Multiobjective optimization and the entropy region

Laszlo Csirmaz

Central European University, Budapest

Prague Stochastics, Aug 28, 2014

Outline

Polymatroids

- The **ground set** N is any finite set, $N = \{1, 2, \ldots, N\}$.
- The **rank function** f assigns non-negative values to the subsets $I \subseteq N$, that is, $f : 2^N \to \mathbb{R}^{\geq 0}$.
- $\langle f, N \rangle$ is a **polymatroid** if it satisfies the Shannon inequalities:

$$
\begin{aligned}
&f(\emptyset) = 0, \\
&f(A) \geq f(B) \quad \text{if } A \supseteq B, \\
&f(A) + f(B) \geq f(A \cup B) + f(A \cap B).
\end{aligned}
$$

- $\langle f, N \rangle$ is a **matroid** if $f(A)$ is integer, and $f(A) \leq |A|$.
- $\langle f, N \rangle$ is **entropic** if $f(A) = \boldsymbol{H}(\xi_A)$, where $(\xi_i : i \in N)$ are discrete random variables with some joint distribution.
- Pointwise limit of entropic polymatroids are **almost entropic**.

Regions

The rank function f is a vector indexed by non-empty subsets of N.

- $\boldsymbol{H}_N \subseteq \mathbb{R}^{2^N-1}$ is the region of **polymatroids**.
  – a full-dimensional closed convex pointed cone.
- $\boldsymbol{H}_N^{\text{ent}} \subseteq \boldsymbol{H}_N$ is the **entropy region**.
- $cl(\boldsymbol{H}_N^{\text{ent}})$ is the **pointwise closure** of $\boldsymbol{H}_N^{\text{ent}}$.

Theorem (Zhang–Yeung 1998, Matúš 2007)

- *$cl(\boldsymbol{H}_N^{\text{ent}})$ is a convex full-dimensional cone in $\mathbb{R}^{2^N-1}$.*
- *The interior of $cl(\boldsymbol{H}_N^{\text{ent}})$ is entropic.*
- $\boldsymbol{H}_2^{\text{ent}} = cl(\boldsymbol{H}_2^{\text{ent}}) = \boldsymbol{H}_2$.
- $\boldsymbol{H}_3^{\text{ent}} \neq cl(\boldsymbol{H}_3^{\text{ent}}) = \boldsymbol{H}_3$.
- $\boldsymbol{H}_N^{\text{ent}} \neq cl(\boldsymbol{H}_N^{\text{ent}}) \neq \boldsymbol{H}_N$ *for* $N \geq 4$.
- $cl(\boldsymbol{H}_N^{\text{ent}})$ *is* **not** *polyhedral for* $N \geq 4$. □

The boundary of the entropy region

Definition

$\boldsymbol{H}_N^k \subseteq \boldsymbol{H}_N^{\text{ent}}$ is the subregion where the distribution $(\xi_i : i \in N)$ has alphabet size k.

Facts

H_N^k is closed; $\boldsymbol{H}_N^k \subseteq \boldsymbol{H}_N^{k+1}$; and $\boldsymbol{H}_N^{\text{ent}} = \bigcup_k \boldsymbol{H}_N^k$. □

Research Problems

1. For fixed N, is the convergence $\boldsymbol{H}_N^k \to \boldsymbol{H}_N^{\text{ent}}$ uniform?
2. Give an estimate for the thickness of $\boldsymbol{H}_N^{\text{ent}} - \boldsymbol{H}_N^k$ (in different metrics) as a function of k.
3. Give a description of $cl(\boldsymbol{H}_N^{\text{ent}}) - \boldsymbol{H}_N^{\text{ent}}$ in the case $N = 3$. Where is it fractal-like?

Outline

Searching for new entropy inequalities

Known methods to get new entropy inequalities are:

1. Zhang–Yeung method (1998)
2. Makarychev *et al.* technique (2002)
3. Matúš' polymatroid convolution (2007)
4. Maximum entropy extension (2014)

Equivalence of #1 and #2 for **balanced** inequalities was shown by Tarik Kaced (2013).

Research problem

Show that methods #3 and #4 are actually **stronger** than the other two.

Searching for new entropy inequalities

Known methods to get new entropy inequalities are:

1. Zhang–Yeung method (1998)
2. Makarychev *et al.* technique (2002)
3. Matúš' polymatroid convolution (2007)
4. Maximum entropy extension (2014)

Equivalence of #1 and #2 for **balanced** inequalities was shown by Tarik Kaced (2013).

Research problem

Show that methods #3 and #4 are actually **stronger** than the other two.

We focus on method #1, others raise similar issues.

Zhang–Yeung method

In nutshell

1. Start with a pool of some (at least four) random variables;
2. split the random variables into two sets: $\vec{x}_1$ and $\vec{y}$,
3. make an independent copy $\vec{x}_2$ of $\vec{x}_1$ over $\vec{y}$ to get the new pool of random variables $\langle \vec{x}_1, \vec{x}_2, \vec{y} \rangle$;
4. iterate steps 2 and 3 several times;
5. collect the constraints:
   - Shannon inequalities for the final variable set;
   - equalities among entropy values expressing:
     all conditional independence; identical distribution of $(\vec{x}_1, \vec{y})$ and $(\vec{x}_2, \vec{y})$; symmetry of $\vec{x}_1$ and $\vec{x}_2$ over $\vec{y}$;
6. extract all consequences for the original variables.

Zhang–Yeung method

In nutshell

1. Start with a pool of some (at least four) random variables;
2. split the random variables into two sets: $\vec{x}_1$ and $\vec{y}$,
3. make an independent copy $\vec{x}_2$ of $\vec{x}_1$ over $\vec{y}$ to get the new pool of random variables $\langle \vec{x}_1, \vec{x}_2, \vec{y} \rangle$;
4. iterate steps 2 and 3 several times;
5. collect the constraints:
   – Shannon inequalities for the final variable set;
   – equalities among entropy values expressing:
     all conditional independence; identical distribution of $(\vec{x}_1, \vec{y})$ and $(\vec{x}_2, \vec{y})$; symmetry of $\vec{x}_1$ and $\vec{x}_2$ over $\vec{y}$;
6. extract all consequences for the original variables.

Numerically intractable even for three full iterations.

Zhang–Yeung method

Remedy (Dougherty *et al*)

- discard some of the copied variables in $\vec{x}_2$; and/or
- glue together some variables in $\vec{x}_2$.

Zhang–Yeung method

Remedy (Dougherty *et al*)

- discard some of the copied variables in $\vec{x}_2$; and/or
- glue together some variables in $\vec{x}_2$.

Example **copy string** with three iterations, **initial** random variables `abcd` and **auxiliary** variables `rstuv`:

```
rs=cd:ab; tu=cr:ab; v=(cr):abtu
```

The set of **constraints** is composed of

- all Shannon inequalities,
- all conditional independence, and
- equality arising from identical distributions and symmetry,

written for entropies of the subsets of the initial and auxiliary variables (`abcd`+`rstuv`).

Zhang–Yeung method – geometrical view

Given a copy string for initial variables `abcd`, we use the notation

- $\mathbf{x} \in \mathbb{R}^p$ for the entropies of subsets of `abcd` ($p = 15$);
- $\mathbf{y} \in \mathbb{R}^m$ for the vector of all other entropies;
- $\mathcal{M}(\mathbf{x}, \mathbf{y})$ for the collection of constraints.

$\mathcal{M}(\mathbf{x}, \mathbf{y})$ is **linear** and **homogeneous**, thus can be written as

$$P\mathbf{x} + M\mathbf{y} \geq 0$$

for some $p \times n$ and $m \times n$ matrices P and M determined by the copy string.

- $\mathcal{P} = \{ (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{p+m} : \mathbf{x} \geq 0,\ \mathbf{y} \geq 0,\ P\mathbf{x} + M\mathbf{y} \geq 0 \}$ is the **feasible region**, a convex pointed polyhedral cone;
- $\mathcal{Q} = \{\mathbf{x} \in \mathbb{R}^p : \text{ for some } \mathbf{y} \in \mathbb{R}^m, (\mathbf{x}, \mathbf{y}) \in \mathcal{P}\}$ is the **projection of** $\mathcal{P}$, a convex, pointed polyhedral cone.

$\Rightarrow$

Geometrical view

$\Rightarrow$

- $\mathcal{P} = \{ (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{p+m} : \mathbf{x} \ge 0,\ \mathbf{y} \ge 0,\ P\mathbf{x} + M\mathbf{y} \ge 0 \}$,
- $\mathcal{Q} = \{\mathbf{x} \in \mathbb{R}^p :$ for some $\mathbf{y} \in \mathbb{R}^m$, $(\mathbf{x}, \mathbf{y}) \in \mathcal{P}\}$.

Linear consequences of $P\mathbf{x} + M\mathbf{y} \ge 0$ are the non-negative linear combinations of the rows of (P, M). Such an inequality bounds $\mathcal{Q}$ iff in it all $\mathbf{y}$ coordinates are zero. Thus the collection of linear inequalities bounding $\mathcal{Q}$ – the **dual cone of** $\mathcal{Q}$ – is

- $\mathcal{Q}^\circ = \{P^T\mathbf{v} \in \mathbb{R}^p : \mathbf{v} \in \mathbb{R}^n,\ \mathbf{v} \ge 0,\ M^T\mathbf{v} = 0\}$.

Observations

a) If $\mathbf{x} \in \mathbb{R}^p$ is entropic, then for some $\mathbf{y} \in \mathbb{R}^m$, $(\mathbf{x}, \mathbf{y}) \in \mathcal{P}$.
Therefore the entropy region is contained in the projection $\mathcal{Q}$.
b) The "strongest" entropy inequalities which can be extracted from a copy string are the extremal rays of $\mathcal{Q}^\circ$. □

Creating new information inequalities

In theory it is as easy as . . .

1. Choose your favorite **copy string**.
2. **Generate** the matrices (P, M) describing the linear homogeneous constraints arising from your copy string.
3. **Compute** the extremal rays of $\mathcal{Q}^\circ$ using your favorite computer algebra package.

Creating new information inequalities

In theory it is as easy as . . .

1. Choose your favorite **copy string**.
2. **Generate** the matrices (P, M) describing the linear homogeneous constraints arising from your copy string.
3. **Compute** the extremal rays of $\mathcal{Q}^\circ$ using your favorite computer algebra package.

In practice there are annoying nuisances . . .

1. When things get interesting, M becomes really large
   (over 28000 Shannon inequalities just for 4+7 variables).
2. Even if the size is not a problem, M is highly degenerate
   (hated by all packages).
3. The computational problem is numerically unstable
   (and integer arithmetic takes ages).

Improving the performance

Use what is known about $\boldsymbol{H}_4$

Where to look:

[1] Frantisek Matúš and Milan Studený, Conditional independencies among four random variables I, *Combinatorics, Probability and Computing*, no 4, (1995) pp. 269-278.

Outline

Entropy expressions

Fix four random variables as a, b, c, d.
For any subset J of $abcd$, J also denotes its entropy, $\boldsymbol{H}(J)$.

Definition

For any permutation of the variables a, b, c, d we define

- $(a, b) \stackrel{\text{def}}{=} a + b - ab$; ⇐ mutual info
- $(a, b \,|\, c) \stackrel{\text{def}}{=} ac + bc - abc - c$; ⇐ cond. mutual info
- $(a, b \,|\, cd) \stackrel{\text{def}}{=} acd + bcd - abcd - cd$;
- $(a \,|\, bcd) \stackrel{\text{def}}{=} abcd - bcd$; ⇐ cond. entropy
- $[abcd\,] \stackrel{\text{def}}{=} -(a, b) + (a, b \,|\, c) + (a, b \,|\, d) + (c, d)$. ⇐ Ingleton

The **Ingleton** expression is symmetric in ab and cd:

$$[abcd\,] = [\overset{\curvearrowright}{ba}cd\,] = [ab\overset{\curvearrowright}{dc}\,] = [\overset{\curvearrowright}{ba}\overset{\curvearrowright}{dc}\,].$$

Why Ingleton is so important

Definition

$\square \subset cl(\boldsymbol{H}_4^{\text{ent}})$ where **all six** Ingleton expressions are ≥ 0;

$\square_{ab} \subset cl(\boldsymbol{H}_4^{\text{ent}})$ where $[abcd] \leq 0$, i.e., this Ingleton is violated;
$\square_{ac} \subset cl(\boldsymbol{H}_4^{\text{ent}})$ where $[acbd] \leq 0$; etc.

Theorem (Matus – Studeny, 1995)

- $cl(\boldsymbol{H}_4^{\text{ent}}) = \square \cup \square_{ab} \cup \cdots \cup \square_{cd}$.
- *Any two of $\square$, $\square_{ab}$, $\ldots$, $\square_{cd}$ have disjoint interior; common points are on the boundary of $\square$.*
- *$\square$ is a full dimensional closed polyhedral cone, bounded by the six Ingleton, and certain other Shannon facets.*
- *Internal points and vertices of $\square$ are linearly representable.*
- *$\square_{ab}$, $\ldots$, $\square_{cd}$ are isomorphic; isomorphisms are provided by permutations of abcd.* □

If . . .

If we know $\square_{ab}$, then

we know everything.*

*at least about $cl(\boldsymbol{H}_4^{\text{ent}})$.

The case of five variables

Research problem

Give a similar decomposition of the 31-dimensional cone $\boldsymbol{H}_5$.

- $\boldsymbol{H}_5$ has a 120-fold symmetry;
- it has 117978 vertices [2];
- the vertices fall into 1319 equivalence classes[2] (into 15 equivalence classes in case of four variables);
- the linearly representable core of $\boldsymbol{H}_5$ is known precisely[3].

[2] M. Studený, R. R. Bouckaert, T. Kočka:
Extreme supermodular set functions over five variables

[3] R. Dougherty, C. Freiling, K. Zeger:
Linear rank inequalities on five or more variables

Natural coordinates

$\square_{ab} \subset \boldsymbol{H}_4$ is contained in the simplicial cone determined by these facets (proved in [1]):

1.	$[abcd]$,
2., 3.	$(a,b\,|\,c)$, $(a,b\,|\,d)$,
4–7.	$(a,c\,|\,b)$, $(b,c\,|\,a)$, $(a,d\,|\,b)$, $(b,d\,|\,a)$,
8., 9.	$(c,d\,|\,a)$, $(c,d\,|\,b)$,
10.	(c,d),
11.	$(a,b\,|\,cd)$,
12–15.	$(a\,|\,bcd)$, $(b\,|\,acd)$, $(c\,|\,adb)$, $(d\,|\,abc)$,

Natural coordinates

Use the facet equations as the coordinates for the entropies.

Entropy inequalities in natural coordinates

There are **six** natural coordinate systems corresponding to the six non-equivalent Ingleton expressions. Each entropy inequality can be written using any of the natural coordinates.

General form of a linear inequality

$$\lambda_1[abcd] + \lambda_2(a,b|c) + \lambda_3(a,b|d) + \cdots + \lambda_{15}(d|abc) \geq 0. \quad (1)$$

Claim

1. $\lambda_2 \geq 0$, $\lambda_3 \geq 0$, ..., $\lambda_{15} \geq 0$.
2. *The Ingleton coeff is* > 0 *in some natural coordinate system.*
3. *Can be strenghtened by setting* $\lambda_{12}, \lambda_{13}, \lambda_{14}, \lambda_{15}$ *to zero.*

Proof.

1. (1) must hold for the entropic vector $(0,\dots,0,1,0,\dots)$.
2. If not, then all points satisfying (1) are in $\mathcal{I}$.
3. Equivalent to balancing (1). □

Outline

What we've had

$\Rightarrow$

- $\mathbf{x} \in \mathbb{R}^p$ are the entropies of abcd ,
- $\mathbf{y} \in \mathbb{R}^m$ are all other entropies,
- the constraints are given by the matrices (P, M),
- $\mathcal{P} = \{ (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{p+m} : \mathbf{x} \geq 0, \mathbf{y} \geq 0, P\mathbf{x} + M\mathbf{y} \geq 0 \}$,
- $\mathcal{Q} = \{\mathbf{x} \in \mathbb{R}^p : \text{ for some } \mathbf{y} \in \mathbb{R}^m, (\mathbf{x}, \mathbf{y}) \in \mathcal{P}\}$.
- $\mathcal{Q}^\circ = \{P^T\mathbf{v} \in \mathbb{R}^p : \mathbf{v} \in \mathbb{R}^n, \mathbf{v} \geq 0, M^T\mathbf{v} = 0\}$,

What we've had, and what we've got

⇒

- $\mathbf{x} \in \mathbb{R}^p$ are the entropies of abcd **in natural coordinates**,
- $\mathbf{y} \in \mathbb{R}^m$ are all other entropies,
- the constraints are given by the matrices (P, M),
- $\mathcal{P} = \big\{ (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{p+m} : \mathbf{x} \ge 0,\ \mathbf{y} \ge 0,\ P\mathbf{x} + M\mathbf{y} \ge 0 \big\}$,
- $\mathcal{Q} = \big\{\mathbf{x} \in \mathbb{R}^p :$ for some $\mathbf{y} \in \mathbb{R}^m,\ (\mathbf{x}, \mathbf{y}) \in \mathcal{P}\big\}$.
- $\mathcal{Q}^\circ = \big\{P^T\mathbf{v} \in \mathbb{R}^p : \mathbf{v} \in \mathbb{R}^n,\ \mathbf{v} \ge 0,\ M^T\mathbf{v} = 0\big\}$,

The gains are

1. the first (Ingleton) coordinate in $\mathcal{Q}^\circ$ can be fixed to be 1;
2. the last four coordinates in $\mathcal{Q}^\circ$ can be requested to be zero.

These conditions can be moved from P to M to get (P_*, M_*).
The relevant part of $\mathcal{Q}^\circ$ with coordinates $\lambda_2, \dots, \lambda_{11}$ is

- $\mathcal{Q}^* = \big\{ P_*^T\mathbf{v} \in \mathbb{R}^{10} : \mathbf{v} \in \mathbb{R}^n,\ \mathbf{v} \ge 0,\ M_*^T\mathbf{v} = \mathbf{e}_{\text{Ing}} \big\}$.

⇒

The optimization problem

⇒

- $\mathcal{Q}^* = \{\, P_*^T\mathbf{v} \in \mathbb{R}^{10} \;:\; \mathbf{v} \in \mathbb{R}^n,\ \mathbf{v} \ge 0,\ M_*^T\mathbf{v} = \mathbf{e}_{\text{Ing}} \,\}$, where $\mathbf{e}_{\text{Ing}}$ is the Ingleton unit vector.

Observations

a) If $\boldsymbol{\lambda} \in \mathcal{Q}^*$, then $\boldsymbol{\lambda} \ge 0$.
b) $\mathcal{Q}^*$ is upward closed: if $\boldsymbol{\lambda} \in \mathcal{Q}^*$, and $\boldsymbol{\lambda} \le \boldsymbol{\lambda}'$, then $\boldsymbol{\lambda}' \in \mathcal{Q}^*$.

The vertices of $\mathcal{Q}^*$ are the coefficients of the “strongest” entropy inequalities which can be extracted from the copy string,

The optimization problem

$\Rightarrow$

- $\mathcal{Q}^* = \{ P_*^T\mathbf{v} \in \mathbb{R}^{10} : \mathbf{v} \in \mathbb{R}^n,\ \mathbf{v} \ge 0,\ M_*^T\mathbf{v} = \mathbf{e}_{\text{Ing}} \}$, where $\mathbf{e}_{\text{Ing}}$ is the Ingleton unit vector.

Observations

a) If $\boldsymbol{\lambda} \in \mathcal{Q}^*$, then $\boldsymbol{\lambda} \ge 0$.
b) $\mathcal{Q}^*$ is upward closed: if $\boldsymbol{\lambda} \in \mathcal{Q}^*$, and $\boldsymbol{\lambda} \le \boldsymbol{\lambda}'$, then $\boldsymbol{\lambda}' \in \mathcal{Q}^*$.

The vertices of $\mathcal{Q}^*$ are the coefficients of the "strongest" entropy inequalities which can be extracted from the copy string, and the vertices of $\mathcal{Q}^*$ are the solutions of

Multiobjective optimization problem

Find the minimum of: $P_*^T\mathbf{v} \in \mathbb{R}^{10}$ $\Leftarrow$ 10 objectives

subject to: $\mathbf{v} \ge 0$, and $M_*^T\mathbf{v} = \mathbf{e}_{\text{Ing}}$ $\Leftarrow$ constraints

Outline

Benson's outer approximation algorithm

The problem is to find the vertices of the polytope

$$\mathcal{Q}^* = \{ P_*^T \mathbf{v} \,:\, \mathbf{v} \ge 0,\ M_*^T\mathbf{v} = \mathbf{e} \}.$$

Benson's idea: Given the internal point $\mathbf{x}_i \in \mathcal{Q}^*$, and the external point $\mathbf{x}_o \notin \mathcal{Q}^*$, find

$$\max_\mu \{ 0 \le \mu \le 1 \,:\, \mu\mathbf{x}_o + (1-\mu)\mathbf{x}_i \in \mathcal{Q}^* \}.$$

a) This is an $n+11$-dimensional LP problem.

b) The (dual of the) solution gives a proof for maximality, which is a facet of $\mathcal{Q}^*$ separating $\mathbf{x}_i$ and $\mathbf{x}_o$.

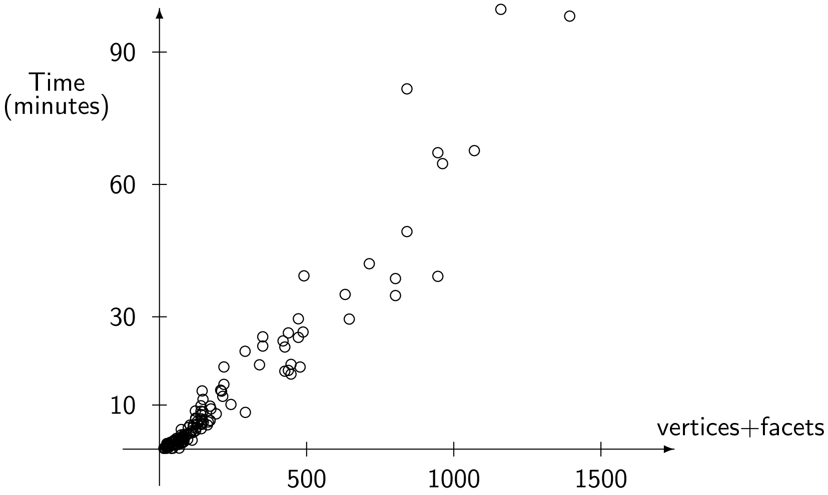
The algorithm

Use this idea to get all facets of $\mathcal{Q}^*$, maintaining the vertices of the approximating polytope bounded by the facets obtained so far.

Some results for Dougherty *et al* out of 133

Copy string	Size of M_*	Vertices	Facets	Time
`r=c:ab;s=r:ac;t=r:ad`	561×80	5	20	0:01
`rs=cd:ab;t=r:ad;u=s:adt`	1509×172	40	132	6:19
`rs=cd:ab;t=a:bcs;u=(cs):abrt`	1569×178	47	76	6:51
`rs=cd:ab;t=a:bcs;u=b:adst`	1512×178	177	261	17:40
`rs=cd:ab;t=a:bcs;u=t:acr`	1532×178	85	134	18:27
`rs=cd:ab;t=(cr):ab;u=t:acs`	1522×172	181	245	22:58
`r=c:ab;st=cd:abr;u=a:bcrt`	1346×161	209	436	29:18
`rs=cd:ab;t=a:bcs;u=c:abrst`	1369×166	355	591	38:59
`rs=cd:ab;t=a:bcs;u=c:abrt`	1511×178	363	599	1:04:32
`rs=cd:ab;t=a:bcs;u=s:abcdt`	1369×166	355	591	1:07:01
`rs=cd:ab;t=a:bcs;u=(at):bcs`	1555×177	484	676	1:39:30
`rs=cd:ab;t=a:bcs;u=a:bcst`	1509×177	880	1238	4:30:26
`rs=cd:ab;t=a:bcs;u=a:bdrt`	1513×177	2506	2708	5:11:25

Running time vs. vertices + facets



Some results with five auxiliary variables

Copy string	Size of M_*	Vertices	Facets	Time
`rs=cd:ab;tu=cr:ab;v=(cs):abtu`	4055×370	19	58	1:10:10
`rs=ad:bc;tu=ar:bc;v=r:abst`	4009×370	40	103	3:24:37
`rs=cd:ab;t=(cr):ab;uv=cs:abt`	3891×358	30	102	3:34:31
`rs=cd:ab;tu=cr:ab;v=t:adr`	3963×362	167	235	9:20:19
`rs=cd:ab;tu=dr:ab;v=b:adsu`	4007×370	318	356	13:20:08
`rs=cd:ab;tv=dr:ab;u=a:bcrt`	4007×370	318	356	14:34:42
`rs=cd:ab;tu=cs:ab;v=a:bcrt`	4007×370	297	648	22:02:39
`rs=cd:ab;t=a:bcs;uv=bt:acr`	3913×362	779	1269	37:15:33
`rs=cd:ab;tu=cr:ab;v=a:bcstu`	3987×362	4510	7966	427:43:30
`rs=cd:ab;tu=cs:ab;v=a:bcrtu`	3893×362	10387	13397	716:36:32

Using five auxiliary variables, more than 260 new entropy inequalities were generated. One of them is

$$2[abcd\,] + (a,b\,|\,c) + 3(a,c\,|\,b) + (b,c\,|\,a) + 3(c,d\,|\,a) \geq 0.$$

Thank you for your attention