Automated Reasoning by Convex Optimization

Proof Simplicity, Duality and Sparsity

Chee Wei Tan

Table of Contents

Hilbert's Twenty-Fourth Problem

"Criteria of simplicity, or proof of the greatest simplicity of certain proofs. Develop a theory of the method of proof in mathematics in general. Under a given set of conditions there can be but one simplest proof."

The Problem

Design a scalable algorithm to automatically construct the simplest analytic proof or disproof for the given information inequality.

S. W. Ho, L. Ling, C. W. Tan and R. W. Yeung, Proving and Disproving Information Inequalities: Theory and Scalable Algorithms, IEEE Transactions on Information Theory, 2020.

Examples

Prove: $H(X,Y,Z) - H(X|Y,Z) - H(Y|X,Z) - H(Z|X,Y) \geq 0$

Proof #1

$$= I(X;Y) + I(X;Z|Y) + I(Y;Z|X) \geq 0$$

Proof #2

$$\begin{aligned} = {} & 0.8(I(X;Y) + I(X;Z|Y) + I(Y;Z|X)) + \\ & 0.1(I(X;Z) + I(X;Y|Z) + I(Y;Z|X)) + \\ & 0.1(I(Y;Z) + I(X;Z|Y) + I(X;Y|Z)) \geq 0 \end{aligned}$$

Proof Simplicity

What is the “criteria of simplicity”?

> “The elegance of a mathematical theorem is directly proportional to the number of independent ideas one can see in the theorem and inversely proportional to the effort it takes to see them”
>
> — George Pólya

In our problem, we can quantify the simplicity of a proof by the number of elemental inequalities involved.

Table of Contents

A Toy Example

$$H(A|B) \leq H(A)$$

Rewrite using joint-entropies

$$H(A) + H(B) - H(A,B) \geq 0$$

Define $\mathbf{h} = \begin{bmatrix} H(A) \\ H(B) \\ H(A,B) \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$

Prove

$$\mathbf{b}^T \mathbf{h} \geq 0$$

Shannon's Elemental Inequalities Constraints

$\mathbf{h} = \begin{bmatrix} H(A) \\ H(B) \\ H(A,B) \end{bmatrix}$ should follow some constraints. e.g.,

$$\begin{aligned} I(A;B) = H(A) + H(B) - H(A,B) &\geq 0 \\ H(A|B) = H(A,B) - H(B) &\geq 0 \\ H(B|A) = H(A,B) - H(A) &\geq 0 \end{aligned}$$

We group them into constraint matrix **D** ($\mathbf{Dh} \geq \mathbf{0}$).

Problem specific constraints: **E** ($\mathbf{Eh} = \mathbf{0}$).

The Primal and Dual Linear Programs

Primal:

$$\begin{array}{ll} \min & \mathbf{b}^T\mathbf{h} \\ \text{s.t.} & \mathbf{D}\mathbf{h} \geq \mathbf{0} \\ & \mathbf{E}\mathbf{h} = \mathbf{0} \\ \text{var.} & \mathbf{h} \end{array}$$

$\Uparrow$

Geometric aspect

$\Downarrow$

Verify the information inequality

Dual:

$$\begin{array}{ll} \max & \mathbf{y}^T\mathbf{0} \\ \text{s.t.} & \mathbf{D}^T\mathbf{y} = \mathbf{b} + \mathbf{E}^T\boldsymbol{\mu} \\ & \mathbf{y} \geq \mathbf{0} \\ \text{var.} & \mathbf{y}, \boldsymbol{\mu} \end{array}$$

$\Uparrow$

Algebraic aspect

$\Downarrow$

Construct an analytic proof for the information inequality

Geometric Aspect (The Primal Problem)

$$p^* = \min_{\mathbf{h}} \{\mathbf{b}^T\mathbf{h} | \mathbf{D}\mathbf{h} \geq \mathbf{0}, \mathbf{E}\mathbf{h} = \mathbf{0}\}$$

- $p^* \geq 0 \quad \Longrightarrow \quad$ The inequality is true (it's Shannon-type).
- $p^* < 0 \quad \Longrightarrow \quad$ The inequality is false or it's a non-Shannon-type inequality.

A verifier that *verifies* information inequalities.

Algebraic Aspect (The Dual Problem)

Dual constraints:

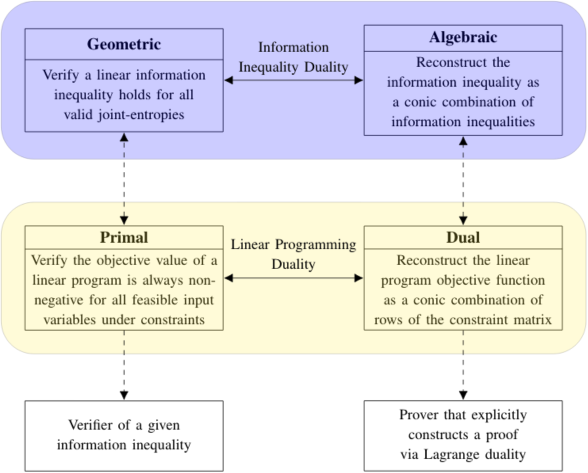
$$\mathbf{D}^T\mathbf{y} = \mathbf{b} + \mathbf{E}\mu$$
$$\mathbf{y} \geq \mathbf{0}$$

Intuition: If the dual problem is feasible, reconstruct the original inequality as

$$\mathbf{b}^T\mathbf{h} = \mathbf{y}^{*T}\mathbf{D}\mathbf{h} - \mu^{*T}\mathbf{E}\mathbf{h}$$
$$\geq \mathbf{0} \quad \text{(the proof!)}$$

A prover that constructs analytic proofs to information inequalities.

Linear Programming Framework



Back to the Toy Example

Prove $H(A|B) \leq H(A) \Longrightarrow \mathbf{b}^T\mathbf{h} \geq 0$,

where $\mathbf{h} = \begin{bmatrix} H(A) \\ H(B) \\ H(A,B) \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$ and $\mathbf{D} = \begin{bmatrix} 1 & 1 & -1 \\ 0 & -1 & 1 \\ -1 & 0 & 1 \end{bmatrix}$

Optimal dual solution: $\mathbf{y} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$

The input inequality can be written as the first row of $\mathbf{D}$, which is non-negative. (Recall that $\mathbf{b} = \mathbf{D}^T\mathbf{y}$, $\mathbf{y} \geq \mathbf{0}$ and $\mathbf{D}^T\mathbf{h} \geq \mathbf{0}$)

Generated proof:

$$H(A|B) \leq H(A) \Rightarrow H(A) + H(B) - H(A,B) \geq 0 \Rightarrow I(A;B) \geq 0$$

Table of Contents

Definition

A *shortest proof* of an information inequality is considered as a proof involving the least number of elemental inequalities. For a given Shannon-type information inequality, there often exist multiple shortest proofs.

Obtaining a Shortest Proof

$$\begin{aligned} \min \quad & \| \begin{bmatrix} \mathbf{y}^T & \mu^T \end{bmatrix}^T \|_0 \\ \text{s.t.} \quad & \mathbf{D}^T\mathbf{y} = \mathbf{b} + \mathbf{E}^T\mu, \ \mathbf{y} \geq \mathbf{0} \end{aligned}$$

Convex Relaxation

$$\begin{aligned} \min \quad & \mathbf{1}^T\mathbf{y} + \mathbf{1}^T\mathbf{z} \\ \text{s.t.} \quad & \mathbf{D}^T\mathbf{y} = \mathbf{b} + \mathbf{E}^T\mu, \ \mathbf{y} \geq \mathbf{0} \\ & -\mathbf{z} \leq \mu \leq \mathbf{z} \end{aligned}$$

Table of Contents

Scalability Issue

- We need vertex solutions to generate short, concise proofs or counter-examples. $\Longrightarrow$ Simplex-based methods
- The simplex method has exponential worst case time complexity, while the size of our LP grows exponentially with n. $\Longrightarrow$ Doubly exponential complexity
- The LP is sparse and highly degenerate, so the performance of the simplex method deteriorates.

Solution: Use iterative algorithms and perform “crossover” as a post-processing step.

Problem Reformulation

$$\begin{array}{ll} \min & \mathbf{b}^T\mathbf{h} \\ \text{s.t.} & \mathbf{B}\mathbf{h} + \mathbf{y} = \mathbf{c} \\ \text{var.} & \mathbf{h}, \mathbf{y}, \end{array}$$

where $\mathbf{B} = \begin{bmatrix} \mathbf{D} \\ \mathbf{D} \\ \mathbf{E} \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} -\mathbf{u} \\ \mathbf{v} \\ \mathbf{0} \end{bmatrix}$, $\mathbf{u}, \mathbf{v} \geq \mathbf{0}$ and $\mathbf{c} = \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \\ \mathbf{0} \end{bmatrix}$.

The ρ-augmented Lagrangian is

$$L_\rho = \mathbf{b}^T\mathbf{h} + \nu^T(\mathbf{B}\mathbf{h} + \mathbf{y} - \mathbf{c}) + \frac{\rho}{2}||\mathbf{B}\mathbf{h} + \mathbf{y} - \mathbf{c}||^2.$$

AITIP Algorithm

ALGORITHM 1: AITIP Algorithm

repeat

1. $\mathbf{h}$-update: $\mathbf{h}^{k+1} = -\frac{1}{\rho}(\mathbf{B}^T\mathbf{B})^{-1}(\mathbf{b} + \mathbf{B}^T\boldsymbol{\nu}^k + \rho\mathbf{B}^T\mathbf{y}^k - \rho\mathbf{B}^T\mathbf{c})$
2. $\mathbf{u}$-update: $\mathbf{u}^{k+1} = (\mathbf{D}\mathbf{h}^{k+1} + \frac{1}{\rho}\boldsymbol{\nu}_u^k)_+$
3. $\mathbf{v}$-update: $\mathbf{v}^{k+1} = (\mathbf{1} - \mathbf{D}\mathbf{h}^{k+1} - \frac{1}{\rho}\boldsymbol{\nu}_v^k)_+$
4. $\boldsymbol{\nu}$-update: $\boldsymbol{\nu}^{k+1} = \boldsymbol{\nu}^k + \rho(\mathbf{B}\mathbf{h}^{k+1} + \mathbf{y}^{k+1} - \mathbf{c})$

until *Stopping criteria is met*;

Each step has a closed-form solution.

Toward High Scalability

- Each step is just a (sparse) linear algebra computation, which can be easily parallelized to multiple cores or GPU (cuBLAS and cuSPARSE).
- The LLT (Cholesky) decomposition of the large matrix $\mathbf{B}^T\mathbf{B}$ matrix can be done beforehand. $\Longrightarrow$ Do a large part of the computation beforehand to improve the runtime performance.

These two points naturally lead to cloud computing.

Table of Contents

The AITIP Software-as-a-Service

`https://aitip.org`

Examples:

- `I(A1;A2)>=I(A1;A3)` s.t. `A1->A2->A3` (data processing inequality)
- `H(X1,X2,X3,X4)<=H(X1)+H(X2)+H(X3)+H(X4)` (independence bound for entropy)

Table of Contents

Open Issues and Future Work

- Solving large LPs is still challenging
  - Dimension grows exponentially
  - Highly degenerate problems
- The logical correctness in mathematical reasoning cannot be susceptible to computational issues
  - Inaccurate numerical approximation
  - Floating point errors
- An end-to-end method to explore the proof space, refine problem-specific constraints and to automatically construct a formal proof or valid counterexample requires more in-depth investigation