Fourier-Motzkin Elimination Software for Information Theoretic Inequalities

Ido B. Gattegno, Ziv Goldfeld, and Haim H. Permuter
Ben-Gurion University of the Negev

I. Abstract

We provide open-source software implemented in MATLAB, that performs Fourier-Motzkin elimination (FME) and removes constraints that are redundant due to Shannon-type inequalities (STIs). The FME is often used in information theoretic contexts to simplify rate regions, e.g., by eliminating auxiliary rates. Occasionally, however, the procedure becomes cumbersome, which makes an error-free hand-written derivation an elusive task. Some computer software have circumvented this difficulty by exploiting an automated FME process. However, the outputs of such software often include constraints that are inactive due to information theoretic properties. By incorporating the notion of STIs (a class of information inequalities provable via a computer program), our algorithm removes such redundant constraints based on non-negativity properties, chain-rules and probability mass function factorization. This newsletter first illustrates the program's abilities, and then reviews the contribution of STIs to the identification of redundant constraints.

II. The Software

The Fourier-Motzkin elimination for information theory (FME-IT) program is implemented in MATLAB and available, with a graphic user interface (GUI), at http://www.ee.bgu.ac.il/~fmeit/. The Fourier-Motzkin elimination (FME) procedure [1] eliminates variables from a linear constraints system to produce an equivalent system that does not contain those variables. The equivalence is in the sense that the solutions of both systems over the remaining variables are the same. To illustrate the abilities of the FME-IT algorithm, we consider the Han-Kobayashi (HK) inner bound on the capacity region of the interference channel [2] (here we use the formulation from [3, Theorem 6.4]). The HK coding scheme insures reliability if certain inequalities that involve the partial rates R_{10}, R_{11}, R_{20} and R_{22}, where

$$R_{jj} = R_j - R_{j0},\ j = 1,2, \tag{1}$$

are satisfied. To simplify the region, the rates R_{jj} are eliminated by inserting (1) into the rate bounds and adding the constraints

$$R_{j0} \le R_j,\ j = 1,2. \tag{2}$$

The inputs and output of the FME-IT program are illustrated in Fig. 1. The resulting inequalities of the HK coding scheme are fed into the textbox labeled as 'Inequalities'. The non-negativity of all the terms involved is accounted for by checking the box in the upper-right-hand corner. The terms designated for elimination and the target terms (that the program isolates in the final output) are also specified. The joint probability mass function (PMF) is used to extract statistical relations between random variables. The relations are described by means of equalities between entropies. For instance, in the HK coding scheme, the joint PMF factors as

$$P_{Q,U_1,U_2,X_1,X_2,Y_1,Y_2} = P_Q P_{X_1,U_1|Q} P_{X_2,U_2|Q} P_{Y_1,Y_2|X_1,X_2}, \tag{3}$$

and implies that $(X_2,U_2) - Q - (X_1,U_1)$ and $(Y_1, Y_2) - (X_1, X_2) - (Q, U_1,U_2)$ form Markov chains. These relations are captured by the following equalities:

$$H(X_2,U_2 | Q) = H(X_2,U_2 | Q,U_1,X_1) \tag{4a}$$

$$H(Y_1,Y_2 | X_1,X_2) = H(Y_1,Y_2 | Q,U_1,U_2,X_1,X_2). \tag{4b}$$

The output of the program is the simplified system from which redundant inequalities are removed. Note that although the first and the third inequalities are redundant [4, Theorem 2], they are not captured by the algorithm. This is since their redundancy relies on the HK inner bound being a union of polytops over a domain of joint PMFs, while the FME-IT program only removes constraints that are redundant for every fixed PMF. An automation of the FME for information theoretic purposes was previously provided in [5]. However, unlike the FME-IT algorithm, the implementation in [5] cannot identify redundancies that are implied by information theoretic properties.

III. Theoretical Background

A. Preliminaries

We use the following notation. Calligraphic letters denote discrete sets, e.g., χ. The empty set is denoted by ϕ, while $\mathcal{N}_n \triangleq \{1,2,\ldots,n\}$ is a set of indices. Lowercase letters, e.g. x, represent variables. A column vector of n variables $(x_1,\ldots,x_n)^\top$ is denoted by $\mathbf{x}_{\mathcal{N}_n}$, where $\mathbf{x}^\top$ denoted the transpose of $\mathbf{x}$. A substring of $\mathbf{x}_{\mathcal{N}_n}$ is denoted by $\mathbf{x}_\alpha = (x_i \in \Omega | i \in \alpha, \phi \ne \alpha \subseteq \mathcal{N}_n)$, e.g., $\mathbf{x}_{\{1,2\}} = (x_1, x_2)^\top$. Whenever the dimensions are clear from the context, the subscript is omitted. Non-italic capital letters, such as A, denote matrices. Vector inequalities, e.g., $\mathbf{v} \ge \mathbf{0}$, are in the componentwise sense. Random variables are denoted by uppercase letters, e.g., X, and similar conventions apply for random vectors.

B. Redundant Inequalities

Some of the inequalities generated by the FME may be redundant. Redundancies may be implied either by other inequalities or by information theoretic properties. To account for the latter, we combine the notion of Shannon-type inequalities (STIs) with a method that identifies redundancies by solving a linear programming (LP) problem.

1) *Identifying Redundancies via Linear Programming:* Let $\mathrm{A}\mathbf{x} \ge \mathbf{b}$ be a system of linear inequalities. To test whether the i-th inequality is redundant, define

- $\mathrm{A}^{(i)}$ - a matrix obtained by removing the i-th row of A;
- $\mathbf{b}^{(i)}$ - a vector obtained by removing the i-th entry of b;
- $\mathbf{a}_i^\top$ - the i-th row of A;
- b_i - the i-th entry of b.

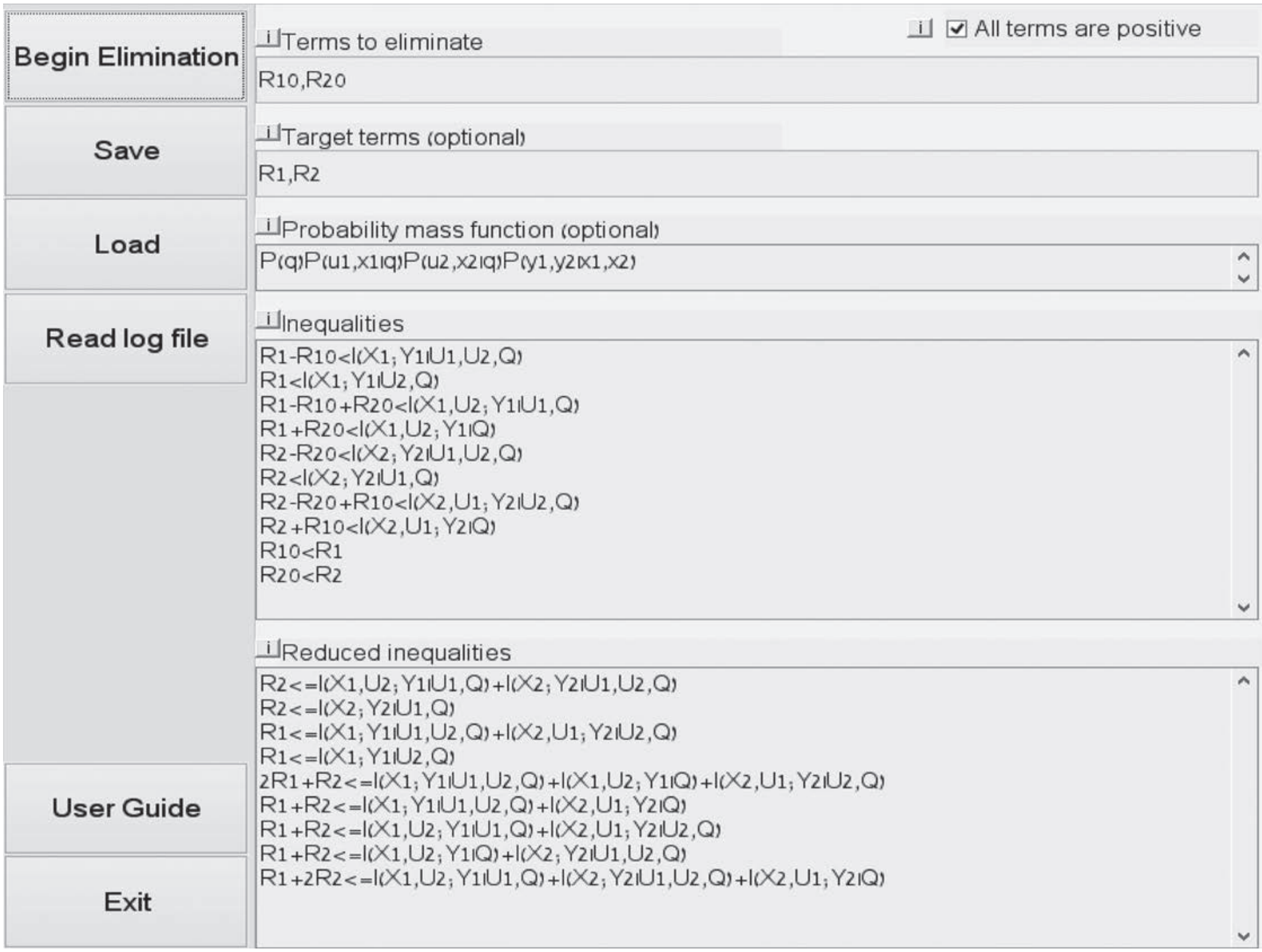

Fig. 1 FME-IT input and output - HK inner bound.

The following lemma states a sufficient and necessary condition for redundancy.

Lemma 1 (Redundancy identification) *The i-th linear constraint in a system* $\mathrm{A}\mathbf{x} \geq \mathbf{b}$ *is redundant if and only if*

$$\rho_i^* = \min_{\substack{\mathbf{x}: \\ \mathrm{A}^{(i)}\mathbf{x} \geq \mathbf{b}^{(i)}}} \mathbf{a}_i^\top \mathbf{x} \tag{5}$$

satisfies $\rho_i^* \geq b_i$.

Lemma 1 lets one determine whether a certain inequality is implied by the remaining inequalities in the system by solving a LP problem. When combined with the notion of STIs, the lemma can also be used to identify redundancies due to information theoretic properties.

2) *Shannon-Type Inequalities:* In [6], Yeung characterized a subset of information inequalities named STIs, that are provable using the ITIP computer program [7] (see also [8]).

Given a random vector $\mathbf{X}_{\mathcal{N}_n}$ that takes values in $\chi_1 \times \ldots \times \chi_n$, define $\mathbf{h}_\ell \triangleq \big(H(\mathbf{X}_\alpha)\,|\,\phi \neq \alpha \subseteq \mathcal{N}_n\big)$[1]. The entries of $\mathbf{h}_\ell$ are *labels* that correspond to the joint entropies of all substrings of $\mathbf{X}_{\mathcal{N}_n}$. Every linear combination of Shannon's information measures is uniquely representable as $\mathbf{b}^\top \mathbf{h}_\ell$, where b is a vector of coefficients. This representation is called the *canonical form*. Fixing the PMF of $\mathbf{X}_{\mathcal{N}_n}$ to p, $\mathbf{h}_\ell(p) \in \mathbb{R}^{2^n-1}$ denotes the evaluation of $\mathbf{h}_\ell$ with respect to p.

We represent a linear information inequality as $\mathbf{f}^\top \mathbf{h}_\ell \geq 0$, where f is a vector of coefficients, and say that it *always holds* if it holds for every PMF. Formally, if

$$\min_{p \in \mathcal{P}} \mathbf{f}^\top \mathbf{h}(p) = 0, \tag{6}$$

where $\mathcal{P}$ is the set of all PMFs on $\mathbf{X}_{\mathcal{N}_n}$, then $\mathbf{f}^\top \mathbf{h}_\ell \geq 0$ always holds.

Since the minimization problem in (6) is intractable, Yeung suggested a simple affine space that contains the set where the canonical vectors take values. This space is described by all basic inequalities, which are non-negativity inequalities on all involved entropy and mutual information terms. The description is further simplified by introducing a minimal set of information inequalities, referred to as *elemental inequalities*.

[1]We order the elements of $\mathbf{h}_\ell$ lexicographically.

Definition 1 (Elemental inequality) *The set of elemental inequalities is given by:*

$$H(X_i \mid \mathbf{X}_{\mathcal{N}_n \setminus \{i\}}) \geq 0 \tag{7a}$$

$$I(X_i; X_j \mid \mathbf{X}_{\mathcal{K}}) \geq 0, \tag{7b}$$

where $i, j \in \mathcal{N}_n, i \neq j, \mathcal{K} \subseteq \mathcal{N}_n \setminus \{i, j\}$.

The left-hand side of every elemental inequality is a linear combination of the entries of $\mathbf{h}_\ell$. Therefore, the entire set can be described in matrix form as

$$\mathrm{G}\mathbf{h}_\ell \geq \mathbf{0}, \tag{8}$$

where G is a matrix whose rows are coefficients. Consequently, the cone

$$\Gamma_n = \left\{ \mathbf{h} \in \mathbb{R}^{2^n - 1} \,\middle|\, \mathrm{G}\mathbf{h} \geq \mathbf{0} \right\}, \tag{9}$$

contains the region where $\mathbf{h}_\ell(p)$ take values. The converse, however, does not hold in general.

Based on Γ_n, one may prove that an information inequality always holds by replacing the convoluted minimization problem from (6) with a LP problem. To state this result, we describe the probabilistic relations that stem from the factorization of the underlying PMF by means of linear equalities between entropies (such as in (3)) as

$$\mathrm{Q}\mathbf{h}_\ell = \mathbf{0}, \tag{10}$$

where Q is a matrix of coefficients.

Theorem 1 (Constrained STIs [6, Theorem 14.4]) *Let* $\mathbf{b}^\top \mathbf{h}_\ell \geq 0$ *be an information inequality, and let*

$$\rho^* = \min_{\substack{\mathbf{h}: \\ \mathrm{G}\mathbf{h} \geq \mathbf{0} \\ \mathrm{Q}\mathbf{h} = \mathbf{0}}} \mathbf{b}^\top \mathbf{h}. \tag{11}$$

If $\rho^* = 0$, *then* $\mathbf{b}^\top \mathbf{h}_\ell \geq 0$ *holds for all PMFs for which* $\mathrm{Q}\mathbf{h}_\ell = \mathbf{0}$, *and is called a constrained STI.*

IV. The Software Algorithm

The algorithm is executed in three stages. In the first stage, the input system of linear inequalities is transformed into matrix form. Assume the input system contains L variables. Denote by $\mathbf{r}_0$ the L-dimensional vector whose entries are the variables of the system. The input inequalities are represented as

$$\mathrm{A}_0 \mathbf{r}_0 + \mathrm{B}_0 \mathbf{h}_\ell \geq \mathbf{c}_0, \tag{12}$$

where $\mathbf{c}_0$ is a vector of constants and $\mathbf{h}_\ell$ is the vector of joint entropies as defined in Subsection III-B2. The rows of the matrices A_0 and B_0 hold the coefficients of the rates and the information measures, respectively, in each inequality. We rewrite (12) as

$$\mathrm{A}_1 \mathbf{x}_1 \geq \mathbf{c}_0, \tag{13a}$$

where

$$\mathrm{A}_1 \triangleq [\mathrm{A}_0 \,|\, \mathrm{B}_0] \tag{13b}$$

$$\mathbf{x}_1 \triangleq (\mathbf{r}_0^\top \; \mathbf{h}_\ell^\top)^\top. \tag{13c}$$

Henceforth, the elements of $\mathbf{h}_\ell$ are also treated as variables.

The second stage executes FME. Suppose we aim to eliminate the first $L_0 < L$ variables in the original $\mathbf{r}_0$. To do so, we run the FME on the first L_0 elements of $\mathbf{x}_1$ (see (13c)) and obtain the system

$$\mathrm{A}\mathbf{x} \geq \mathbf{c}, \tag{14}$$

where x is the reduced version of $\mathbf{x}_1$ after the elimination. The matrix A and the vector c are determined by the FME procedure.

The third stage identifies and removes redundancies. Let

$$\tilde{\mathrm{G}} \triangleq [0 \,|\, \mathrm{G}], \tag{15}$$

where G is the matrix from (8), and

$$\tilde{\mathrm{A}} \triangleq \begin{bmatrix} \mathrm{A} \\ \tilde{\mathrm{G}} \end{bmatrix} \tag{16a}$$

$$\tilde{\mathbf{c}} \triangleq \left(\mathbf{c}^\top \mathbf{0}^\top\right)^\top. \tag{16b}$$

Further, to account for constraints that are induced by the underlying PMF factorization, set

$$\tilde{\mathrm{Q}} \triangleq [0 \,|\, \mathrm{Q}], \tag{17}$$

where Q is the matrix from (10). Applying Lemma 1 (redundancy identification)[2] on each of the rows of

$$\tilde{\mathrm{A}}\mathbf{x} \geq \tilde{\mathbf{c}} \tag{18a}$$

under the constraint

$$\tilde{\mathrm{Q}}\mathbf{x} = \mathbf{0}, \tag{18b}$$

while relying on the machinery of Theorem 1, removes the redundant inequalities and results in the reduced system.

References

[1] A. Schrijver, *Theory of linear and integer programming*. John Wiley & Sons, 1998.

[2] Te Sun, Han and Kobayashi, Kingo, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.

[3] El Gamal, Abbas and Kim, Young-Han, *Network information theory*. Cambridge University Press, 2011.

[4] H.-F. Chong, M. Motani, H. K. Garg, and H. E. Gamal, "On the Han-Kobayashi Region for the Interference Channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3188–3195, July 2008.

[2]We use an extended version of Lemma 1 that accounts also for equality constraints [9, Theorem 2.1].

[5] Joffrey Villard, "GUI for automatic Fourier-Motzkin elimination," http://www.joffrey-villard.fr/fmg?lang=en.

[6] R. W. Yeung, *Information theory and network coding*. Springer Science & Business Media, 2008.

[7] R. W. Yeung and Y.-O. Yan, "Information theoretic inequality prover (ITIP)," http://user-www.ie.cuhk.edu.hk/~ITIP/.

[8] Rethnakaran Pulikkoonattu, Etienne Perron and Suhas Diggavi, "X Information Theoretic Inequalities Prover," http://xitip.epfl.ch/.

[9] S. Jibrin and D. Stover, "Identifying redundant linear constraints in systems of linear matrix inequality constraints," *Journal of Interdisciplinary Mathematics*, vol. 10, no. 5, pp. 601–617, May 2007.

President's Column *continued from page 1*

of Error for Classical and Classical-Quantum Channels." The 2015 IT Society Paper Award was also announced; the award winning paper is titled "A Family of Optimal Locally Recoverable Codes" by Itzhak Tamo and Alexander Barg. The inaugural James L. Massey Research and Teaching Award for Young Scholars went to Young-Han Kim. The 2015 IT Society Aaron D. Wyner Distinguished Service Award went to Han Vinck. And the 2015 Shannon Award, announced at ISIT 2014, was presented to Robert Calderbank. The 2015 Wolf Student Paper Awards were announced later in the week; the recipients were Tarun Jog for the paper "On the Geometry of Convex Typical Sets," Marco Mondelli for the paper "Unified Scaling of Polar Codes: Error Exponent, Scaling Exponent, Moderate Deviations, and Error Floors," and Yihong Wu and Pengkun Yang for their paper "Optimal Entropy Estimation on Large Alphabets via best Polynomial Approximation."

The ISIT 2015 banquet ended with the announcement of the 2016 Shannon Award. The 2016 Shannon Lecturer will be Alexander Semenovich Holevo. Professor Holevo was chosen for our community's highest honor for his contributions to the field of quantum information theory. Over forty years ago, Holevo initiated the study of capacities for quantum channels. Very recently, he and his co-authors have settled a decades-old conjecture resolving the capacities of quantum Gaussian channels.

Looking ahead, we are moving forward on a number of plans to celebrate Shannon's 100th birthday. You can read updates on these plans in the Centenary section on Claude Shannon's Wikipedia page (https://en.wikipedia.org/wiki/Claude_Shannon#Shannon_Centenary). Bell Labs will develop an exhibit on Shannon's time there. Many universities around the world have signed on to hold "Shannon Day" programs—celebrating the Shannon centenary through public outreach events targeted to the general population and in particular to young people. These include Technische Universität Berlin, University of South Australia (UniSA), UNICAMP (Universidade Estadual de Campinas), University of Toronto, Chinese University of Hong Kong, Cairo University, Telecom ParisTech, National Technical University of Athens, Indian Institute of Science, Indian Institute of Technology Bombay, Nanyang Technological University, University of Maryland, University of Illinois at Chicago, Ecole Polytechnique Federale de Lausanne, The Pennsylvania State University (Penn State), University of California Los Angeles, Massachusetts Institute of Technology, and University of Illinois at Urbana-Champaign. The organizers would love to sign on more locations. Please see their request for participation elsewhere in this issue.

As noted earlier, the Information Theory Society is also working to create a documentary about Shannon's life and work. The Board of Governors allocated seed funding for this project at their meeting in June, and we are working to raise the remainder of the funds through science funding agencies, foundations, companies, and individuals. Several proposals are currently under review, but further funds will be required. I would love to hear from companies, foundations, and individuals interested in helping with this important and historic project.

As always, I want to thank all of the Society's many volunteers; the Society simply wouldn't exist without you. If you are not currently an active participant in the society, I strongly encourage you to get involved; I am happy to help you find a way to engage that matches your interests and availability. The vibrance of our field and community depend both on technical contributions and on the time, energy, and ideas of our members. I welcome your questions, comments, and suggestions. Please contact me at effros@caltech.edu