A framework for generating inequality conjectures[1]

Rahul Sarkar
Institute for Computational and Mathematical Engineering
Stanford University
(Joint work with Challenger Mishra & Subhayan Roy Moulik, University of Cambridge)

Online machine learning seminar
Sep 27, 2023

[1] *Mathematical conjecture generation using machine intelligence*. Challenger Mishra, Subhayan Roy Moulik, Rahul Sarkar (2023), arXiv:2306.07277

My main motivation

There exist a lot of examples in mathematics where one has a nice inequality derived under a specific set of assumptions.

One would like to know how a particular inequality generalizes / changes when some of the assumptions are slightly altered or replaced by another set of assumptions.

Can one design a scheme that helps a mathematician in such a search? The goal is to really make the daily job of an "inequality hunting" mathematician easier & faster. Also add a computational toolbox that they can leverage.

We make some preliminary headways towards this problem. Work in progress...

The space of relations

- Suppose $\mathcal{D}$ is a finite dimensional vector spaces over $\mathcal{R}$. We assume that $\mathcal{D}$ has dimension n.
- Let $\widehat{\mathcal{D}} \subseteq \mathcal{D}$ be a compact subset. We refer to $\widehat{\mathcal{D}}$ as the *feature space*
- Let $C(\widehat{\mathcal{D}})$ denote the Banach space of continuous, real valued functions on $\widehat{\mathcal{D}}$ equipped with the supremum norm, i.e. if $f \in C(\widehat{\mathcal{D}})$, then

$$\|f\|_{C(\widehat{\mathcal{D}})} := \sup_{x \in \widehat{\mathcal{D}}} |f(x)|.$$

- We define a *space of relations* $\mathcal{R} := C(\widehat{\mathcal{D}}) \times C(\widehat{\mathcal{D}})$, which is again a Banach space with the norm

$$\|(f,g)\|_{\mathcal{R}} := \|f\|_{C(\widehat{\mathcal{D}})} + \|g\|_{C(\widehat{\mathcal{D}})}, \quad (f,g) \in \mathcal{R}. \tag{1}$$

Inequality conjectures form a Banach manifold

What is a conjecture?

Definition

Let $(f, g) \in \mathcal{R}$. We say that the tuple (f, g) is a *conjecture* if and only if $f(x) < g(x)$ for all $x \in \widehat{\mathcal{D}}$. The set $\mathcal{C} := \mathcal{C}_< \sqcup \mathcal{C}_>$ is called the *space of conjectures*, where $\mathcal{C}_< := \{(f, g) \in \mathcal{R} : f(x) < g(x),\ \forall x \in \widehat{\mathcal{D}}\}$ and $\mathcal{C}_> := \{(f, g) \in \mathcal{R} : f(x) > g(x),\ \forall x \in \widehat{\mathcal{D}}\}$.

Manifold structure for the space of conjectures:

Lemma

$\mathcal{C}, \mathcal{C}_<$ and $\mathcal{C}_>$ are open subsets of $\mathcal{R}$ and each of them is a Banach manifold.

Why a Banach manifold

Idea of the proof is simple:

- $\mathcal{R}$ is a Banach manifold.
- Any open subset of $\mathcal{R}$ is a Banach manifold.
- The conditions $f < g$ and $f > g$ are *open conditions* (due to compactness of $\widehat{D}$).
- Thus $\mathcal{C}_>$ and $\mathcal{C}_<$ are both open sets.

Note: Because of the compactness assumption $\inf_{x \in \widehat{\mathcal{D}}} |f(x) - g(x)| = \epsilon > 0$, and is always attained by some $x \in \widehat{\mathcal{D}}$.

Group actions of the space of relations and conjectures

The space of conjectures $\mathcal{C}$ enjoys certain nice geometrical properties. Let $\mathrm{GL}(2,\mathbb{R})$ denote the set of all invertible 2×2 real valued matrices. We will denote $I := \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. If $A \in \mathrm{GL}(2,\mathbb{R})$ has the property that $A(f,g) \in \mathcal{C}$ for every $(f,g) \in \mathcal{C}$, we say that A leaves $\mathcal{C}$ *invariant*. In particular, we are interested in all the elements of $\mathrm{GL}(2,\mathbb{R})$ that leave $\mathcal{C}$ invariant. For every $\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} =: A \in \mathrm{GL}(2,\mathbb{R})$, we first define the linear map

$$A : \mathcal{R} \to \mathcal{R}, \quad (f,g) \mapsto (\bar{f},\bar{g}),$$
$$\begin{pmatrix} \bar{f} \\ \bar{g} \end{pmatrix} := \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} f \\ g \end{pmatrix}.$$

One can check that this defines a group action on $\mathcal{R}$

$$F : \mathrm{GL}(2,\mathbb{R}) \times \mathcal{R} \to \mathcal{R}, \quad (A,(f,g)) \mapsto A(f,g).$$

Moreover every $A \in \mathrm{GL}(2,\mathbb{R})$ defines a homeomorphism on $\mathcal{R}$

Which linear transformations leave $\mathcal{C}$ invariant?

Not every $A \in \mathrm{GL}(2,\mathbb{R})$ leaves $\mathcal{C}$ invariant. One can find two sets of non-trivial subgroups of $\mathrm{GL}(2,\mathbb{R})$, both of which leave $\mathcal{C}$ invariant:

Example 2.4 (Dilations)**.** Consider the group of positive dilations $\mathcal{T} := \{\lambda \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) : \mathbb{R} \ni \lambda > 0\}$. Then $\mathcal{T}$ is a subgroup of $\mathrm{GL}(2,\mathbb{R})$, and if $A \in \mathcal{T}$, it is clear that $A(f,g) \in \mathcal{C}$ if and only if $(f,g) \in \mathcal{C}$.

Example 2.5. Consider the 2-parameter set $\mathcal{H}$ defined by

$$\mathcal{H} := \left\{ \begin{pmatrix} p & q-1 \\ p-1 & q \end{pmatrix}, \; p,q \in \mathbb{R}, \; p+q \neq 1 \right\}. \tag{7}$$

The largest group of invertible matrices preserving $\mathcal{C}$

Let $\mathcal{G}$ be the largest set of invertible 2×2 matrices that leaves $\mathcal{C}$ invariant. Then $\mathcal{G}$ has a nice characterization:

Lemma

The set $\mathcal{G}$ satisfies the following properties:

(a) *Every element $A \in \mathcal{G}$ can be expressed in the form*

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a & c \\ 0 & b \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \quad a, b, c \in C(\widehat{D}), \;\; a(x), b(x) \neq 0, \;\; \textit{for any } x \in \widehat{D}.$$

(b) *$\mathcal{G}$ is a group, and a Banach manifold.*

Aside

One can also define the space of conjectures to be functions $f \in C(\widehat{D})$ such that $f(x) > 0$ or $f(x) < 0$ for all $x \in \widehat{D}$.

This leads to a slightly smaller search space (about 2x).

On the other hand, if the conjecture space has very complicated geometry, it may be better to have some redundancy in the parameterization (a form of relaxation for non-convex optimization problems).

We have been experimenting with both kinds of oracles. So far it is hard to say which one is better. Performance wise, this version is 2x faster.

The computational oracle: version zero

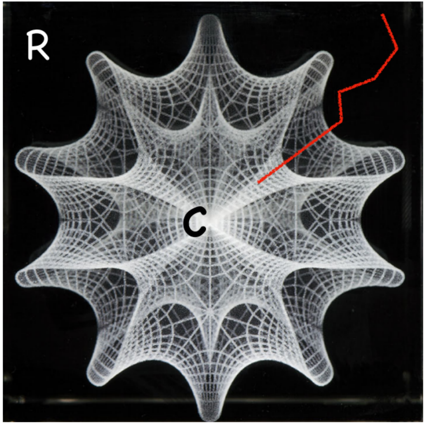
The conjecture space $\mathcal{C}$ is a subset of the space of relations.

Basic idea: Start from some $(f, g) \in \mathcal{R}$, and perform gradient descent to converge to some point in $\mathcal{C}$, by minimizing a loss function that is non-zero when $(f, g) \notin \mathcal{C}$.

Algorithm 1 Oracle

1: **Inputs:** *Mathematical features* $\{x_i\}_{i=1}^N$, *function class* $\mathcal{F}_d$, and *hyperparameters*: tolerance (tol), batch-size (b), maximum epochs (emax), learning rate (η).
2: $\theta \leftarrow$ random real vector.
3: **Parameterise:** $c_\theta := (f_\theta, g_\theta) := \mathcal{P}(\theta)$; $f_\theta, g_\theta \in \mathcal{F}(K^{G_f}[x])$.
4: **for** $i \leftarrow 1$ to emax **do**
5:     **for** $j \leftarrow 1$ to b **do**
6:         **if** $\mathcal{L}(c_\theta) \leq$ tol **then**
7:             return θ
8:         $\theta \leftarrow \theta - \eta\, \mathcal{M}_\theta^{-1} \nabla \mathcal{L}(\theta)$
9:         $\omega(c_\theta) \leftarrow \frac{2}{b} \sum_{i \in \mathrm{rand}} \mathrm{sgn}(f_\theta(x_i) - g_\theta(x_i))$
10:        $\mathcal{L}(c_\theta) \leftarrow (1 - \omega(c_\theta)^2)^2$
11: **Output:** $c_\theta = \mathcal{P}(\theta)$; $f_\theta < g_\theta$.

Cartoon visualization



Drawing courtesy Challenger Mishra.

Number theoretic conjectures

Some number theoretic conjectures about the prime counting function:

#	Conjecture (c_θ)
1	$\pi(ab) \geq \pi(a) + \pi(b)$
2	$\pi(ab) + \pi(a+b) + \pi(a) + \pi(b) \leq 2ab$
3	$4(\pi(a) + \pi(b)) + \pi(a+b) \leq 4\pi(ab)$
4	$\pi(ab) + 2\pi(a+b) \geq \pi(a) + \pi(b)$
5	$\pi(a+b) \leq \pi(a) + \pi(b)$
6	$\pi(a\ b\ c) \geq \pi(a)\ \pi(b)\ \pi(c)$
7†	$\pi(a\ b\ c \ldots) \geq \pi(a)\ \pi(b)\ \pi(c) \ldots$
8	$\pi(a+b+c) \leq \pi(a) + \pi(b) + \pi(c)$
9†	$\pi(a+b+c+\ldots) \leq \pi(a) + \pi(b) + \pi(c) + \ldots$
10	$\pi(a)\pi(b)\pi(c) \leq \sqrt{(\pi(a)+\pi(b)+\pi(c))^2 + (\pi(abc))^2}$
11†⋆	$(\ (\Sigma_i \pi(a_i))^2 + (\pi(\Pi_i a_i))^2)^{1/3} \leq \Pi_i \pi(a_i) \leq (\ (\Sigma_i \pi(a_i))^2 + (\pi(\Pi_i a_i))^2\ )^{1/2}$
12	$\pi(ab+bc+ca)^3 \geq 2\pi(abc)^2 + \pi(abc) + \pi(a+b+c)^2$
13	$\pi(a+b+c)^2 + \pi(abc) \geq \pi(ab+bc+ca)^3 + 2\pi(abc)^2$
14	$\pi(ab+bc+ca)^7 \geq \pi(a+b+c)^7$
15	$\pi\left(\alpha_1\alpha_2\alpha_3\right)^2 + \pi\left(\alpha_1\alpha_2 + \alpha_2\alpha_3 + \alpha_3\alpha_1\right) \geq \pi\left(\alpha_1\alpha_2 + \alpha_2\alpha_3 + \alpha_3\alpha_1\right)^3 + \pi\left(\alpha_1 + \alpha_2 + \alpha_3\right)$
16	$\pi\left(\alpha_1\alpha_2\alpha_3\right)^3 + 4\pi\left(\alpha_1\alpha_2\alpha_3\right)^2 + 4\pi\left(\alpha_1+\alpha_2+\alpha_3\right)^3 + 3 + ldots \geq \pi\left(\alpha_1\alpha_2 + \ldots\right)$
17°	$\pi\left(\chi_2\right)^3 - \pi\left(\chi_3\right)^2 + \pi\left(\chi_3\right) \geq \pi\left(\chi_1\right)$
18	$\pi\left(\chi_1\right)^3 \geq \pi\left(\chi_2\right)^3$
19	$5\pi\left(\chi_2\right)^3 \geq \pi\left(\chi_1\right)$
20	$\pi\left(\chi_1\right) + \pi\left(\chi_3\right) < \pi\left(\chi_2\right)$
21	$5\pi\left(\pi\left(\chi_1\right)\right) + 2\pi\left(\pi\left(\chi_3\right)\right) \geq \pi\left(\pi\left(\chi_2\right)\right)$
22	$\pi\left(\pi\left(\chi_2\right)\right) \geq 2\pi\left(\pi\left(\chi_3\right)\right) + 3\pi\left(\pi\left(\chi_1\right)\right)$
22	$11\left(\pi(ab) + \frac{ab}{\log(ab)}\right) > 9\pi(a+b) + \frac{9(a+b)}{\log(a+b)}$
23	$\pi\left(x + \sqrt{x}\right) \leq 3\pi\left(x\right) + 1$
24 ††	$\pi\left(x\right)^2 > x^3 + 2x + 2$
25	$\pi\left(x + \sqrt{x}\right) < 3\pi(x)$
26	$\pi\left(x + \sqrt{x}\right) < \frac{12}{5}\pi(x) + 1$

A closer look at some of the conjectures

#	Conjecture (c_θ)
1	$\pi(ab) \geq \pi(a) + \pi(b)$
2	$\pi(ab) + \pi(a+b) + \pi(a) + \pi(b) \leq 2ab$
3	$4(\pi(a) + \pi(b)) + \pi(a+b) \leq 4\pi(ab)$
4	$\pi(ab) + 2\pi(a+b) \geq \pi(a) + \pi(b)$
5	$\pi(a+b) \leq \pi(a) + \pi(b)$
6	$\pi(a\ b\ c) \geq \pi(a)\ \pi(b)\ \pi(c)$
$7^\dagger$	$\pi(a\ b\ c \ldots) \geq \pi(a)\ \pi(b)\ \pi(c) \ldots$
8	$\pi(a+b+c) \leq \pi(a) + \pi(b) + \pi(c)$
$9^\dagger$	$\pi(a+b+c+\ldots) \leq \pi(a) + \pi(b) + \pi(c) + \ldots$

\# 5 on the list is the famous Second Hardy-Littlewood conjecture.
\# 1 on the list is a new conjecture.

An example new result

Theorem

If x, y are positive integers, then $\pi(xy) \geq \pi(x) + \pi(y), \quad \forall\, x, y \geq 17$. In addition, the inequality holds for all $2 \leq x, y < 17$.

Proof idea:

- Prove the case for $x, y \geq 17$. Uses the Rosser-Schoenfeld formula.
- Enumeratively check the cases $2 \leq x, y \leq 17$.

Proof of theorem

Proof.

By the Rosser-Schoenfeld formula, we have for any $\alpha \geq 1.25506$

$$\frac{x}{\log x} < \pi(x) < \frac{\alpha x}{\log x}, \quad \forall\, x \geq 17. \tag{2}$$

Now choose $\alpha = 1.26$. Then note that for all $x, y \geq 17$ we have: (i) $\log x < \frac{x}{2\alpha}$, and (ii) $\log x + \log y \leq \log x \log y$. Using the above facts, we get for all $x, y \geq 17$

$$\begin{aligned} \pi(x) + \pi(y) &< \alpha\left(\frac{x}{\log x} + \frac{y}{\log y}\right) = \alpha\left(\frac{x\log y + y\log x}{\log x \log y}\right) < \left(\frac{\frac{xy}{2} + \frac{xy}{2}}{\log x \log y}\right) \\ &= \left(\frac{xy}{\log x \log y}\right) \leq \left(\frac{xy}{\log x + \log y}\right) = \frac{xy}{\log(xy)} < \pi(xy). \end{aligned} \tag{3}$$

The case $2 \leq x, y < 17$ has been enumeratively checked using a computer. $\square$

Some new group theory conjectures

Conjectures: Cayley graphs of finite simple groups

$\mathcal{H} := \langle \sigma, \tau \rangle, \quad S = \{\sigma, \tau\}, \quad \mathcal{D} := \mathrm{diam}(\mathrm{Cay}(\mathcal{H}, S \cup S^{-1})).$

$\tau_1 := tr(\sigma), \quad \tau_2 := tr(\tau), \quad \mathcal{O}_1 := o(\sigma), \ \mathcal{O}_2 := o(\sigma).$

#	Conjecture (c_θ)
1	$\mathcal{D} \geq (\tau_1 + \tau_2)/2$
2	$\mathcal{D} \leq (\mathcal{O}_1 + \mathcal{O}_2)$
3	$\mathcal{D} \geq (\mathcal{O}_1 + \mathcal{O}_2)/4 + 5(\tau_1 + \tau_2)/12 - 0.120225 \log_2(G)$
4	$\mathcal{D} \geq (4.32809 \log(G) + 4\,(o_1 + o_2) + 5\,(\tau_1 + \tau_2))/3$

A new theorem for necessary conditions on generators of finite non-Abelian simple groups [He, Jejjala, CM, Sharnoff].

Babai '92: $\mathcal{D} \leq (\log_2|\mathcal{H}|)^c$, for some universal constant c.

What I'm working on: A problem from Fourier analysis

Basic question: When you are computing Fourier transform of a function under some nice assumptions (say Schwartz function), what kind of error do you make when you compute it using the DFT (discrete Fourier transform)?

Answer is well-known, and one of the lemmas that I use in one version of the proof is the following:

Lemma

Let $B := \{(x_1, \ldots, x_d) \subseteq \mathbb{R}^d : |x_i| \leq b_i,\ b_i > 0,\ \forall i = 1, \ldots, d\}$, and define $\delta := \max_i\{b_i\} / \min_i\{b_i\}$. Suppose $x \in \mathbb{R}^d \setminus int(2B)$ and $\omega \in B$, where $int(\cdot)$ denotes interior of a set. Then we have the estimate $|x| \leq 2\sqrt{1 + \delta^2(d-1)}|x + \omega|$.

One can ask general versions of this lemma

- What happens when you change the shape of the convex body (or bounded non-convex bodies) to something other than a cube?
- For example, consider the family of astroids in $\mathbb{R}^2$

$$|x/a|^{1/n} + |y/b|^{1/n} = 1, \quad 0 < n \in \mathbb{Z}, \quad a, b, > 0.$$

- What will be the dependence on dimension d?

One can ask all kinds of questions, and it is not clear the formula should generalize. But we have a starting point that we know is true, and would like to use it to find appropriate generalizations (or help us finding the right generalizations).

Outlook and limitations

1. How to connect to automatic theorem provers, and proof assistants?
2. Extend the framework to handle other kinds of conjectures (like equality conjectures) - finite precision issues.
3. Using the symmetries of the conjecture space to cut down the search space. We already do this by factoring in the dilation group. But what about other symmetries?
4. What to do if problem itself has inbuilt symmetry?
5. What to do when it is not easy to generate training data exactly (precision issues or numerical accuracy issues) – for example, say you wanted to recover the isoperimetric inequaltiies.

Questions

Contact (good till Feb 2024): rsarkar@stanford.edu

Soon to move to UC Berkeley, Math Department!