Parallel optimization on the Entropic Cone

Benoît Legat Raphaël Jungers
Université catholique de Louvain
ICTEAM
4 Av. G. Lemaître
1348 Louvain-la-Neuve, Belgium
benoit.legat@student.uclouvain.be raphael.jungers@uclouvain.be

Abstract

We introduce a parallelizable algorithm for approximate optimization on the entropic cone. We also present the toolbox `EntropicCone.jl`. Its aim is to improve the computational repoducibility of the recent progress on the approximation of the entropic cone and to make them easily accessible for its many applications. These applications include the capacity region of multi-source network coding, converse theorems for multi-terminal problems of information theory, bounds on the information ratios in secret sharing schemes and conditional independence among subvectors of a random vector.

1 The problem setting

Given n random variables, we can compute the entropy of any of the 2^n subsets of these n variables. The set $\mathcal{E}_n \triangleq \mathbb{R}^{2^n-1}$ of vectors indexed by the nonempty subsets of $[n] \triangleq \{1, \ldots, n\}$ is called the *entropy space*.The entropy vector of a set of n random variables is the entropy vector h such that h_I is the entropy of the set $\{\, X_i \mid i \in I \,\}$.

We denote the set of vectors of $\mathbb{R}^{2^n-1}$ that are entropic as:

$$\mathcal{H}_n \triangleq \{\, h \in \mathbb{R}^{2^n-1} \mid \exists X_1, \ldots, X_n, \forall \emptyset \neq S \subseteq [n], h_S = H_b(\{\, X_i \mid i \in S \,\}) \,\}.$$

It is known that the set $\mathcal{H}_n$ is not a cone for $n \geq 3$ but its closure $\operatorname{cl} \mathcal{H}_n$ is a convex cone [20]. The difference between $\mathcal{H}_n$ and $\operatorname{cl} \mathcal{H}_n$ is only on the boundary of $\operatorname{cl} \mathcal{H}_n$. More precisely, it has been shown that the relative interior of $\operatorname{cl} \mathcal{H}_n$ is contained in $\mathcal{H}_n$ [15].

For $n \leq 3$, $\operatorname{cl} \mathcal{H}_n$ is equal to the *polymatroid cone* $\mathcal{P}_n$. This is the set of entropy vectors h that are

- ◇ *nonnegative*: $h_I \geq 0$ for any $I \subseteq [n]$,
- ◇ *nondecreasing*: $h_I \leq h_J$ for any $I \subseteq J \subseteq [n]$ and
- ◇ *submodular*: $h_J + h_K \geq h_{J \cup K} + h_{J \cap K}$ for any $J, K \subseteq [n]$.

These three sets of conditions are linear inequalites on the entropy vector h. Since $\mathcal{P}_n$ is defined by a finite subset of linear inequalities, it is a polyhedral cone.

For $n \geq 4$, $\operatorname{cl} \mathcal{H}_n$ is a strict subset $\mathcal{P}_n$. Moreover, $\operatorname{cl} \mathcal{H}_n$ is not polyhedral [14] and not even semialgebraic* [17].

The entropic cone has a varietie of applications including the capacity region of multi-source network coding [1], converse theorems for multi-terminal problems of information theory [19], bounds on the information ratios in secret sharing schemes [2]

*A set is *semialgebraic* if it is the projection of an algebraic set. A set is *algebraic* if it can be defined by finitely many polynomial inequalities.

and conditional independence among subvectors of a random vector [18]. This motivates the research on filling the gap between $\mathcal{P}_n$ and $\operatorname{cl}\mathcal{H}_n$.

In Section 2, we review the methods used to find tighter approximations of $\mathcal{H}_n$ than $\mathcal{P}_n$. The current methods are linear and generate polyhedral outer approximations. However, as one can anticipate, the number of facets of a polyhedral approximation that would be "everywhere close" to $\mathcal{H}_n$ would have a sizeable amount of facets since it is high-dimensional and not semialgebraic. In practical applications, one is often looking at the simpler problem of solving an optimization problem involving the entropic cone so we are only looking for an outer approximation that is "close" to $\mathcal{H}_n$ "near" the optimum. In Section 3, we show a parallelizable algorithm for this problem.

2 Generating Non-Shannon Inequalities

The inequalities that are valid for $\mathcal{P}_n$ are called *Shannon inequalities* and those that are valid for $\mathcal{H}_n$ but not for $\mathcal{P}_n$ are called *non-Shannon inequalities*.

The current approach in generating an outer bound for $\mathcal{H}_n$ is to generate non-Shannon inequalities and to intersect $\mathcal{P}_n$ with the halfspaces they define.

There are currently four known methods for generating non-Shannon inequalities. The first one, which is the most commonly used, was introduced by Zhang and Yeung in order to generate the first non-Shannon inequality [21]; its description and analysis can be found in [13]. The three other methods are respectively described in [12], [15] and [6].

The first two methods are equivalent [10] and it is still unknown whether the third and four methods can generate inequalities that cannot be generated by the first two ones. It is also unknown whether these four methods can generate all non-Shannon inequalities. In this paper, we will only use the first method which is described in Section 2.1.

2.1 Entropic Cone and adhesivity

We define the *inner-adhesivity* and *self-adhesivity* operators respectively as

$$\mathsf{ia}_{J,K|I}(h) = \{\, g \in \mathcal{E}_{n'} \mid \forall I \subseteq L \subseteq J \cup K, g_L = h_{L\cap J} + h_{L\cap K} - h_I \,\}, \qquad I = J \cap K,$$
$$\mathsf{sa}_{J|I}(h) = \{\, g \in \mathcal{E}_{n'} \mid \forall I \subseteq L \subseteq J' \cup K, g_L = h_{L\cap J'} + h_{L\cap K} - h_I \,\}, \qquad I \subseteq J,$$

where for inner-adhesivity, $n' = |J \cup K|$ and for self-adhesivity $n' = n + |J \setminus I|$, $J' = [n]$ and $K = ([n'] \setminus J') \cup I$.

Definition 1. We say that a family of sets $\mathcal{S}_n \subseteq \mathcal{E}_n$ is *inner-adhesive* if for any n, $x \in \mathcal{S}_n$ and $J, K \subseteq [n]$, there exists $y \in \mathcal{S}_{|J\cup K|}$ such that $y \in \mathsf{ia}_{J,K|J\cap K}(x)$ and we say that it is *self-adhesive* if for any n, $x \in \mathcal{S}_n$ and $I \subseteq J \subseteq [n]$, there exists $y \in \mathcal{S}_{n+|J\setminus I|}$ such that $y \in \mathsf{sa}_{J|I}(x)$.

The following theorem gives the relation between adhesivity and the entropic cone.

Theorem 1. The entropic cone $\mathcal{H}_n$ is inner-adhesive and self-adhesive.

Proof. Consider $h \in \mathcal{H}_n$ and n jointly distributed random variables X_i of joint probability mass function p such that h is their entropy vector. Let p_I denote the probability mass function of the marginal distribution of I. For inner-adhesivity, the entropy vector $g \in \mathcal{H}_{|\mathcal{S}\cup\mathcal{T}|}$ of the probability function

$$\frac{p_J(x_J)p_K(x_K)}{p_I(x_I)} = p_{J|I}(x_J|x_I)p_{K|I}(x_K|x_I)p_I(x_I)$$

belongs to $\mathsf{ia}_{J,K|I}(h)$ and for self-adhesivity, the entropy vector $g \in \mathcal{H}_{n+|J\setminus I|}$ of the probability function

$$\frac{p_{J'}(x_{J'})p_J(x_K)}{p_I(x_I)} = p_{J'|I}(x_{J'}|x_I)p_{J|I}(x_K|x_I)p_I(x_I),$$

where $J' = [n]$, belongs to $\mathsf{sa}_{J|I}(h)$. □

Let $\mathsf{ia}_{J,K|I}^{-1}(\mathcal{S})$ (resp. $\mathsf{sa}_{J|I}^{-1}(\mathcal{S})$) be the set of vectors x such that there exists $y \in \mathcal{S}$ such that $y \in \mathsf{ia}_{J,K|I}(x)$ (resp. $y \in \mathsf{sa}_{J|I}(\mathcal{S})(x)$). By Theorem 1, given an integer n_0 and a sequence $(J_1, I_1), \ldots, (J_m, I_m)$ such that $I_i \subseteq J_i \subseteq [n_i]$ and $n_i = n_{i-1} + |J_i \setminus I_i|$ for $i = 1, \ldots, m$, the set[†]

$$(\mathsf{sa}_{J_1|I_1}^{-1} \circ \cdots \circ \mathsf{sa}_{J_m|I_m}^{-1})(\mathcal{P}_{n_m})$$

provides an outer approximations of $\mathcal{H}_{n_0}$. Therefore an outer approximation of $\mathcal{H}_{n_0}$ can be obtained by projecting the polyhedral cone of dimension $\sum_{i=0}^{m} 2^{n_i} - 1$ given by

$$\{\, (h_0, h_1, \ldots, h_m) \in \mathcal{P}_{n_0} \times \mathcal{P}_{n_1} \times \cdots \times \mathcal{P}_{n_m} | h_i \in \mathsf{sa}_{J|I}(h_{i-1}), i = 1, \ldots, m \,\}. \quad (1)$$

on the first $2^{n_0} - 1$ variables.

This method was used to generate hundreds of non-Shannon inequalities in [8]. Using Benson's algorithm [3] to compute the projection of (1), even more non-Shannon inequalities were uncovered in [7, 11].

As mentioned earlier, $\operatorname{cl} \mathcal{H}_n$ is strictly included in $\mathcal{P}_n$ for $n \geq 4$. As the dimension of $\mathcal{H}_n$ is exponential in n, the methods are usually benchmarked using $\mathcal{H}_4$. An important numerical quantity, related with the geometric properties of $\mathcal{H}_4$ is the *Ingleton score* defined as

$$\mathbb{I}^* \triangleq \inf_{0 \neq h \in \mathcal{H}_4} \mathbb{I}_{ij}(h)$$

where $\mathbb{I}_{ij}(h) = \langle \square_{ij}, h \rangle / h_{[n]}$ [8, Definition 3] and $\square_{ij}$ is the Ingleton dual entropy vector [9]. The current best lower bound on $\mathbb{I}^*$ is equal to -0.15789 [8]. Upper bounds on $\mathbb{I}^*$ can be obtained from exhibiting four jointly distributed variables for which the entropy vector has low Ingleton score. The current best upper bound on $\mathbb{I}^*$ is equal to -0.09243 [16].

3 Parallelizable optimization on the Entropic Cone

In this section, we show how to decompose polyhedra such as described by (1) to solve optimization problems on it in an efficient and parallelizable manner using ideas from *Stochastic Programming* [5]. In stochastic programming, large scale linear programs are decomposed into smaller linear programs linked together by a markov chain. The linear program at each state u of the markov chain is:

$$\begin{aligned} Q(x, u) = \text{minimize } & c^T y + \mathcal{Q}_u(y) \\ \text{s.t. } & W_u y = h_u - T_u x, \\ & x \geq 0 \end{aligned}$$

where $\mathcal{Q}_u(y)$ is the sum of $Q(x, v)$ for each state v accessible from u weighted by the probability to go from state u to state v. When the program is infeasible for some x,

[†]Of course this also works if we include inner-adhesive operations in the sequence, we have only included self-adhesivity in the sequence to keep simple notation.

$Q(x, u) = \infty$. At the initial state of the markov chain, there is no term $-T_u x$ and the solution at this stage is the solution of the original large scale linear program. Note that if v is accessible from different states u, u', the number of variables at u and u' must match for $Q(\cdot, v)$ to be well-defined.

For the adhesive operations, we propose to define a state for each dimension n and adhesive operation. That is, for every n, J, K, we have a state for the operation $\mathsf{ia}_{J,K|J\cap K}$ and for every n, J, I with $I \subseteq J$, we have a state for the operation $\mathsf{sa}_{J|I}$. The linear program at the initial state is

$$\begin{aligned} &\text{minimize } c^T h + \mathcal{Q}_0(h) \\ &\text{s.t. } h \in \mathcal{P}_{n_0}, \end{aligned}$$

the linear program for each state representing an inner-adhesivity is

$$\begin{aligned} Q(h, (n, \mathsf{ia}_{J,K|J\cap K})) = \text{minimize } & \mathcal{Q}_{(n, \mathsf{ia}_{J,K|J\cap K})}(g) \\ \text{s.t. } & g \in \mathsf{ia}_{J,K|J\cap K}(h) \\ & g \in \mathcal{P}_{|J\cup K|} \end{aligned}$$

and the linear program for each state representing an self-adhesivity is

$$\begin{aligned} Q(h, (n, \mathsf{sa}_{J|I})) = \text{minimize } & \mathcal{Q}_{(n, \mathsf{sa}_{J|I})}(g) \\ \text{s.t. } & g \in \mathsf{sa}_{J|I}(h) \\ & g \in \mathcal{P}_{n+|J\setminus I|}. \end{aligned}$$

A state (n, a) is accessible from a state u if the dimension‡ of the linear program at u is $2^n - 1$. Since there is no objective in states other than the initial state, the probability assigned for each transition does not matter. The only relevant information in $Q(h, (n, \mathsf{a}))$ is whether it is infinite or zero.

In stochastic programming, the domain of $Q(\cdot, (n, \mathsf{a}))$ is approximated by a polyhedron by starting with the approximation $\mathbb{R}^n$ and adding *feasibility cuts*. The feasibility cuts are computed as follows: the linear program is solved for some h, if it is infeasible, an unbounded ray of the dual linear program is computed and used to generate a feasibility cut.

These observations lead to Algorithm 1 for approximate optimization on the entropic cone that is inspired from the *Stochastic Dual Dynamic Programming* algorithm used in stochastic programming. Note that the initial node can have a nonlinear objective function and additionnal nonlinear constraints. This algorithm is easily parallelizable as the linear programs of the different states only need to communicate cuts and optimal solutions.

We developed a new toolbox `EntropicCone.jl` in Julia [4] for working with the entropic cone. This algorithm is one of the features implemented in the toolbox. We tested Algorithm 1 to find lower bounds for the Ingleton score and obtained the best known lower bound -0.15789 in under a minute.

4 Conclusion

Searching for non-Shannon inequalities to provide tighter outer approximations of the entropic cone may seem computationally demanding due to the use of a projection algorithm in high-dimensional space. However, if we restrict ourself to the optimization

‡the number of variables of the linear program

Algorithm 1 Approximate minimization of $c(h)$ subject to $x \in \mathcal{H}_n \cap \mathbb{F}$ for parameters $n_{\max}, K, m, \rho$.

Given a maximal value $n_{\max}$ for n, generate all states such that the dimension of the linear program is at most $2^{n_{\max}} - 1$
for $k = 1, 2, \ldots, K$ **do**
  Pick a set P of ρ random paths of length m starting at the initial state.
  Solve the optimization program

$$\begin{aligned} &\text{minimize } c(h) \\ &\quad \text{s.t. } h \in \mathcal{P}_{n_0} \cap \mathbb{F} \cap \operatorname{dom}(\mathcal{Q}_0), \end{aligned}$$

  where the domain of $\mathcal{Q}_0(h)$ is approximated by the feasibility cuts.
  if the program is infeasible **then**
    return Infeasible
  end if
  for $i = 1, 2, \ldots, m$ **do**
    for all $p \in P$ **do**
      Solve $Q(h, (n_{p,i}, \mathsf{a}_{p,i}))$ where h is the value of the optimal solution of the previous state in the path.
      if the program is infeasible **then**
        Add a feasibility cut for $\operatorname{dom}(Q(\cdot, (n_{p,i}, \mathsf{a}_{p,i})))$
        Remove p from P
      end if
    end for
  end for
end for

of a (possibly nonlinear) objective on the entropic cone (possibly under additionnal constraints), the algorithms used in stochastic programming can be used to provide a parallelizable algorithm that can provide bounds on the objective.This method is able to obtain current best lower bound on the Ingleton score in under a minute.

References

[1] Riccardo Bassoli, Hugo Marques, Jose Rodriguez, Kenneth W Shum, and Rahim Tafazolli. Network coding theory: A survey. *Communications Surveys & Tutorials, IEEE*, 15(4):1950–1978, 2013.

[2] Amos Beimel. Secret-sharing schemes: a survey. In *Coding and cryptology*, pages 11–46. Springer, 2011.

[3] Harold P Benson. An outer approximation algorithm for generating all efficient extreme points in the outcome set of a multiple objective linear programming problem. *Journal of Global Optimization*, 13(1):1–24, 1998.

[4] Jeff Bezanson, Stefan Karpinski, Viral B Shah, and Alan Edelman. Julia: A fast dynamic language for technical computing. *arXiv preprint arXiv:1209.5145*, 2012.

[5] John R Birge and Francois Louveaux. *Introduction to stochastic programming.* Springer Science & Business Media, 2011.

[6] Laszlo Csirmaz. Book inequalities. *Information Theory, IEEE Transactions on*, 60(11):6811–6818, 2014.

[7] László Csirmaz. Using multiobjective optimization to map the entropy region. *Computational Optimization and Applications*, 63(1):45–67, 2016.

[8] Randall Dougherty, Chris Freiling, and Kenneth Zeger. Non-Shannon information inequalities in four random variables. *arXiv preprint arXiv:1104.3602*, 2011.

[9] AW Ingleton. Conditions for representability and transversality of matroids. In *Théorie des Matroïdes*, pages 62–66. Springer, 1971.

[10] Tarik Kaced. Equivalence of two proof techniques for non-Shannon-type inequalities. In *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pages 236–240. IEEE, 2013.

[11] Andreas Löhne and Benjamin Weißing. The vector linear program solver Bensolve–notes on theoretical background. *arXiv preprint arXiv:1510.04823*, 2015.

[12] Konstantin Makarychev, Yury Makarychev, Andrei Romashchenko, and Nikolai Vereshchagin. A new class of non-Shannon-type inequalities for entropies. *Communications in Information and Systems*, 2(2):147–166, 2002.

[13] František Matúš. Adhesivity of polymatroids. *Discrete Mathematics*, 307(21):2464–2477, 2007.

[14] František Matúš. Infinitely many information inequalities. In *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*, pages 41–44. IEEE, 2007.

[15] František Matúš. Two constructions on limits of entropy functions. *Information Theory, IEEE Transactions on*, 53(1):320–330, 2007.

[16] F. Matúš and L. Csirmaz. Entropy region and convolution. *ArXiv e-prints*, October 2013.

[17] Carolina Mejia and J Andres Montoya. The almost-entropic regions are not semialgebraic. *arXiv preprint arXiv:1510.02658*, 2015.

[18] Milan Studený. *Probabilistic conditional independence structures.* Springer Science & Business Media, 2006.

[19] Raymond W Yeung. *A first course in information theory.* Springer Science & Business Media, 2012.

[20] Zhen Zhang and Raymond W Yeung. A non-Shannon-type conditional inequality of information quantities. *Information Theory, IEEE Transactions on*, 43(6):1982–1986, 1997.

[21] Zhen Zhang and Raymond W Yeung. On characterization of entropy function via information inequalities. *Information Theory, IEEE Transactions on*, 44(4):1440–1452, 1998.