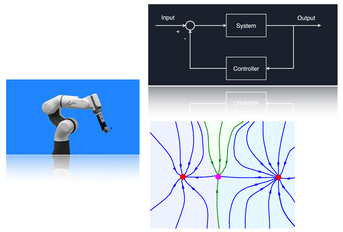
Learning Polynomial Problems with $SL(2,\mathbb{R})$-Equivariance

Hannah Lawrence*, Mitchell Tong Harris*
Massachusetts Institute of Technology

Polynomial Problems

Applications: (problems solved via "sum of squares" [Parrilo])
- operations research
- control theory
- certifying dynamical system stability
- robot path planning
- designing nonlinear controllers



Polynomial minimization

$$\alpha = \min_x \quad p(x)$$

Two machine learning pipelines:
- Predict minimum α
- Predict M to show $p - \alpha \geq 0$

Gives a "sum of squares" proof of lower bound if $M \succeq 0$

Certifying nonnegativity

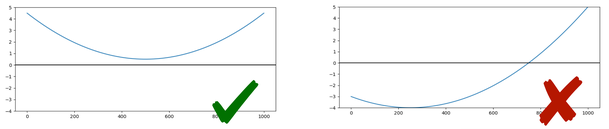
$$\max_{M \in S^{2d}} \log\det M$$

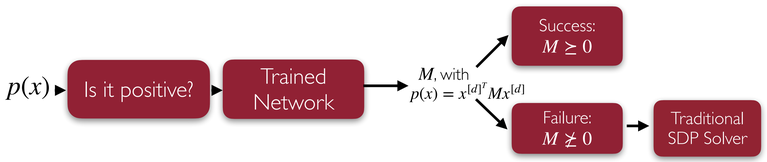
s.t. $x^{[d]^T} M x^{[d]} = p(x)$ and $M \succeq 0$

M "represents" p $\quad$ p is nonnegative

where $x^{[d]} = \begin{pmatrix} 1 & x & x^2 \cdots & x^d \end{pmatrix}^T$

Fundamental problem: is $p(x) \geq 0$ for all $x \in \mathcal{X}$?





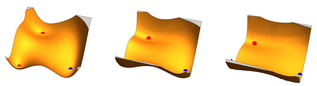
Symmetries

General Linear Group, $GL(2)$

$A \in \mathbb{R}^{2\times 2}$ s.t. $\det(A) \neq 0$

Special Linear Group, $SL(2)$

$A \in \mathbb{R}^{2\times 2}$ s.t. $\det(A) = 1$, i.e. coordinate changes that preserve area



Local and global minima of a polynomial after transformation by $A_1, A_{1.2}, A_{1.4}$ for

$$A_n = \begin{bmatrix} n & 0 \\ 0 & \frac{1}{n} \end{bmatrix} \in SL(2)$$

The minimum value is **invariant** to a change of coordinates.

If $g \in SL(2)$, define $g^{[d]}$ s.t. $(gx)^{[d]} = g^{[d]} x^{[d]}$.

M is the optimizer for $p(x)$

$\Longrightarrow$

$g^{[d]^T} M g^{[d]}$ is the optimizer for $p(gx)$.

The maximizer is **equivariant** to a change of coordinates.

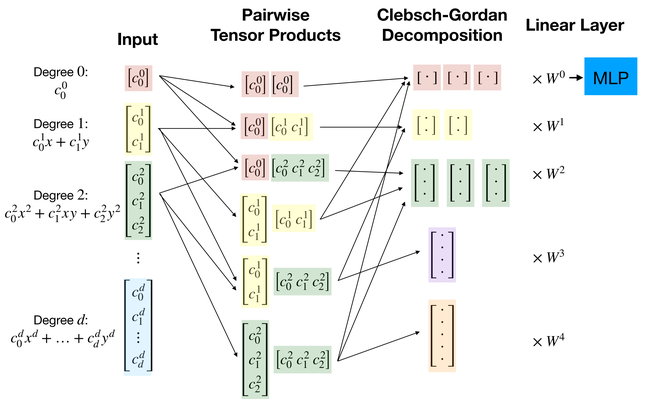
$SL(2)$-Equivariant Architecture

Key idea: decomposing tensor products into irreducible representations

- $V(d)$ = bivariate homogeneous polynomials (binary forms) of degree d = irreducible representations
  - For example: $V(2) = \{ax^2 + bxy + cy^2 : a,b,c \in \mathbb{R}\} = \{(a \quad b \quad c) : a,b,c \in \mathbb{R}\}$
- When $d_1 \geq d_2$, the Clebsch-Gordan decomposition,

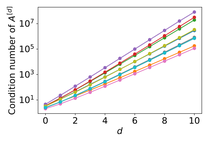
$$V(d_1) \otimes V(d_2) = V(d_1 + d_2) \oplus V(d_1 + d_2 - 2) \cdots \oplus V(d_1 - d_2),$$

can be computed explicitly by the classical transvectant from the invariant theory literature.



What's hard about non-compactness?

Representations of $SL(2)$ can be arbitrarily poorly conditioned



Here, each line is a randomly chosen element of $SL(2)$; d = degree of the representation ($d = 0$ is original 2×2 matrix). Condition number = max singular value / min singular value; high condition number means matrix-vector products lose numerical accuracy.

→ **condition number grows exponentially in d**

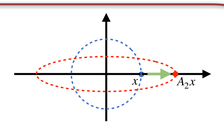
Non-compact equivariance = out-of-distribution robustness

There is no finite, invariant measure over $SL(2)$, so not all points in the orbit of a non-zero datapoint can be equally likely:

$\exists g, x : \mathbb{P}(x) \neq \mathbb{P}(gx)$

Example:

$$x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, A_n = \begin{bmatrix} n & 0 \\ 0 & \frac{1}{n} \end{bmatrix} \in SL(2)$$



In this example, the orbit of x is the entire x-axis, excluding the origin. There is no uniform measure on the whole x-axis, so some points in the orbit must be more likely than others.

→ Contrasts with most applications involving compact groups, such as rotations
→ Encouraging equivariance prioritizes datapoints that aren't necessarily as likely
→ Moreover, standard loss functions are not $SL(2)$-invariant

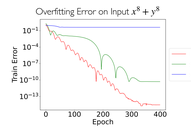
Impossibility Results

Theorem (Informal): There is a non-polynomial equivariant function that the $SL(2)$-equivariant architecture cannot approximate.

Proof. The problematic function is exactly the "certifying non-negativity" function we wished to learn,

$f(p) = \text{argmax}\, \text{logdet}\, Q$ s.t. $p(\vec{x}) = \vec{x}^{[d]^T} Q \vec{x}^{[d]}$ and $Q \succeq 0$.

We prove that the sparsity pattern of $f(x^8 + y^8)$ cannot be matched by this architecture, for any learned parameters. □



Corollary (Informal): There is an $SL(2)$-equivariant function that cannot be approximated by products of equivariant polynomials and arbitrary invariants.

Proof. The architecture can parametrize any equivariant polynomial multiplied by an arbitrary invariant [Bogatskiy et al], but this architecture cannot represent the stated equivariant function. □

Implication: there is no general, equivariant version of the Weierstrass approximation theorem for $SL(2,\mathbb{R})$.

Experiments

Degree	6	8	10	12	14
MLP NMSE	9.5e-6	6.0e-5	2.9e-5	2.3e-5	1.1e-5
MLP times (min)	0.062	0.082	0.17	0.22	0.29
SCS NMSE	2.7e-5	6.2e-5	1.2e-4	2.7e-4	1.2e-3
SCS times (min)	3.50	4.94	9.11	18.8	37.4

MLP is faster than a convex solver!

Methods:
- MLP with no augmentation, rotation-augmentation, or $SL(2)$-augmentation
  - $SL(2)$-augmentation distribution can vary by allowable condition number
- $SO(2)$-equivariant net with or without $SL(2)$-augmentation
- $SL(2)$-equivariant net

Dataset:
- Synthetically generated positive polynomials: $A_{ij} \sim N(0,1), p = \vec{x}^{[d]^T}(A^T A + 10^{-8} I)\vec{x}^{[d]}$
- Solve for max-determinant certificate using Mosek, a second-order convex solver
- Consider test distribution shift by randomly drawn $SL(2)$ matrices of varying conditioning

- On test instances drawn from the original distribution, MLP augmented by **rotations** is best
- MLP augmented by **well-conditioned** elements of $SL(2)$ is best under distribution shift
- Augmentation by **poorly-conditioned** elements of $SL(2)$ impedes performance

Conclusions

- Machine learned methods can substantially accelerate positivity verification
- The standard toolkit of equivariant polynomial approximation (via tensor products of irreps) does not suffice for $SL(2)$

Future Work

- Methods for $SL(2)$-equivariance that avoid the roadblock of polynomial approximation, e.g. frames
- Scaling to higher-degree polynomials
- Deployment on non-synthetic applications-driven data

References

1. Parrilo, Pablo A. Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. California Institute of Technology, 2000.
2. Bogatskiy, Alexander, et al. "Lorentz group equivariant neural network for particle physics." International Conference on Machine Learning. PMLR, 2020.

* equal contribution
hanlaw@mit.edu, mitchh@mit.edu