NeurIPS 2024

University of Southampton

Rethinking Deep Thinking

Stable Learning of Algorithms using Lipschitz Constraints

Jay Bear, Adam Prügel-Bennett, and Jonathon Hare

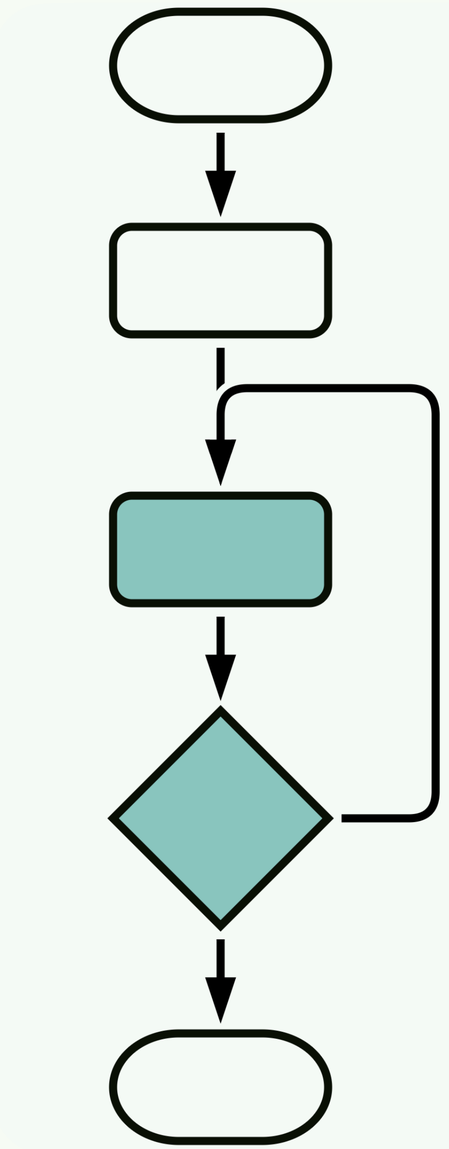
Motivation

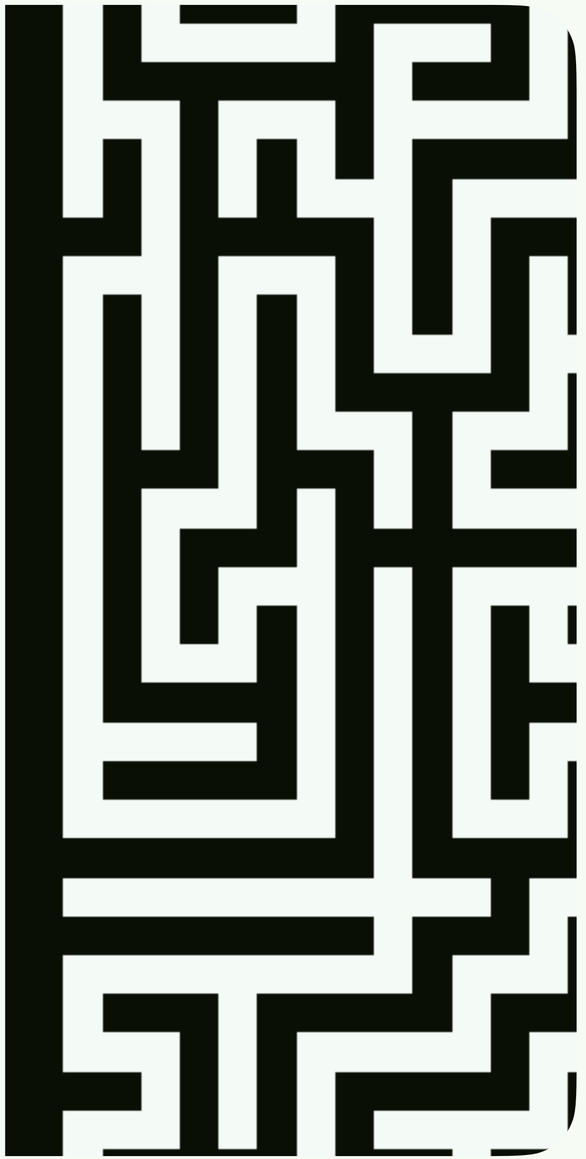
Iteration is a key ingredient in a vast number of important algorithms.

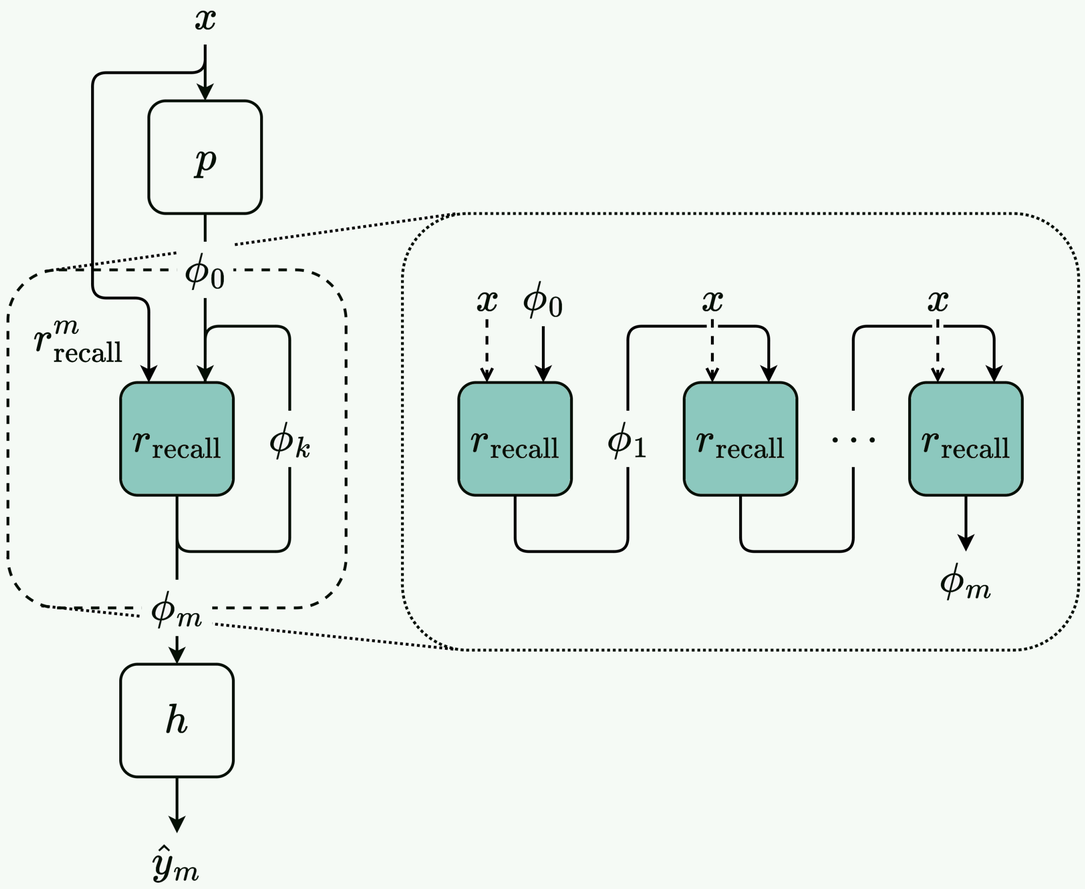
Incremental progress towards a solution is demonstrated in many of these.

Can deep learning models be used to learn the complex steps of algorithms?

Is there a way to guarantee a solution or approximation is reached?







Deep Thinking

Deep Thinking (DT) networks were designed to learn algorithms by using recurrence to induce iterative behavior.

Can be trained with few iterations on easy problems then solve harder problems by increasing the number of iterations.

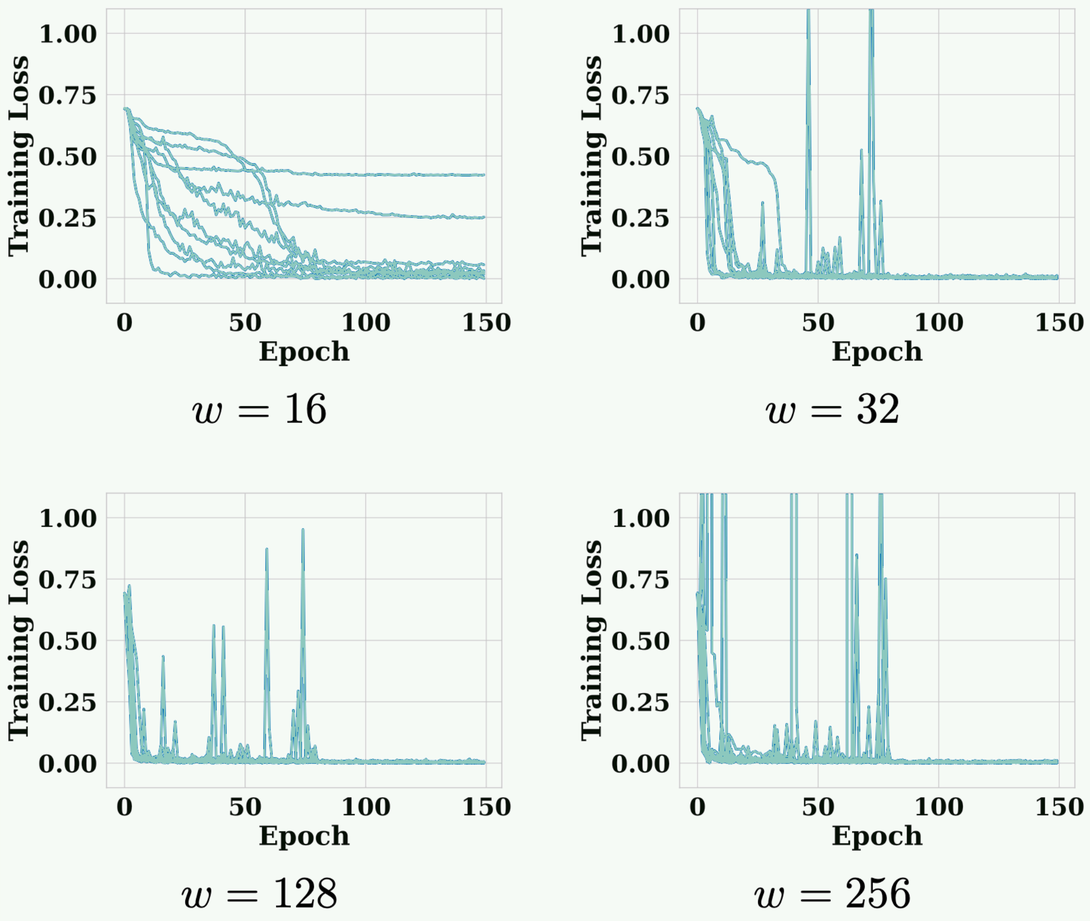
Improved by adding '**recall**' (DT-R) which mitigates decay in performance over extended iterations – **overthinking**.

DT Stability

We measured differing stability in training multiple DT-R models. Models would sometimes fail to learn.

Models with increased width (recurrent channels) often **exploded** or resulted in NaN errors.

Extrapolation (increased iterations) often resulted in explosion even with models that trained well.



Lipschitz Constraints

We propose constraining the minimum **Lipschitz constant** of the recurrent block (denoted G) to be just less than 1, making it a **contraction mapping**.

Keeping it close to 1 motivates incremental changes.

Constrain each composed element (layers, activation functions, *etc.*) to ensure this.

We call this model '**Deep Thinking with Lipschitz Constraints**' (DT-L).

Lipschitz Constant

Let $f : \mathbb{R}^n \to \mathbb{R}^m$. If, for all $x_1, x_2 \in \mathbb{R}^n$,

$$\left\| f(x_1) - f(x_2) \right\| \leq K \left\| x_1 - x_2 \right\|$$

then f is Lipschitz continuous with Lipschitz constant K.

$\exists K < 1 \Rightarrow f$ is a contraction mapping.

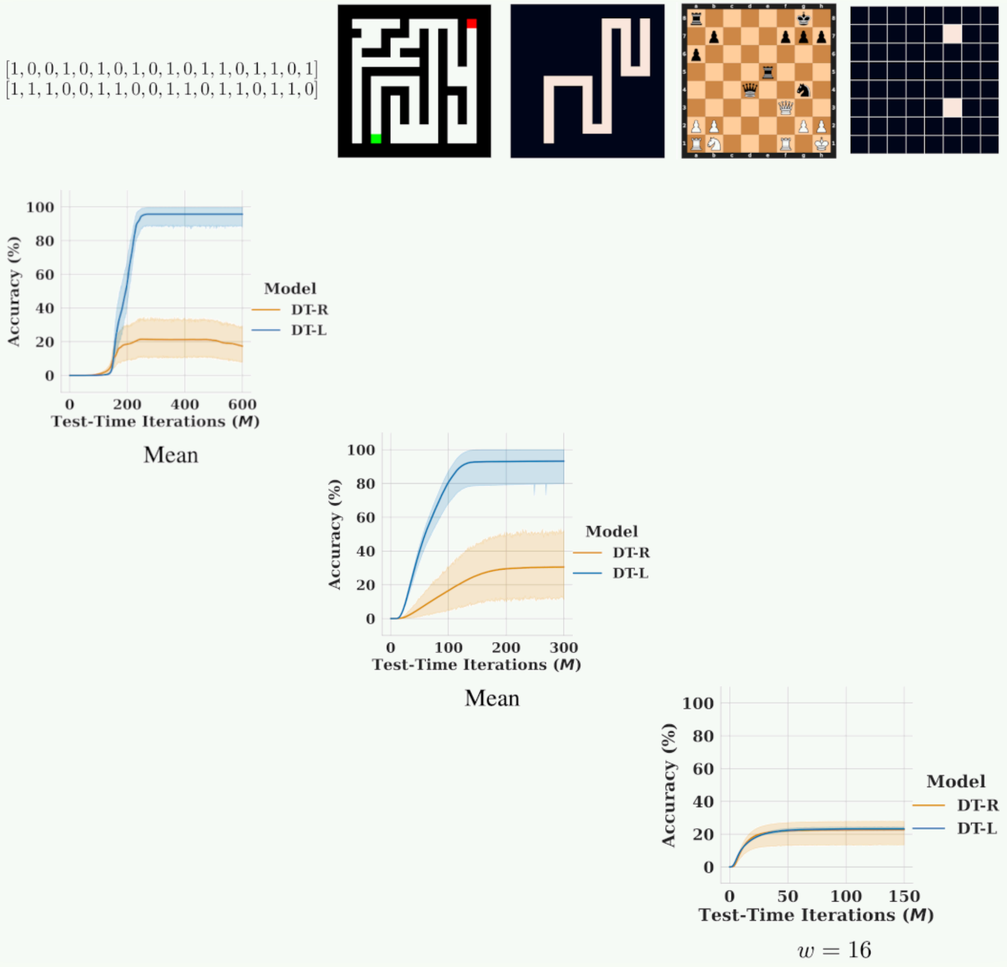
Experiments and Results

We test on all 3 Easy-to-Hard problems; **prefix sums**, **mazes**, and **chess puzzles**.

DT-L shows increased average accuracy for randomly-initialized models.

Chess puzzles showed no improvement – an interesting phenomenon.

No measured cases of explosion.



Questions?

We'll be presenting a poster at **NeurIPS 2024** in Vancouver.

Alternatively, contact us via email.