*
Set $p_{ijk} := \sum_{l=1}^{r} \lambda_l a_{li} b_{lj} c_{lk}$ for $i = 1, \ldots, d_1$, $j = 1, \ldots, d_2$, $k = 1, \ldots, d_3$.
Return P.

EM fixed point ideal

- An EM fixed point for an observed data tensor U is a parameter vector in Θ which stays fixed after one iteration of the EM algorithm.
- An element of Θ to which the EM-algorithm can converge to for some random starting point is an EM fixed point.
- The minimal set of polynomial equations that all EM fixed points satisfy are called the EM fixed point equations.
- They define the EM fixed point ideal.

EM fixed point ideal

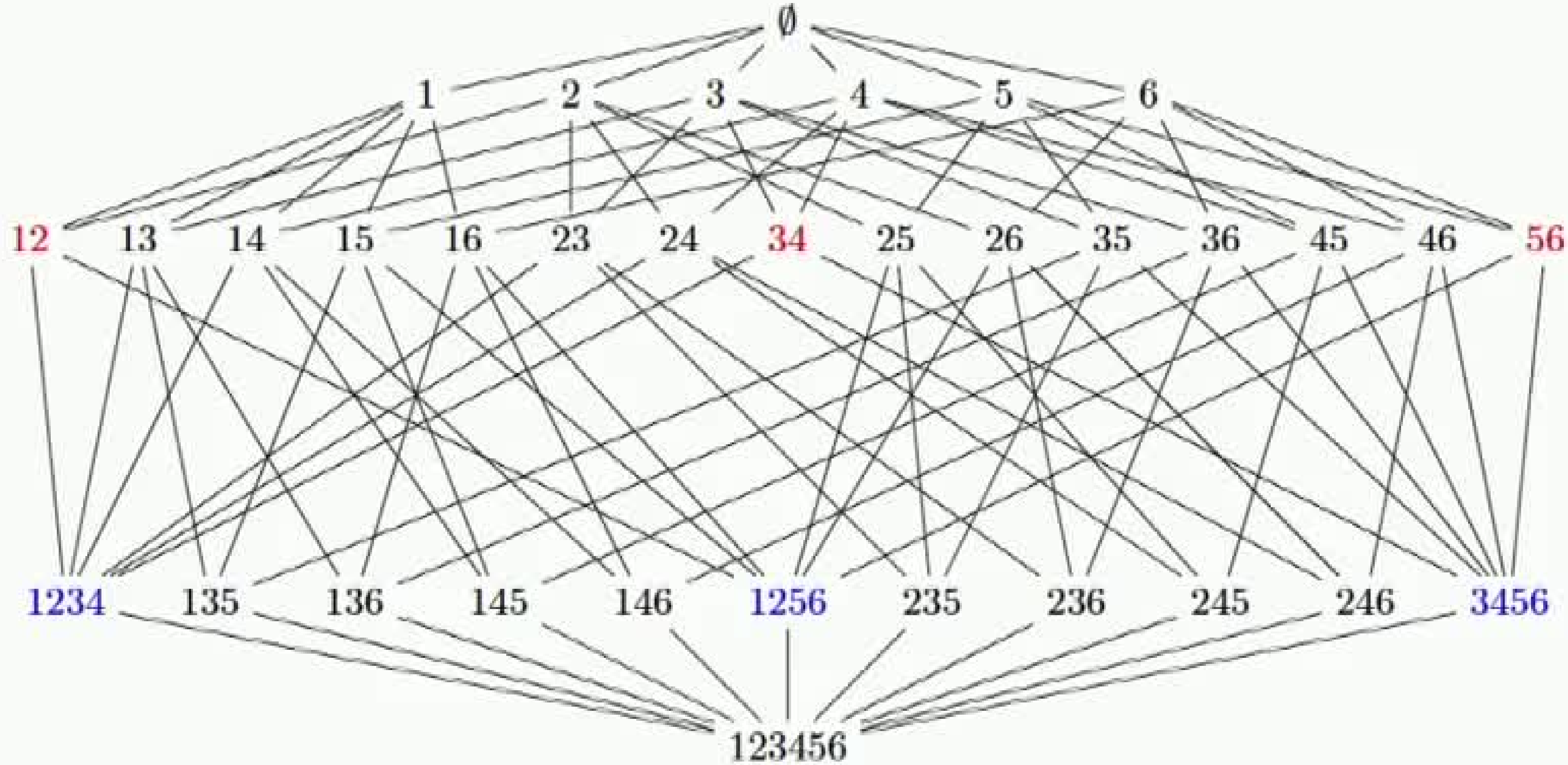
Theorem

The radical of the EM fixed point ideal $\mathcal{F}$ for $\mathcal{M}_3$ has precisely 63 relevant primes consisting of 9 orbital classes.

Table: Minimal primes of EM fixed point ideal for $2 \times 2 \times 2$ tensors of $\text{rank}_+ \leq 2$.

Class S	|S|	a's	b's	c's	deg	codim	rA	rB	rC	orbit
$\{\emptyset\}$	0	0	0	0	60	7	1	1	1	1
	0	0	0	0	48	7	2	2	1	1
	0	0	0	0	48	7	2	1	2	1
	0	0	0	0	48	7	1	2	2	1
	0	0	0	0	1	8	2	2	2	1
$\{a_{00}\}$	1	1	0	0	5	8	2	2	2	12
$\{a_{00}, a_{11}\}$	2	2	0	0	25	8	2	2	2	6
$\{a_{00}, b_{00}\}$	2	1	1	0	11	8	2	2	2	24
$\{a_{00}, b_{00}, c_{00}\}$	3	1	1	1	23	8	2	2	2	16

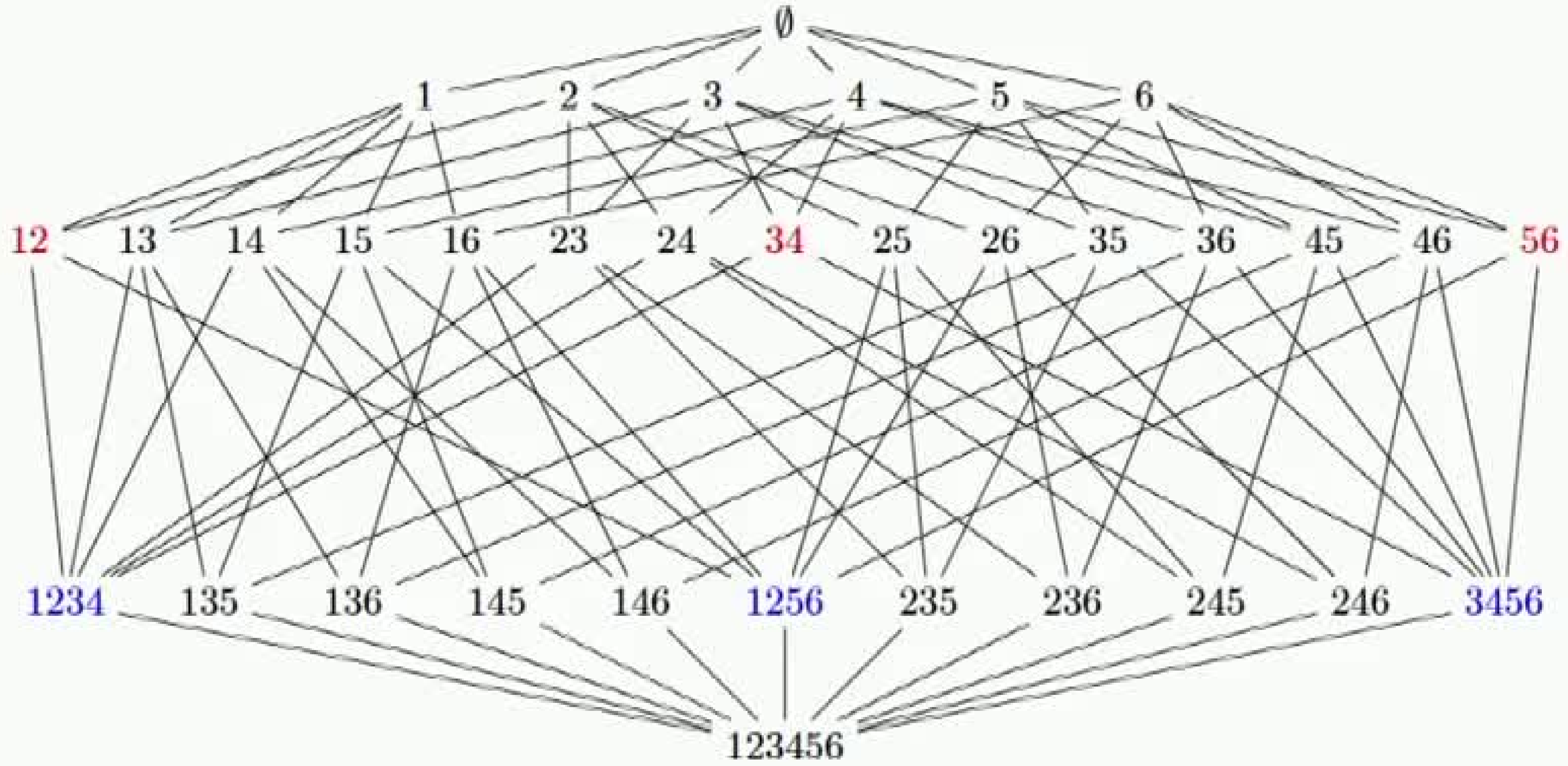
Boundary stratification



Proposition

The dimension of the model $\mathcal{M}_n$ is $2n+1$. The boundary of this semi-algebraic set is defined by $2n$ irreducible components.

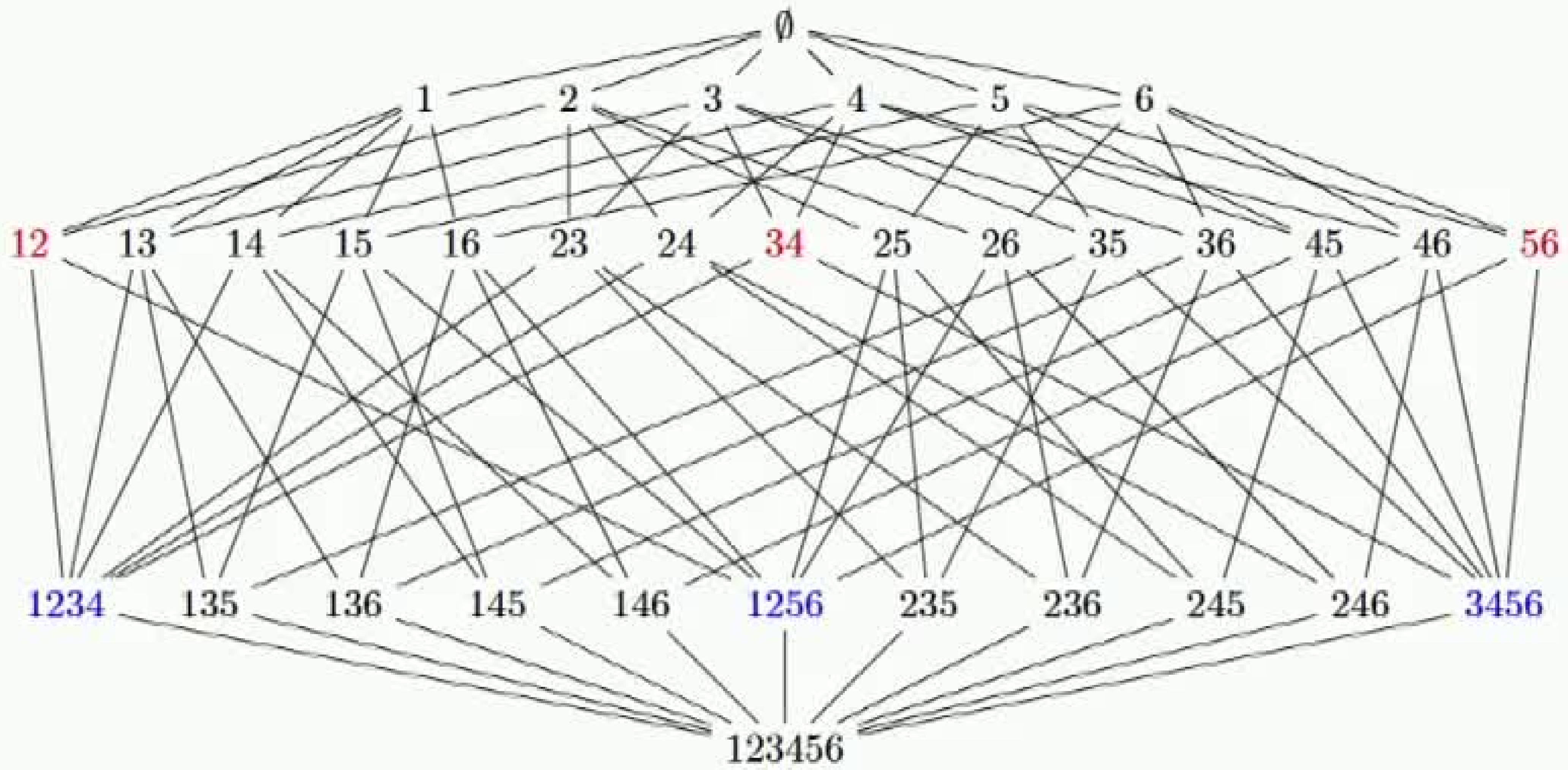
Boundary stratification



Theorem

For $n \leq k \leq 2n+1$, the k-dimensional strata of the nondegenerate part of $\mathcal{M}_n$ are in bijection with the $k-(n+1)$-dimensional faces of the cube C_n, except for $k = 2n-1$ when n additional strata are present, and for $k = n+1$ when $\binom{n}{2}$ additional strata are present.

Class S	|S|	a's	b's	c's	deg	codim	rA	rB	rC	orbit	type
$\{\emptyset\}$	0	0	0	0	60	7	1	1	1	1	3-dimensional
	0	0	0	0	48	7	2	2	1	1	4-dimensional
	0	0	0	0	48	7	2	1	2	1	4-dimensional
	0	0	0	0	48	7	1	2	2	1	4-dimensional
	0	0	0	0	1	8	2	2	2	1	7-dimensional
$\{a_{11}\}$	1	1	0	0	5	8	2	2	2	12	6-dimensional
$\{a_{11}, a_{22}\}$	2	2	0	0	25	8	2	2	2	6	5-dimensional
$\{a_{11}, b_{11}\}$	2	1	1	0	11	8	2	2	2	24	5-dimensional
$\{a_{11}, b_{11}, c_{11}\}$	3	1	1	1	23	8	2	2	2	16	4-dimensional



Example

Consider the minimal prime of the EM fixed point ideal corresponding to $a_{11} = a_{22} = 0$:

$$\begin{aligned} I_1 = \langle & a_{22}, a_{11}, r_{212}r_{221} - r_{211}r_{222}, c_{11}r_{221} + c_{12}r_{222}, b_{11}r_{212} + b_{12}r_{222}, c_{11}r_{211} + c_{12}r_{212}, \\ & b_{11}r_{211} + b_{12}r_{221}, r_{112}r_{121} - r_{111}r_{122}, c_{21}r_{121} + c_{22}r_{122}, b_{21}r_{112} + b_{22}r_{122}, \\ & c_{21}r_{111} + c_{22}r_{112}, b_{21}r_{111} + b_{22}r_{121}\rangle. \end{aligned}$$

We add to the ideal I_1 the ideal of the parametrization map

$$\begin{aligned} I_2 = \langle & -a_{21}b_{21}c_{21} + p_{111}, -a_{21}b_{21}c_{22} + p_{112}, -a_{21}b_{22}c_{21} + p_{121}, -a_{21}b_{22}c_{22} + p_{122}, \\ & -a_{12}b_{11}c_{11} + p_{211}, -a_{12}b_{11}c_{12} + p_{212}, -a_{12}b_{12}c_{11} + p_{221}, -a_{12}b_{12}c_{12} + p_{222}\rangle. \end{aligned}$$

Eliminating parameters $a_{11}, \ldots, c_{22}$ from $I_1 + I_2$, gives the ideal

$$\begin{aligned} J = \langle & p_{212}p_{221} - p_{211}p_{222}, r_{221}p_{221} + r_{222}p_{222}, r_{211}p_{221} + r_{212}p_{222}, r_{212}p_{212} + r_{222}p_{222}, \\ & r_{211}p_{212} + r_{221}p_{222}, r_{221}p_{211} + r_{222}p_{212}, r_{212}p_{211} + r_{222}p_{221}, r_{211}p_{211} - r_{222}p_{222}, \\ & p_{112}p_{121} - p_{111}p_{122}, r_{121}p_{121} + r_{122}p_{122}, r_{111}p_{121} + r_{112}p_{122}, r_{112}p_{112} + r_{122}p_{122}, \\ & r_{111}p_{112} + r_{121}p_{122}, r_{121}p_{111} + r_{122}p_{112}, r_{112}p_{111} + r_{122}p_{121}, r_{111}p_{111} - r_{122}p_{122}, \\ & r_{212}r_{221} - r_{211}r_{222}, r_{112}r_{121} - r_{111}r_{122}\rangle \end{aligned}$$

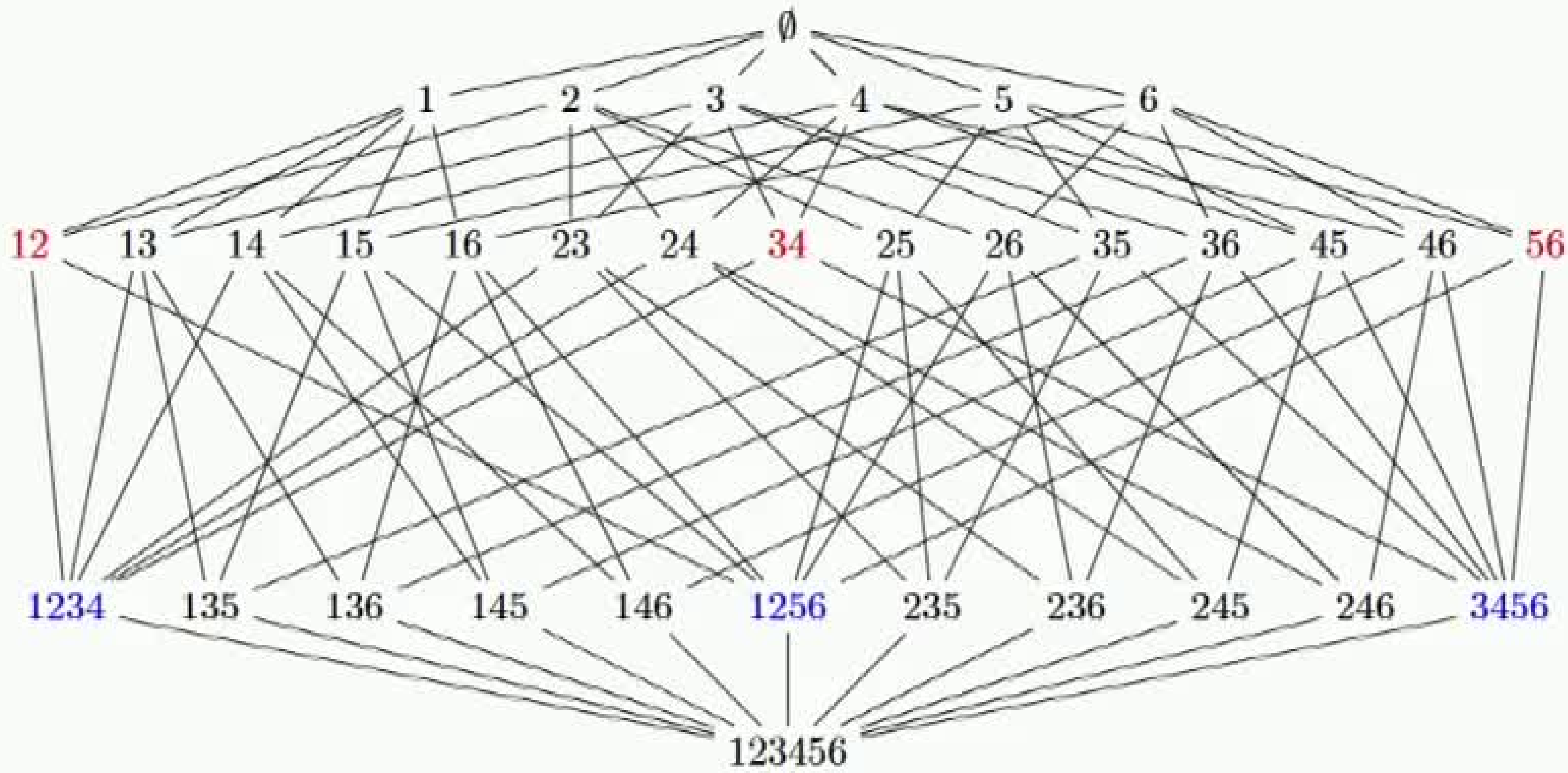
Finally, we substitute to the ideal J the expressions

$$r_{ijk} = u_{+++} - \frac{u_{ijk}}{p_{ijk}}$$

and clear the denominators. To obtain an estimate for p_{111}, we eliminate all other p_{ijk}. This gives the ideal generated by $p_{111}u_{1++}u_{+++} - u_{11+}u_{1+1}$. Hence

$$p_{111} = \frac{u_{11+}u_{1+1}}{u_{1++}u_{+++}}.$$

Class S	|S|	a's	b's	c's	deg	codim	rA	rB	rC	orbit	type
$\{\emptyset\}$	0	0	0	0	60	7	1	1	1	1	3-dimensional
	0	0	0	0	48	7	2	2	1	1	4-dimensional
	0	0	0	0	48	7	2	1	2	1	4-dimensional
	0	0	0	0	48	7	1	2	2	1	4-dimensional
	0	0	0	0	1	8	2	2	2	1	7-dimensional
$\{a_{11}\}$	1	1	0	0	5	8	2	2	2	12	6-dimensional
$\{a_{11}, a_{22}\}$	2	2	0	0	25	8	2	2	2	6	5-dimensional
$\{a_{11}, b_{11}\}$	2	1	1	0	11	8	2	2	2	24	5-dimensional
$\{a_{11}, b_{11}, c_{11}\}$	3	1	1	1	23	8	2	2	2	16	4-dimensional



Example

Consider the minimal prime of the EM fixed point ideal corresponding to $a_{11} = a_{22} = 0$:

$$
\begin{aligned}
I_1 = \langle & a_{22}, a_{11}, r_{212}r_{221} - r_{211}r_{222}, c_{11}r_{221} + c_{12}r_{222}, b_{11}r_{212} + b_{12}r_{222}, c_{11}r_{211} + c_{12}r_{212}, \\
& b_{11}r_{211} + b_{12}r_{221}, r_{112}r_{121} - r_{111}r_{122}, c_{21}r_{121} + c_{22}r_{122}, b_{21}r_{112} + b_{22}r_{122}, \\
& c_{21}r_{111} + c_{22}r_{112}, b_{21}r_{111} + b_{22}r_{121}\rangle.
\end{aligned}
$$

We add to the ideal I_1 the ideal of the parametrization map

$$
\begin{aligned}
I_2 = \langle & -a_{21}b_{21}c_{21} + p_{111}, -a_{21}b_{21}c_{22} + p_{112}, -a_{21}b_{22}c_{21} + p_{121}, -a_{21}b_{22}c_{22} + p_{122}, \\
& -a_{12}b_{11}c_{11} + p_{211}, -a_{12}b_{11}c_{12} + p_{212}, -a_{12}b_{12}c_{11} + p_{221}, -a_{12}b_{12}c_{12} + p_{222}\rangle.
\end{aligned}
$$

Eliminating parameters $a_{11}, \ldots, c_{22}$ from $I_1 + I_2$, gives the ideal

$$
\begin{aligned}
J = \langle & p_{212}p_{221} - p_{211}p_{222}, r_{221}p_{221} + r_{222}p_{222}, r_{211}p_{221} + r_{212}p_{222}, r_{212}p_{212} + r_{222}p_{222}, \\
& r_{211}p_{212} + r_{221}p_{222}, r_{221}p_{211} + r_{222}p_{212}, r_{212}p_{211} + r_{222}p_{221}, r_{211}p_{211} - r_{222}p_{222}, \\
& p_{112}p_{121} - p_{111}p_{122}, r_{121}p_{121} + r_{122}p_{122}, r_{111}p_{121} + r_{112}p_{122}, r_{112}p_{112} + r_{122}p_{122}, \\
& r_{111}p_{112} + r_{121}p_{122}, r_{121}p_{111} + r_{122}p_{112}, r_{112}p_{111} + r_{122}p_{121}, r_{111}p_{111} - r_{122}p_{122}, \\
& r_{212}r_{221} - r_{211}r_{222}, r_{112}r_{121} - r_{111}r_{122}\rangle
\end{aligned}
$$

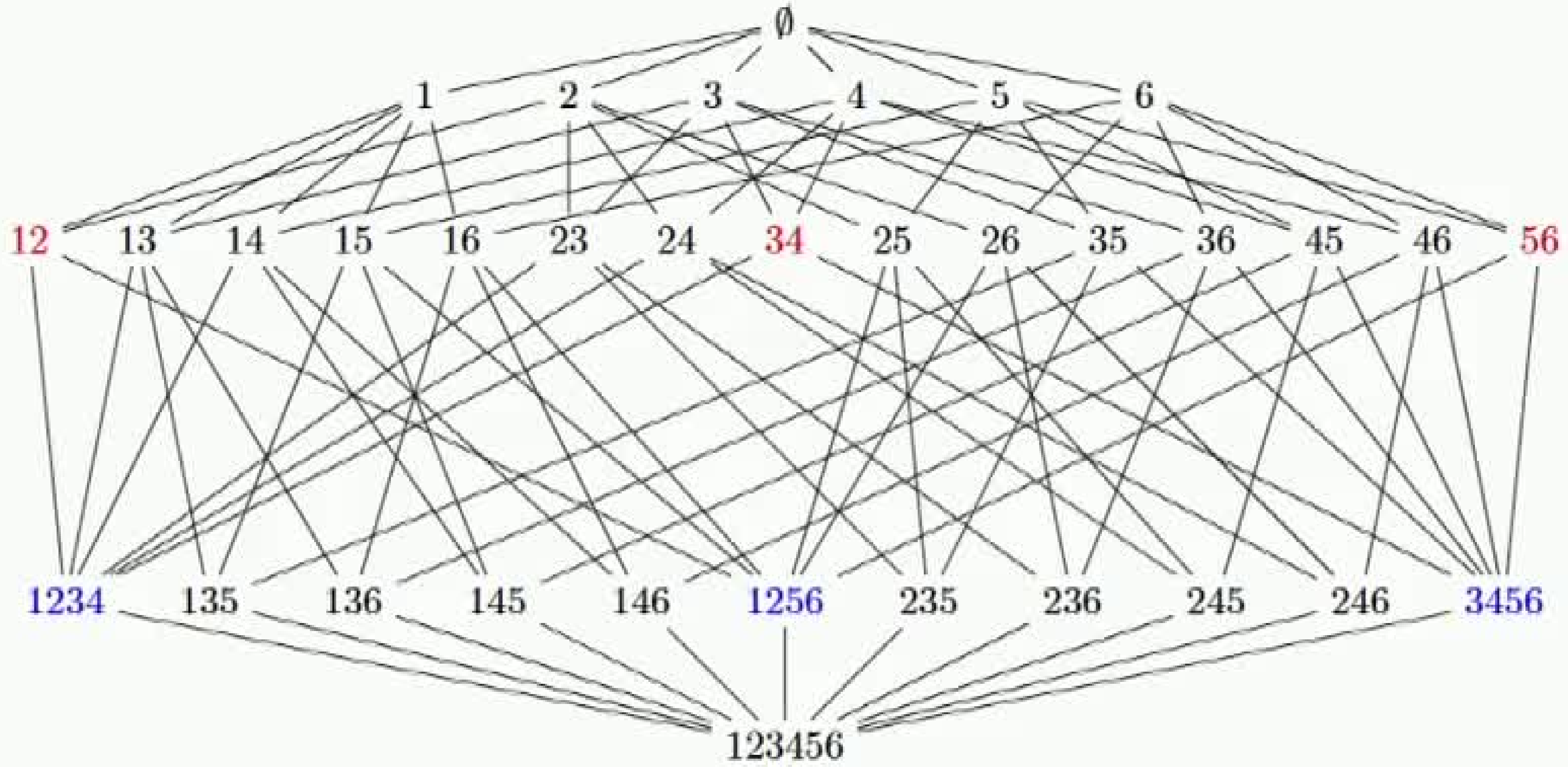
Finally, we substitute to the ideal J the expressions

$$r_{ijk} = u_{+++} - \frac{u_{ijk}}{p_{ijk}}$$

and clear the denominators. To obtain an estimate for p_{111}, we eliminate all other p_{ijk}. This gives the ideal generated by $p_{111}u_{1++}u_{+++} - u_{11+}u_{1+1}$. Hence

$$p_{111} = \frac{u_{11+}u_{1+1}}{u_{1++}u_{+++}}.$$

Class S	|S|	a's	b's	c's	deg	codim	rA	rB	rC	orbit	type
$\{\emptyset\}$	0	0	0	0	60	7	1	1	1	1	3-dimensional
	0	0	0	0	48	7	2	2	1	1	4-dimensional
	0	0	0	0	48	7	2	1	2	1	4-dimensional
	0	0	0	0	48	7	1	2	2	1	4-dimensional
	0	0	0	0	1	8	2	2	2	1	7-dimensional
$\{a_{11}\}$	1	1	0	0	5	8	2	2	2	12	6-dimensional
$\{a_{11}, a_{22}\}$	2	2	0	0	25	8	2	2	2	6	5-dimensional
$\{a_{11}, b_{11}\}$	2	1	1	0	11	8	2	2	2	24	5-dimensional
$\{a_{11}, b_{11}, c_{11}\}$	3	1	1	1	23	8	2	2	2	16	4-dimensional



Example

Consider the minimal prime of the EM fixed point ideal corresponding to $a_{11} = a_{22} = 0$:

$$\begin{aligned} I_1 = \langle & a_{22}, a_{11}, r_{212}r_{221} - r_{211}r_{222}, c_{11}r_{221} + c_{12}r_{222}, b_{11}r_{212} + b_{12}r_{222}, c_{11}r_{211} + c_{12}r_{212}, \\ & b_{11}r_{211} + b_{12}r_{221}, r_{112}r_{121} - r_{111}r_{122}, c_{21}r_{121} + c_{22}r_{122}, b_{21}r_{112} + b_{22}r_{122}, \\ & c_{21}r_{111} + c_{22}r_{112}, b_{21}r_{111} + b_{22}r_{121}\rangle. \end{aligned}$$

We add to the ideal I_1 the ideal of the parametrization map

$$\begin{aligned} I_2 = \langle & -a_{21}b_{21}c_{21} + p_{111}, -a_{21}b_{21}c_{22} + p_{112}, -a_{21}b_{22}c_{21} + p_{121}, -a_{21}b_{22}c_{22} + p_{122}, \\ & -a_{12}b_{11}c_{11} + p_{211}, -a_{12}b_{11}c_{12} + p_{212}, -a_{12}b_{12}c_{11} + p_{221}, -a_{12}b_{12}c_{12} + p_{222}\rangle. \end{aligned}$$

Eliminating parameters $a_{11}, \ldots, c_{22}$ from $I_1 + I_2$, gives the ideal

$$\begin{aligned} J = \langle & p_{212}p_{221} - p_{211}p_{222}, r_{221}p_{221} + r_{222}p_{222}, r_{211}p_{221} + r_{212}p_{222}, r_{212}p_{212} + r_{222}p_{222}, \\ & r_{211}p_{212} + r_{221}p_{222}, r_{221}p_{211} + r_{222}p_{212}, r_{212}p_{211} + r_{222}p_{221}, r_{211}p_{211} - r_{222}p_{222}, \\ & p_{112}p_{121} - p_{111}p_{122}, r_{121}p_{121} + r_{122}p_{122}, r_{111}p_{121} + r_{112}p_{122}, r_{112}p_{112} + r_{122}p_{122}, \\ & r_{111}p_{112} + r_{121}p_{122}, r_{121}p_{111} + r_{122}p_{112}, r_{112}p_{111} + r_{122}p_{121}, r_{111}p_{111} - r_{122}p_{122}, \\ & r_{212}r_{221} - r_{211}r_{222}, r_{112}r_{121} - r_{111}r_{122}\rangle \end{aligned}$$

Finally, we substitute to the ideal J the expressions

$$r_{ijk} = u_{+++} - \frac{u_{ijk}}{p_{ijk}}$$

and clear the denominators. To obtain an estimate for p_{111}, we eliminate all other p_{ijk}. This gives the ideal generated by $p_{111}u_{1++}u_{+++} - u_{11+}u_{1+1}$. Hence

$$p_{111} = \frac{u_{11+}u_{1+1}}{u_{1++}u_{+++}}.$$

EM fixed point ideal

Theorem

The radical of the EM fixed point ideal $\mathcal{F}$ for $\mathcal{M}_{3,3}$ has precisely 317 *relevant primes consisting of* 21 *orbital classes.*

Table: Minimal primes of EM fixed point ideal for $2 \times 2 \times 2$ tensors of $\text{rank}_+ \leq 3$.

Class S	|S|	a's	b's	c's	deg	codim	rA	rB	rC	orbit
$\{\emptyset\}$	0	0	0	0	121	10	1	1	1	1
	0	0	0	0	162	9	1	2	2	1
	0	0	0	0	162	9	2	1	2	1
	0	0	0	0	162	9	2	2	1	1
*	0	0	0	0	38	10	2	2	2	6×1
	0	0	0	0	1	8	2	2	2	1
$\{a_{00}\}$	1	1	0	0	10	10	2	2	2	18
$\{a_{00}, a_{10}\}$	2	2	0	0	5	9	2	2	2	18
$\{a_{00}, b_{00}\}$	2	1	1	0	39	10	2	2	2	36
$\{a_{00}, a_{10}, a_{21}\}$	3	3	0	0	50	11	2	2	2	18
$\{a_{00}, b_{00}, c_{00}\}$	3	1	1	1	60	11	2	2	2	24
$\{a_{00}, a_{10}, b_{00}, b_{10}\}$	4	2	2	0	11	10	2	2	2	36
$\{a_{00}, a_{11}, b_{00}, b_{11}\}$	4	2	2	0	8	11	2	2	2	36
$\{a_{00}, a_{10}, b_{00}, b_{10}, c_{00}, c_{10}\}$	6	2	2	2	23	11	2	2	2	24
$\{a_{00}, a_{10}, b_{00}, b_{11}, c_{00}, c_{11}\}$	6	2	2	2	20	12	2	2	2	72
$\{a_{00}, a_{11}, b_{00}, b_{11}, c_{00}, c_{11}\}$	6	2	2	2	23	12	2	2	2	24

This table contains all the boundary strata of $\mathcal{M}_{3,3}$ from (Seigal,