When Algorithms Learn: Discrete Optimization Meets Machine Learning

Sebastian Pokutta

Technische Universität Berlin
and
Zuse Institute Berlin

pokutta@math.tu-berlin.de
@spokutta

ISCO 2026

May 8th, 2026 · Kuşadası, Turkey

ZUSE INSTITUTE BERLIN

Berlin Mathematics Research Center MATH+

What is this talk about?

Introduction

A personal, highly-biased, and incomplete take of what AI can do in Mathematics et al.

What is this talk about?

Introduction

A personal, highly-biased, and incomplete take of what AI can do in Mathematics et al.

Why? AI methods increasingly shape how scientific discovery is performed, configured, and accelerated.

What is this talk about?

Introduction

A personal, highly-biased, and incomplete take of what AI can do in Mathematics et al.

Why? AI methods increasingly shape how scientific discovery is performed, configured, and accelerated.

Today: Two recent use-cases / perspectives from my own work.

What is this talk about?

Introduction

A personal, highly-biased, and incomplete take of what AI can do in Mathematics et al.

Why? AI methods increasingly shape how scientific discovery is performed, configured, and accelerated.

Today: Two recent use-cases / perspectives from my own work.

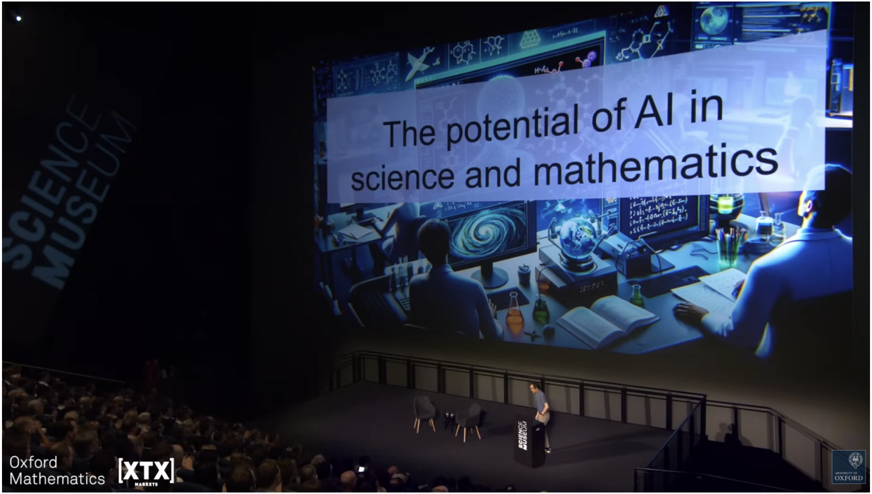
Outline

- A bit of high-level perspective
- Agentic AI Researcher framework
- Neural Colorings of the plane

(Hyperlinked) References are not exhaustive; check references contained therein.

What is this talk about?

AI and Mathematics et al



[The Potential for AI in Science and Mathematics - Terence Tao]

What is this talk about?

AI and Mathematics et al

Various levels of co-creation. [Haase and Pokutta, 2026]

- Digital Pen: basically like autocorrect, bibtex lookup, etc. “2000s”
- AI Task Specialist: ChatGPT, Claude, Gemini, etc. 2022 - 2025
- AI Assistant: Agents with integrated tools, verification, etc. 2025 -
- AI Co-creator: Fully integrated, autonomous, co-creator 2027(??) -

What is this talk about?

AI and Mathematics et al

Various levels of co-creation. [Haase and Pokutta, 2026]

- Digital Pen: basically like autocorrect, bibtex lookup, etc. “2000s”
- AI Task Specialist: ChatGPT, Claude, Gemini, etc. 2022 - 2025
- AI Assistant: Agents with integrated tools, verification, etc. 2025 -
- AI Co-creator: Fully integrated, autonomous, co-creator 2027(??) -

Broadly. Two categories: (a) co-creation agent and (b) tool in larger system.

What is this talk about?

AI and Mathematics et al

Various levels of co-creation. [Haase and Pokutta, 2026]

- Digital Pen: basically like autocorrect, bibtex lookup, etc. "2000s"
- AI Task Specialist: ChatGPT, Claude, Gemini, etc. 2022 - 2025
- AI Assistant: Agents with integrated tools, verification, etc. 2025 -
- AI Co-creator: Fully integrated, autonomous, co-creator 2027(??) -

Broadly. Two categories: (a) co-creation agent and (b) tool in larger system.

Capabilities are impressive but unstable:

- SOTA models achieve post-PhD level scores on benchmarks, yet in day-to-day use make trivial (logical) errors.
- No hard verification of results and randomness across runs.
- Prompting and scaffolding are still a challenge.
- Availability of tools for verification etc crucial.

What is this talk about?

Mathematics with computers is not new

Various high-profile examples from the past.

- Four Color Theorem: massive computer-based case checking

  [Appel and Haken, 1977, Robertson et al., 1997]

- Kepler Conjecture / Hales' Theorem: extensive computer verification

  [Hales et al., 2017]

- Classification of Finite Simple Groups: Formal verification with Lean/Coq
- Boolean Pythagorean Triples Problem: A spectacular 200TB SAT-solver proof.

  [Heule et al., 2016]

What is this talk about?

Mathematics with computers is not new

Various high-profile examples from the past.

- Four Color Theorem: massive computer-based case checking

  [Appel and Haken, 1977, Robertson et al., 1997]

- Kepler Conjecture / Hales' Theorem: extensive computer verification

  [Hales et al., 2017]

- Classification of Finite Simple Groups: Formal verification with Lean/Coq
- Boolean Pythagorean Triples Problem: A spectacular 200TB SAT-solver proof.

  [Heule et al., 2016]

Crucial role in computational mathematics / scientific computing

- Finite Elements
- Numerical Simulations
- Optimization
- Engineering
- ...

The Agentic Researcher

A Practical Guide to AI-Assisted Research
in Mathematics and Machine Learning

joint work with: Max Zimmer, Nico Pelleriti,
Christophe Roux

preprint (2026)
https://arxiv.org/abs/2603.15914

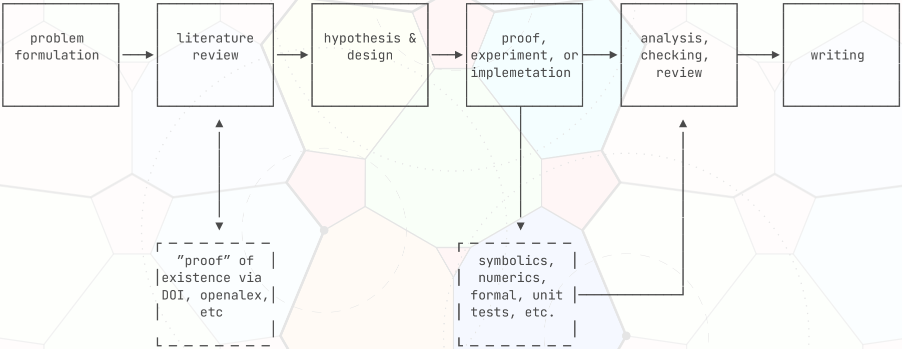
Partially supported by ExC MATH+ Project EF-LiOpt-3
Agent AI in Mathematics

[Zimmer et al., 2026]

AI in the Scientific Discovery Workflow

The agentic researcher framework

AI systems can be used at almost any stage of the scientific workflow.

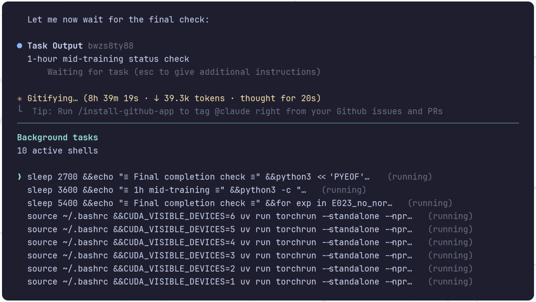


(workflow simplified. for illustration only.)

Note. The researcher needs to be in charge and responsible(!!) throughout.

Building a practical tool

The agentic researcher framework



Screenshot from a run of the framework.

- Born out of the MATH+ project *Agentic AI in Mathematics*.
- Refined over roughly 1.5 years of day-to-day research use.
- Sandboxed CLI agents with project instructions, tools, Git, LaTeX, and GPU-backed experiments.
- Puts a premium on verification: record everything, verify citations, and verify before claiming.

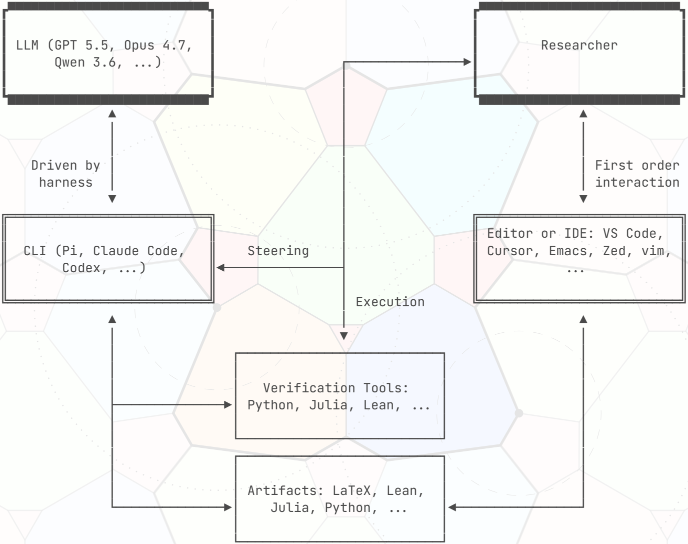
[blog post] [arXiv] [GitHub]

Does it works? Several real-world use-cases in paper

1. Convergence lower bounds for Frank-Wolfe on uniformly convex sets
2. Multi-Variable Dual Tightening for Boscia.jl for MINLPs
3. Weight Reconstruction in LLM pruning
4. ...

Interacting with the System

The agentic researcher framework



Orchestration Dashboard

The agentic researcher framework

ai-box dashboard

7 containers (1 busy, 6 idle) | 1 service | 9 IPC sessions | 19:00:01

Tool Containers

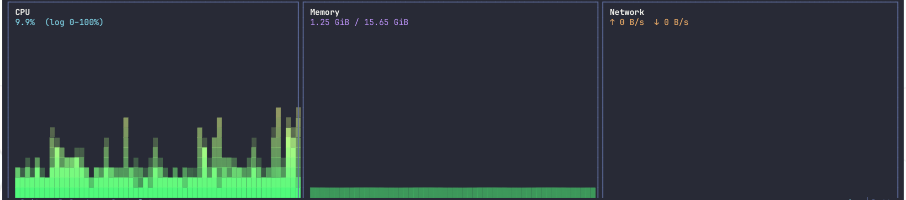
Status	Tool	Container	CPU	Mem	Uptime	Workspace	Entry
● BUSY	bash	ai-box-bash-13804cfe	8.9%	88.76MiB	4d 22h	/workspace/project-monitor	bash -c exec uv run /app/dashboard.py
● IDLE	claude	ai-box-claude-931bce30	0.1%	238.6MiB	4d 11h	~/research/helper-tools	claude --resume session-summary --enable-auto-mode
● IDLE	codex	ai-box-codex-d3b96bc7	0.0%	157.7MiB	20h 48m	~/research/paper-alpha	codex resume 019d8cdd-d0b1-74a3-b766-eb1215fb8802
● IDLE	pi	ai-box-pi-dd9605c9	0.0%	115.9MiB	20h 09m	~/research/helper-tools	pi
● IDLE	codex	ai-box-codex-8356c14f	0.0%	64.22MiB	11h 22m	~/research/paper-beta	codex
● IDLE	codex	ai-box-codex-28c9612b	0.0%	93.67MiB	10h 43m	~/research/talk-prep	codex
● IDLE	claude	ai-box-claude-22319677	0.2%	400.4MiB	8h 02m	~/research/helper-tools	claude --enable-auto-mode --resume e91045dc-bd30-4f47-93e

Services

Status	Name	Image	CPU	Mem	Uptime	Ports
● IDLE	qdrant	qdrant/qdrant	0.6%	120.6MiB	4d 22h	6333-6334

IPC Sessions

ID	Topic	Participants	Msgs	Files	Latest
0405ac53	Formal verification feasibility review	agent-a, agent-b	1	1	agent-b: Feasibility brief attached.
100e9e6f	Interactive article draft with embedded widgets	assistant, researcher	12	14	researcher: Updated draft attached. References ur
1d72f106	Research note restructuring	agent-a, researcher	3	3	agent-a: Restructured draft attached. Key changes
7b7b2528	Homepage layout feedback	agent-a	2	0	agent-a: The homepage agent started a related ses
859f8622	Draft editing round	workspace-agent	2	6	workspace-agent: Updated draft attached. Changes
9148b790	Conference talk details	slide-agent	5	1	slide-agent: Changed the active highlight and spa
d8b2e00c	Intro page content review	agent-a, agent-c	6	0	agent-a: Intro update request queued for the next
dc6393f2	Reference cleanup	agent-d, agent-a	2	2	agent-a: Cleaned bibliography attached.
f358b0e2	Publication wording review	workspace-agent	1	1	workspace-agent: Review notes attached.



q Quit r Refresh s Sort l Log ^p palette

Dashboard showing multiple containerized agents interacting with each other (stylized example).

Frank-Wolfe Beyond 1/t Convergence

preprint
https://arxiv.org/abs/2604.28006

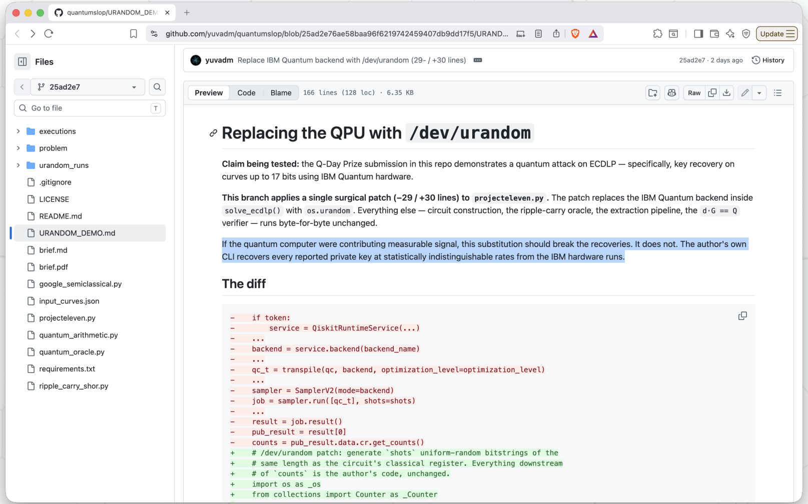
Partially supported by ExC MATH+ Project EF-LiOpt-3

Agent AI in Mathematics

[Pokutta, 2026]

Why Verification is Crucial

The agentic researcher framework


github.com/yuvadm/quantumslop/blob/25ad2e76ae58baa96f6219742459407db9dd17f5/URAND...

yuvadm Replace IBM Quantum backend with /dev/urandom (29- / +30 lines) 25ad2e7 · 2 days ago History

Preview Code Blame 166 lines (128 loc) · 6.35 KB

Replacing the QPU with `/dev/urandom`

Claim being tested: the Q-Day Prize submission in this repo demonstrates a quantum attack on ECDLP — specifically, key recovery on curves up to 17 bits using IBM Quantum hardware.

This branch applies a single surgical patch (−29 / +30 lines) to `projecteleven.py`. The patch replaces the IBM Quantum backend inside `solve_ecdlp()` with `os.urandom`. Everything else — circuit construction, the ripple-carry oracle, the extraction pipeline, the `d·G == Q` verifier — runs byte-for-byte unchanged.

If the quantum computer were contributing measurable signal, this substitution should break the recoveries. It does not. The author's own CLI recovers every reported private key at statistically indistinguishable rates from the IBM hardware runs.

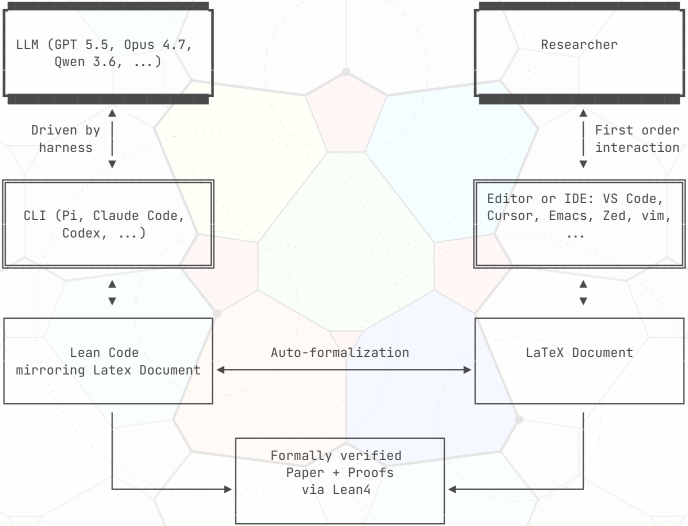
The diff

```
-    if token:
-        service = QiskitRuntimeService(...)
-    ...
-    backend = service.backend(backend_name)
-    ...
-    qc_t = transpile(qc, backend, optimization_level=optimization_level)
-    ...
-    sampler = SamplerV2(mode=backend)
-    job = sampler.run([qc_t], shots=shots)
-    ...
-    result = job.result()
-    pub_result = result[0]
-    counts = pub_result.data.cr.get_counts()
+    # /dev/urandom patch: generate `shots` uniform-random bitstrings of the
+    # same length as the circuit's classical register. Everything downstream
+    # of `counts` is the author's code, unchanged.
+    import os as _os
+    from collections import Counter as _Counter
```

[GitHub Repository]

End-2-End Formal Verification with Lean 4

The agentic researcher framework



When generation is cheap, verification is everything.

[Pokutta, 2026]

The Hadwiger-Nelson Problem

joint work with: Aldo Kiem, Konrad Mundinger,
Christoph Spiegel, Max Zimmer

ICML 2025 (oral)
https://arxiv.org/abs/2404.05509

Partially supported by ExC MATH+ Project EF-LiOpt-3

Agent AI in Mathematics

[Mundinger et al., 2025]

The Hadwiger-Nelson Problem

Problem (Nelson 1950, also: Gardner, Moser, Erdős, Harary, Tutte, ...)

What is the smallest number of colors sufficient for coloring the plane in such a way that no two points of the same color are at a unit distance apart?

The Hadwiger-Nelson Problem

Problem (Nelson 1950, also: Gardner, Moser, Erdős, Harary, Tutte, ...)

What is the smallest number of colors sufficient for coloring the plane in such a way that no two points of the same color are at a unit distance apart?

Infinite graph with vertex set $\mathbb{E}^2$ and edges $\{x, y\}$ for any $x, y \in \mathbb{E}^2$ with $\|x - y\| = 1$

$\Rightarrow$ chromatic number of the plane $\chi(\mathbb{E}^2)$

The Hadwiger-Nelson Problem

Problem (Nelson 1950, also: Gardner, Moser, Erdős, Harary, Tutte, ...)

What is the smallest number of colors sufficient for coloring the plane in such a way that no two points of the same color are at a unit distance apart?

Infinite graph with vertex set $\mathbb{E}^2$ and edges $\{x, y\}$ for any $x, y \in \mathbb{E}^2$ with $\|x - y\| = 1$
$\Rightarrow$ chromatic number of the plane $\chi(\mathbb{E}^2)$

Theorem

Assuming Axiom of Choice (AoC): [Bruijn and Erdos, 1951]

Any graph is k-colorable iff every finite subgraph of it is k-colorable.

The Hadwiger-Nelson Problem

Problem (Nelson 1950, also: Gardner, Moser, Erdős, Harary, Tutte, ...)

What is the smallest number of colors sufficient for coloring the plane in such a way that no two points of the same color are at a unit distance apart?

Infinite graph with vertex set $\mathbb{E}^2$ and edges $\{x, y\}$ for any $x, y \in \mathbb{E}^2$ with $\|x - y\| = 1$

$\Rightarrow$ chromatic number of the plane $\chi(\mathbb{E}^2)$

Theorem

Assuming Axiom of Choice (AoC): [Bruijn and Erdos, 1951]

Any graph is k-colorable iff every finite subgraph of it is k-colorable.

This problem has a long and complicated history... over 14 pages in [Soifer, 2024]

History

The Hadwiger-Nelson Problem

Table 3.1 Who created the chromatic number of the plane problem?

Publication	Year	Author(s)	Problem creator(s) or source named
[Gar2]	1960	Gardner	"**Leo Moser** ...writes..."
[Had4]	1961	Hadwiger (after Klee)	**Nelson**
[E61.22]	1961	Erdős	"I cannot trace the origin of this problem"
[Cro]	1967	Croft	"A long[18]-standing open problem of **Erdős**"
[Woo1]	1973	Woodall	**Gardner**
[Sim]	1976	Simmons	**Erdős**, **Harary**, and **Tutte**
[E80.38] [E81.23] [E81.26]	1980–1981	Erdős	**Hadwiger** and **Nelson**
[CFG]	1991	Croft, Falconer, and Guy	"Apparently due to **E. Nelson**"
[KW]	1991	Klee and Wagon	"Posed in 1960–61 by **M. Gardner** and **Hadwiger**"

p. 24 in [Soifer, 2024]

Lower bounds on $\chi(\mathbb{E}^2)$

Lower bounds through unit distance graphs

The Hadwiger-Nelson Problem

Find unit distance graphs of large chromatic number.

Lower bounds through unit distance graphs

The Hadwiger-Nelson Problem

Find unit distance graphs of large chromatic number.

Definition

A graph $G = (V, E)$ is a unit distance graph if there exists an embedding $f : V \to \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only $\{u, v\} \in E$.

Lower bounds through unit distance graphs

The Hadwiger-Nelson Problem

Find unit distance graphs of large chromatic number.

Definition

A graph $G = (V, E)$ is a unit distance graph if there exists an embedding $f : V \to \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only $\{u, v\} \in E$.

A triangle gives a lower bound of 3.

Find unit distance graphs of large chromatic number.

Definition

A graph $G = (V, E)$ is a unit distance graph if there exists an embedding $f : V \to \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only $\{u, v\} \in E$.

A triangle gives a lower bound of 3.

The Moser spindle gives a lower bound of 4. [Moser and Moser, 1961]

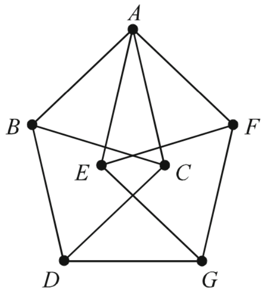
Lower bounds through unit distance graphs

The Hadwiger-Nelson Problem

Find unit distance graphs of large chromatic number.

Definition
A graph $G = (V, E)$ is a unit distance graph if there exists an embedding $f : V \to \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only $\{u, v\} \in E$.

A triangle gives a lower bound of 3.
The Moser spindle gives a lower bound of 4. [Moser and Moser, 1961]



Lower bounds through unit distance graphs

The Hadwiger-Nelson Problem

Find unit distance graphs of large chromatic number.

Definition

A graph $G = (V, E)$ is a unit distance graph if there exists an embedding $f : V \to \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only $\{u, v\} \in E$.

A triangle gives a lower bound of 3.
The Moser spindle gives a lower bound of 4. [Moser and Moser, 1961]

Theorem

There is a unit distance graph on $20\,425$ vertices with chromatic number 5. [De Grey, 2018]

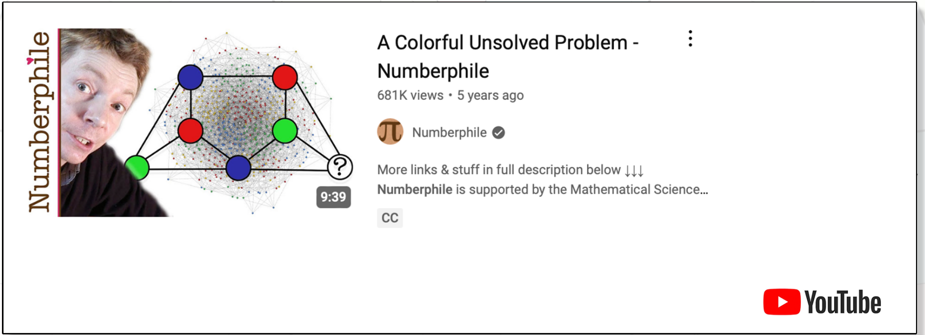
Lower bounds through unit distance graphs

The Hadwiger-Nelson Problem

Find unit distance graphs of large chromatic number.

Definition

A graph $G = (V, E)$ is a unit distance graph if there exists an embedding $f : V \to \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only



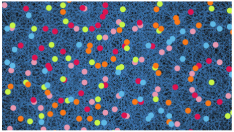
Lower bounds through unit distance graphs

The Hadwiger-Nelson Problem

Find unit distance graphs of large chromatic number.

Definition

A graph $G = (V, E)$ is a **unit distance graph** if there exists an embedding $f : V \to \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only



GRAPH THEORY

Decades-Old Graph Problem Yields to Amateur Mathematician

By EVELYN LAMB | APRIL 17, 2018 | 26 |

...number of vertices? The problem, now known as the Hadwiger-Nelson problem or the problem of finding the chromatic number of the plane, has piqued the interest of many mathematicians, including...

Quantamagazine

Lower bounds through unit distance graphs

The Hadwiger-Nelson Problem

Find unit distance graphs of large chromatic number.

Definition

A graph $G = (V, E)$ is a unit distance graph if there exists an embedding $f : V \to \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only

Aubrey de Grey and Alexander Soifer, *Il Vicino*, January 18, 2020

Ronald L. Graham presents Aubrey D.N.J. de Grey the Prize: $1000, San Diego, September 22, 2018

Lower bounds through unit distance graphs

The Hadwiger-Nelson Problem

Find unit distance graphs of large chromatic number.

Definition

A graph $G = (V, E)$ is a unit distance graph if there exists an embedding $f : V \to \mathbb{E}^2$ of its vertices in the plane s.t. $\|f(u) - f(v)\| = 1$ if and only $\{u, v\} \in E$.

A triangle gives a lower bound of 3.
The Moser spindle gives a lower bound of 4. [Moser and Moser, 1961]

Theorem

There is a unit distance graph on $20\,425$ vertices with chromatic number 5. [De Grey, 2018]

Simpler constructions with...

1. 1581 vertices — for detail see [De Grey, 2018]
2. 627 vertices — [Exoo and Ismailescu, 2020]
3. 553 vertices (as part of Polymath16) — Marijn Heule, for details see [Mixon, 2021]
4. 509 vertices (as part of Polymath16) — Jaan Parts, for details see [Mixon, 2021]

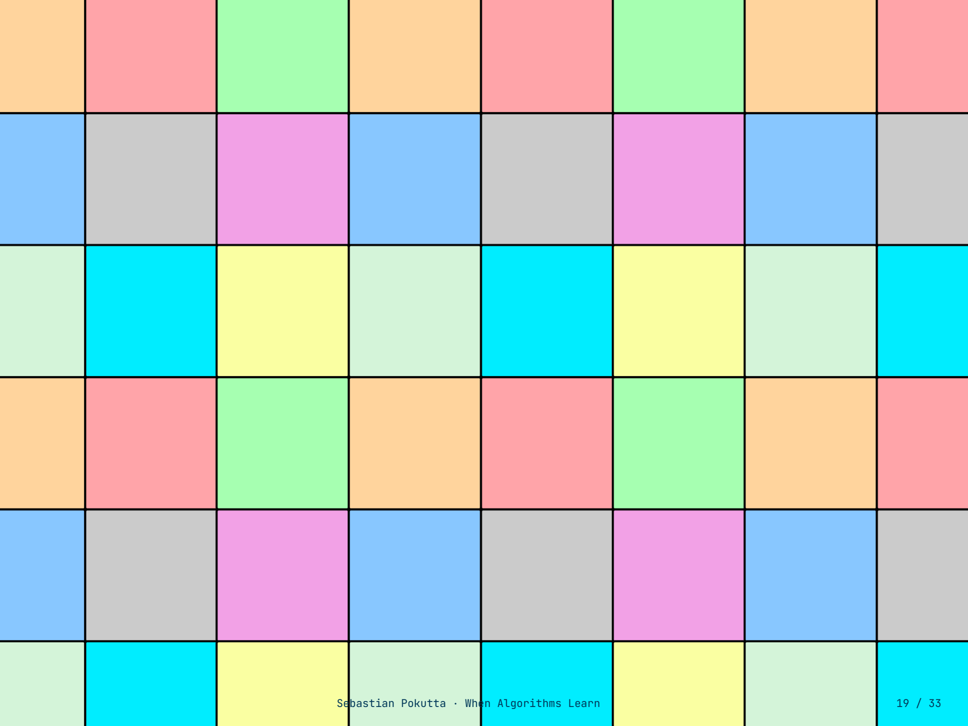
Upper bounds on $\chi(\mathbb{E}^2)$

Upper bounds through colorings

The Hadwiger-Nelson Problem

Explicit colorings $g : \mathbb{E}^2 \to [c] := \{1, \ldots, c\}$, usually derived through tesselations using simple polytopal shapes, give

$$5 \leq \chi(\mathbb{E}^2) \leq \ldots$$

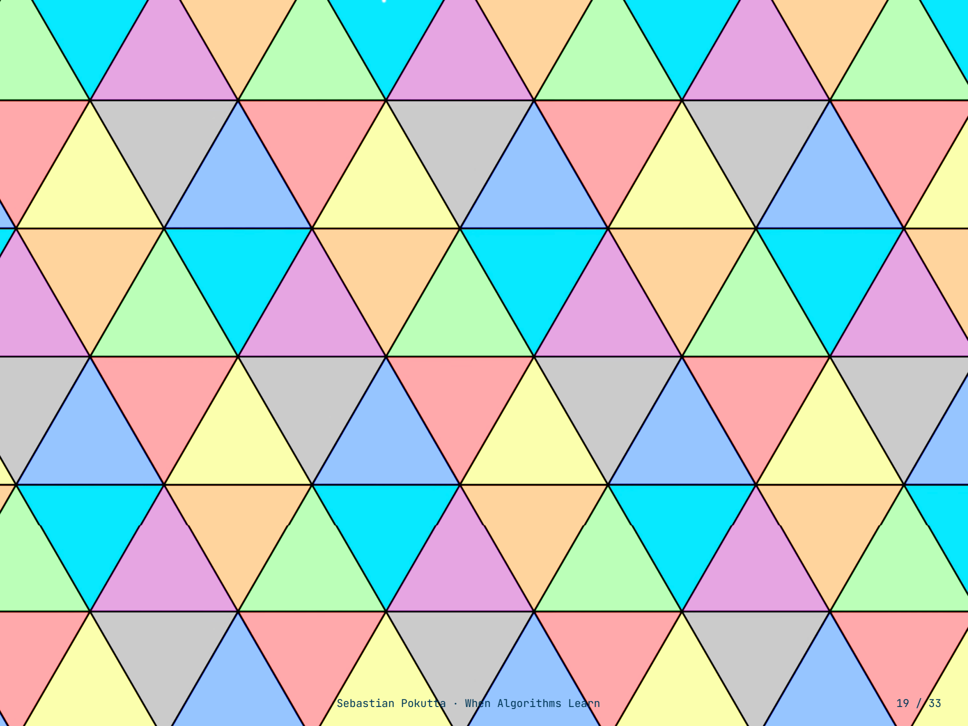


Upper bounds through colorings

The Hadwiger-Nelson Problem

Explicit colorings $g : \mathbb{E}^2 \to [c] := \{1, \ldots, c\}$, usually derived through tesselations using simple polytopal shapes, give

$$5 \leq \chi(\mathbb{E}^2) \leq 9.$$

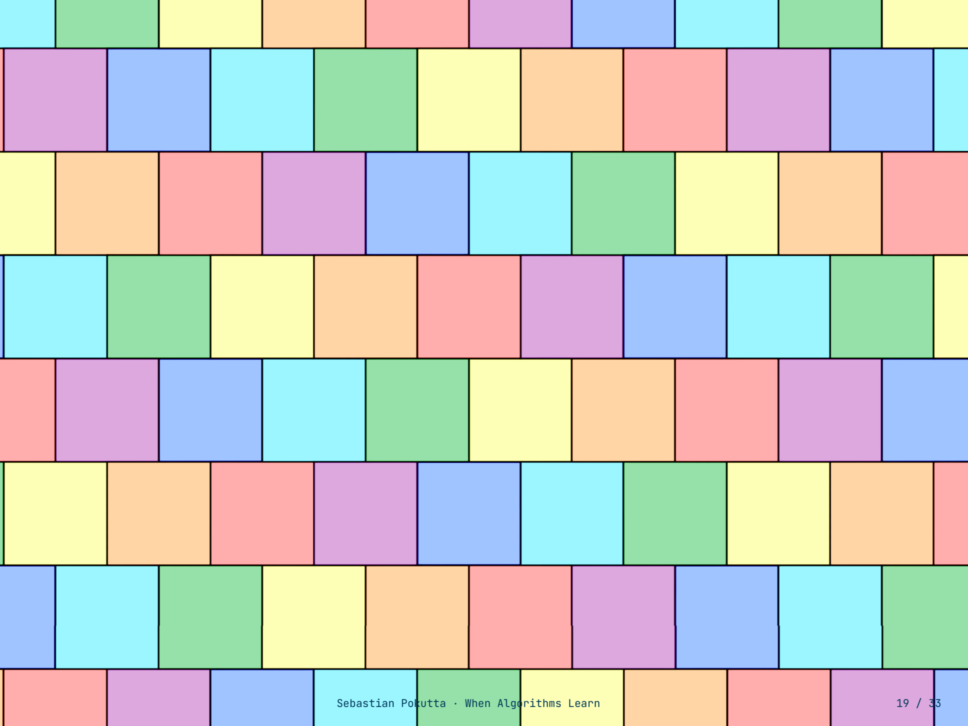


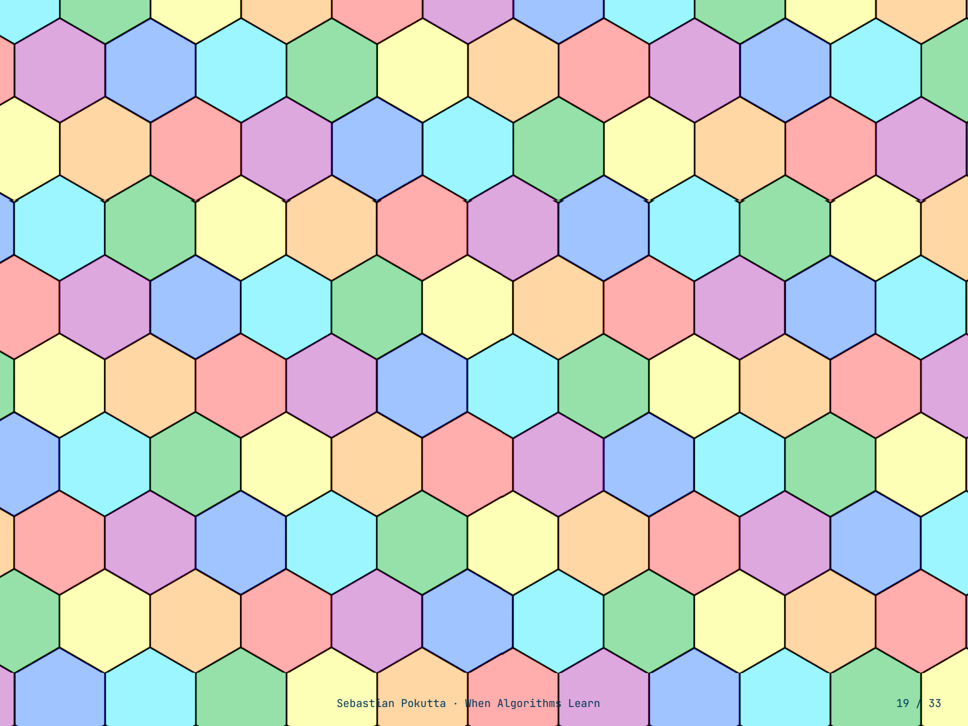
Upper bounds through colorings

The Hadwiger-Nelson Problem

Explicit colorings $g : \mathbb{E}^2 \to [c] := \{1, \ldots, c\}$, usually derived through tesselations using simple polytopal shapes, give

$$5 \leq \chi(\mathbb{E}^2) \leq 8.$$





Explicit colorings $g : \mathbb{E}^2 \to [c] := \{1, \ldots, c\}$, usually derived through tesselations using simple polytopal shapes, give

$$5 \leq \chi(\mathbb{E}^2) \leq 7.$$

Upper bounds through colorings

The Hadwiger-Nelson Problem

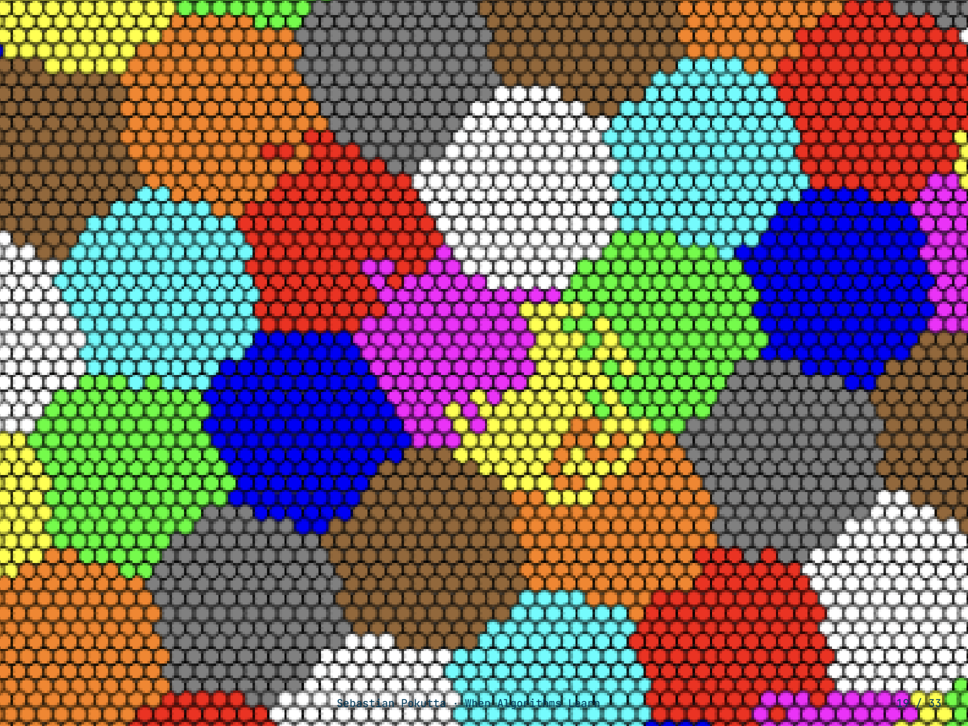
Explicit colorings $g : \mathbb{E}^2 \to [c] := \{1, \ldots, c\}$, usually derived through tesselations using simple polytopal shapes, give

$$5 \leq \chi(\mathbb{E}^2) \leq 7.$$

Question. Can we use computers to find admissible colorings $g : \mathbb{E}^2 \to [c]$, i.e.,

$$\{x \in \mathbb{E}^2 \mid \exists y \in B_1(x) : g(x) = g(y)\} = \emptyset?$$

... attempts, e.g., via discretization and SAT solvers...



Upper bounds through colorings

The Hadwiger-Nelson Problem

Explicit colorings $g : \mathbb{E}^2 \to [c] := \{1, \dots, c\}$, usually derived through tesselations using simple polytopal shapes, give

$$5 \leq \chi(\mathbb{E}^2) \leq 7.$$

Question. Can we use computers to find admissible colorings $g : \mathbb{E}^2 \to [c]$, i.e.,

$$\left\{x \in \mathbb{E}^2 \mid \exists y \in B_1(x) : g(x) = g(y)\right\} = \emptyset?$$

... attempts, e.g., via discretization and SAT solvers...

Idea. Use a parameterized and easily differentiable family $g_\theta : \mathbb{E}^2 \to \Delta_c$ and find

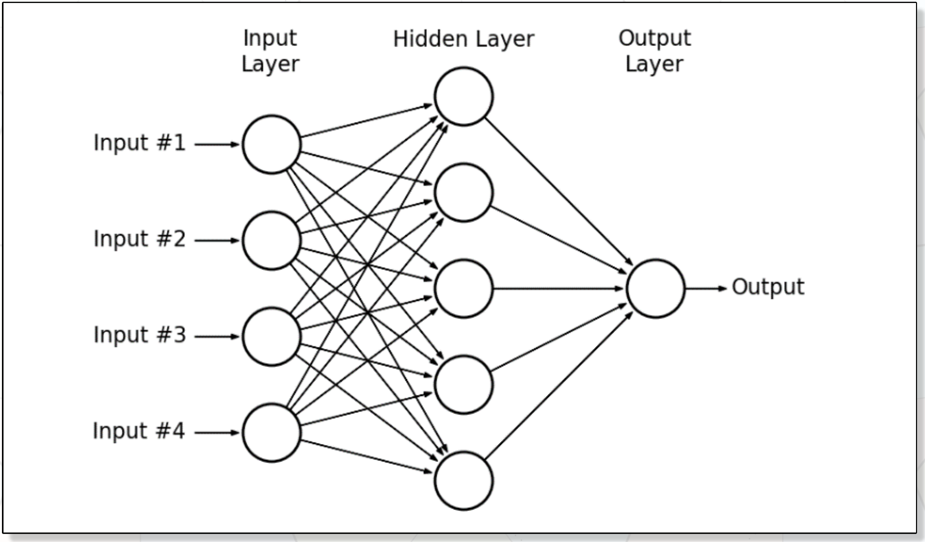
$$\operatorname*{arg\,min}_{\theta} \mathbb{E}\left[\int_{B_1(x)} g_\theta(x) \cdot g_\theta(y)\, dy \,\middle|\, x \in \mathbb{E}^2\right].$$

Key Point. Approach is continuous in nature.

New upper bounds via machine learning?

One second recap: Neural Networks

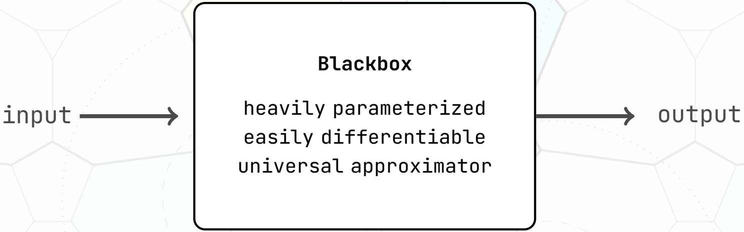
Neural Networks as Colorings



One second recap: Neural Networks

Neural Networks as Colorings

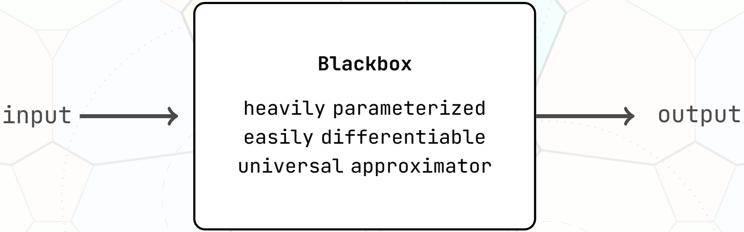
Here. Simply a parameterized continuous function to model the coloring.



One second recap: Neural Networks

Neural Networks as Colorings

Here. Simply a parameterized continuous function to model the coloring.



Theorem (Universal Approximation Theorem)

Feedforward neural networks with certain activation functions are dense (w.r.t. compact convergence) in the space of continuous functions.

Can we improve the upper bound?

Neural Networks as Colorings

Idea. Use gradient descent to train a feedforward network g_θ to minimize

$$L(\theta) = \int_{[-b,b]\times[-b,b]} \int_{B_1(x)} g_\theta(x) \cdot g_\theta(y)\, dy\, dx$$

for some reasonable $b \in \mathbb{R}$?

Can we improve the upper bound?

Neural Networks as Colorings

Idea. Use gradient descent to train a feedforward network g_θ to minimize

$$L(\theta) = \int_{[-b,b]\times[-b,b]} \int_{B_1(x)} g_\theta(x) \cdot g_\theta(y)\, dy\, dx$$

for some reasonable $b \in \mathbb{R}$?

Problem. Still a continuous problem. How to compute?

Can we improve the upper bound?

Neural Networks as Colorings

Idea. Use gradient descent to train a feedforward network g_θ to minimize

$$L(\theta) = \int_{[-b,b]\times[-b,b]} \int_{B_1(x)} g_\theta(x) \cdot g_\theta(y)\, dy\, dx$$

for some reasonable $b \in \mathbb{R}$?

Problem. Still a continuous problem. How to compute?

Stochastic (Batch) Gradient Descent. Sample point $x^{(i)} \in [-b, b] \times [-b, b]$ and $y^{(i)} \in B_1(x)$ for $i = 1, \ldots, m$

Can we improve the upper bound?

Neural Networks as Colorings

Idea. Use gradient descent to train a feedforward network g_θ to minimize

$$L(\theta) = \int_{[-b,b]\times[-b,b]} \int_{B_1(x)} g_\theta(x) \cdot g_\theta(y) \, dy \, dx$$

for some reasonable $b \in \mathbb{R}$?

Problem. Still a continuous problem. How to compute?

Stochastic (Batch) Gradient Descent. Sample point $x^{(i)} \in [-b,b] \times [-b,b]$ and $y^{(i)} \in B_1(x)$ for $i = 1, \ldots, m$ and use that

$$\nabla_\theta L(\theta) \approx \hat{\nabla}_\theta L(\theta) \doteq \frac{1}{m} \sum_{i=1}^{m} \nabla_\theta \, g_\theta(x^{(i)}) \cdot g_\theta(y^{(i)}),$$

Can we improve the upper bound?

Neural Networks as Colorings

Idea. Use gradient descent to train a feedforward network g_θ to minimize

$$L(\theta) = \int_{[-b,b]\times[-b,b]} \int_{B_1(x)} g_\theta(x) \cdot g_\theta(y)\, dy\, dx$$

for some reasonable $b \in \mathbb{R}$?

Problem. Still a continuous problem. How to compute?

Stochastic (Batch) Gradient Descent. Sample point $x^{(i)} \in [-b,b] \times [-b,b]$ and $y^{(i)} \in B_1(x)$ for $i = 1, \ldots, m$ and use that

$$\nabla_\theta L(\theta) \approx \hat{\nabla}_\theta L(\theta) \doteq \frac{1}{m} \sum_{i=1}^{m} \nabla_\theta\, g_\theta(x^{(i)}) \cdot g_\theta(y^{(i)}),$$

where $\nabla_\theta\, g_\theta(x^{(i)}) \cdot g_\theta(y^{(i)})$ is easily computed through backpropagation,

Can we improve the upper bound?

Neural Networks as Colorings

Idea. Use gradient descent to train a feedforward network g_θ to minimize

$$L(\theta) = \int_{[-b,b]\times[-b,b]} \int_{B_1(x)} g_\theta(x) \cdot g_\theta(y)\, dy\, dx$$

for some reasonable $b \in \mathbb{R}$?

Problem. Still a continuous problem. How to compute?

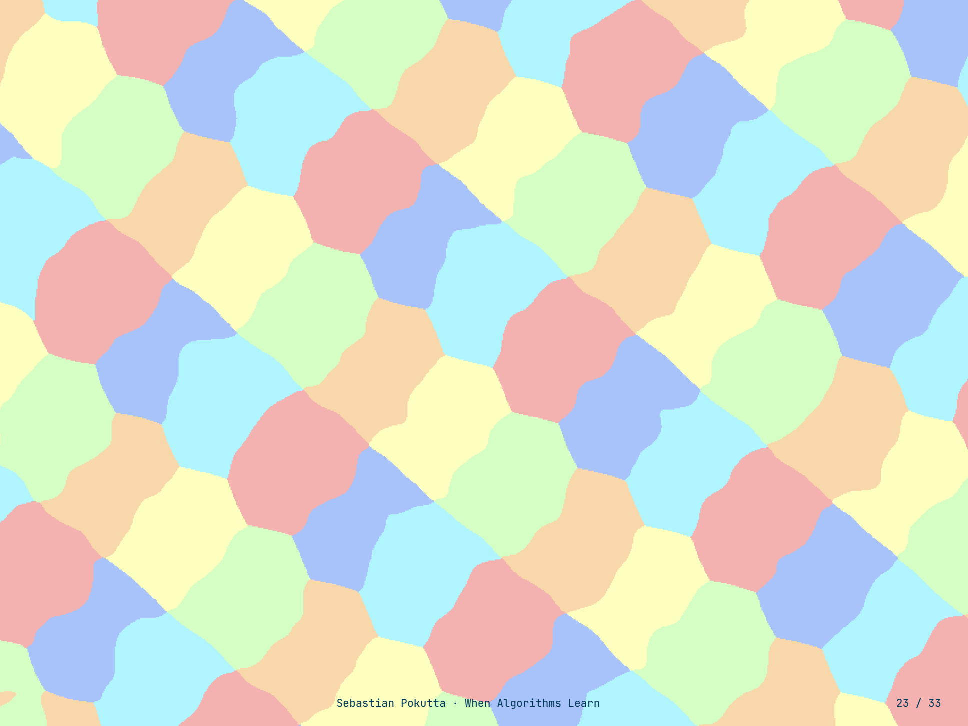
Stochastic (Batch) Gradient Descent. Sample point $x^{(i)} \in [-b,b] \times [-b,b]$ and $y^{(i)} \in B_1(x)$ for $i = 1, \ldots, m$ and use that

$$\nabla_\theta L(\theta) \approx \hat{\nabla}_\theta L(\theta) \doteq \frac{1}{m} \sum_{i=1}^{m} \nabla_\theta\, g_\theta(x^{(i)}) \cdot g_\theta(y^{(i)}),$$

where $\nabla_\theta\, g_\theta(x^{(i)}) \cdot g_\theta(y^{(i)})$ is easily computed through backpropagation, to adjust the parameters θ with an appropriate step size α_k through

$$\theta_{k+1} = \theta_k - \alpha_k \hat{\nabla}_\theta L(\theta).$$

$\Rightarrow$ Very flexible approach "Deep Annealing"
(also: tropicalization of loss function aka softmax... "minimize the max")



Unfortunately this coloring was already known...

Neural Networks as Colorings

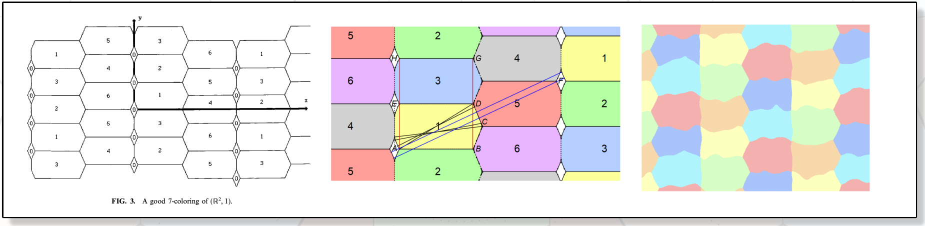
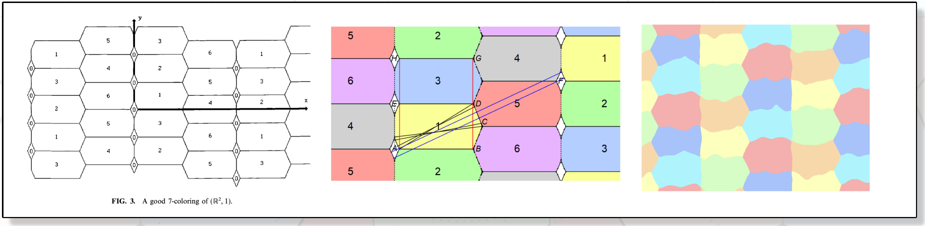


FIG. 3. A good 7-coloring of $(\mathbb{R}^2, 1)$.

Theorem

99.985% of the plane can be colored with 6 colors such that no two points of the same color are a unit distance apart. [Pritikin, 1998, Parts, 2020]

Unfortunately this coloring was already known...



Theorem

99.985% of the plane can be colored with 6 colors such that no two points of the same color are a unit distance apart. [Pritikin, 1998, Parts, 2020]

Corollary

Any unit distance graph with chromatic number 7 must have at least $6\,992$ vertices.

$\Rightarrow$ While coloring was known already maybe on the right track?

Off-diagonal variant

Going off-diagonal

Neural Networks as Colorings

If we cannot solve the original problem, we study variants of it:

Going off-diagonal

Neural Networks as Colorings

If we cannot solve the original problem, we study variants of it:

A c-coloring of the plane has **type** $(d_1, \ldots, d_c)$
if color i does not contain any points at distance d_i.

If we cannot solve the original problem, we study variants of it:

A c-coloring of the plane has **type** $(d_1, \ldots, d_c)$
if color i does not contain any points at distance d_i.

Problem (Soifer in Nash and Rassias' *Open Problems in Mathematics*)

Determine the continuum of six-colorings

$$X_6 = \{d \mid (1,1,1,1,1,d) \text{ can be realized}\}.$$

[Soifer, 1994a, Nash and Rassias, 2016]

Going off-diagonal

Neural Networks as Colorings

If we cannot solve the original problem, we study variants of it:

A c-coloring of the plane has **type** $(d_1, \ldots, d_c)$
if color i does not contain any points at distance d_i.

Problem (Soifer in Nash and Rassias' *Open Problems in Mathematics*)

Determine the continuum of six-colorings

$$X_6 = \{d \mid (1,1,1,1,1,d) \text{ can be realized}\}.$$

[Soifer, 1994a, Nash and Rassias, 2016]

Status. Six-colorings exist for:

1. $d = 1/\sqrt{5}$ [Soifer, 1992]
2. $d = \sqrt{2} - 1$ [Hoffman and Soifer, 1993, 1996]
3. Family with $0.414 \approx \sqrt{2} - 1 \leq d \leq 1/\sqrt{5} \approx 0.447$ [Hoffman and Soifer, 1996, Soifer, 1994b, 2009]

Going off-diagonal

Neural Networks as Colorings

If we cannot solve the original problem, we study variants of it:

A c-coloring of the plane has **type** $(d_1, \ldots, d_c)$
if color i does not contain any points at distance d_i.

Problem (Soifer in Nash and Rassias' *Open Problems in Mathematics*)

Determine the continuum of six-colorings

$$X_6 = \{d \mid (1,1,1,1,1,d) \text{ can be realized}\}.$$

[Soifer, 1994a, Nash and Rassias, 2016]

Status. Six-colorings exist for:

1. $d = 1/\sqrt{5}$ [Soifer, 1992]
2. $d = \sqrt{2} - 1$ [Hoffman and Soifer, 1993, 1996]
3. Family with $0.414 \approx \sqrt{2} - 1 \leq d \leq 1/\sqrt{5} \approx 0.447$ [Hoffman and Soifer, 1996, Soifer, 1994b, 2009]

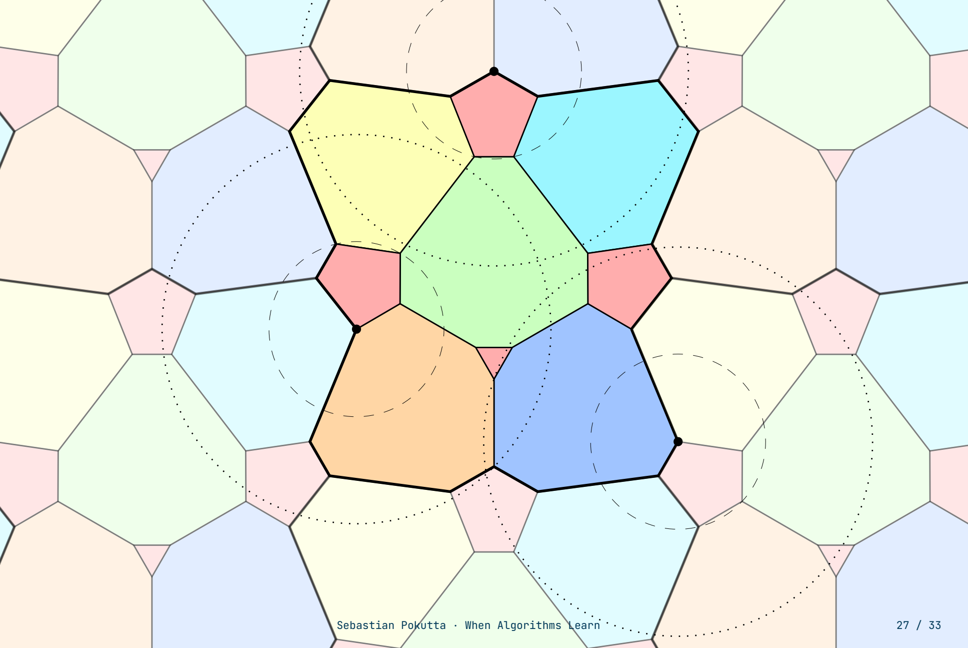
Deep Annealing approach provides two new colorings leading to...

Theorem

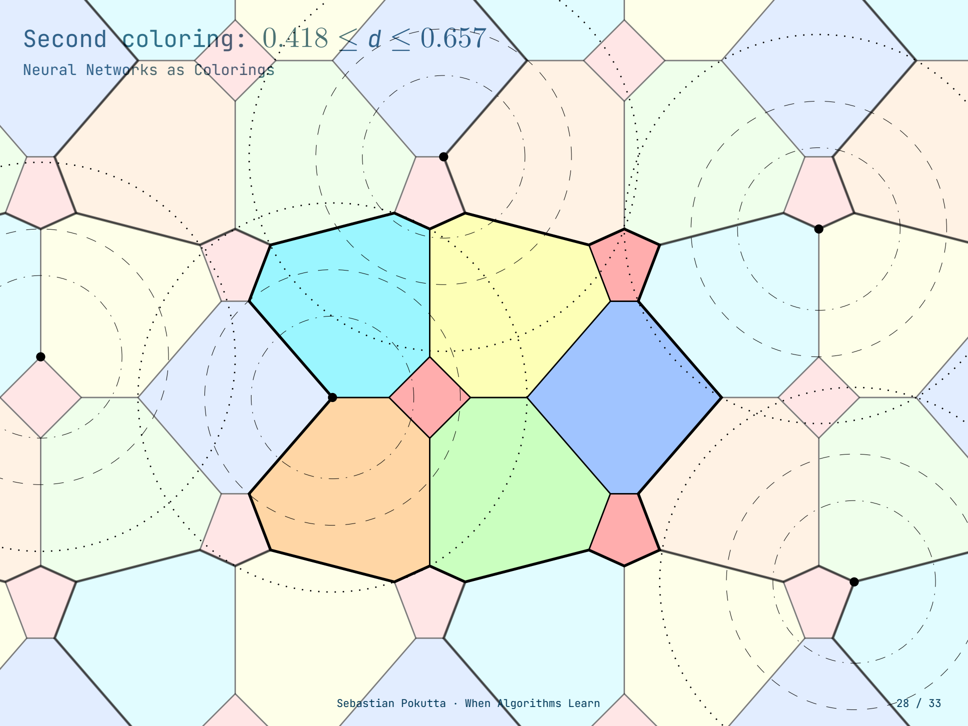
X_6 contains the closed interval [0.354, 0.657]. [Mundinger et al., 2024, 2025]

First coloring: $0.354 \leq d \leq 0.553$

Neural Networks as Colorings

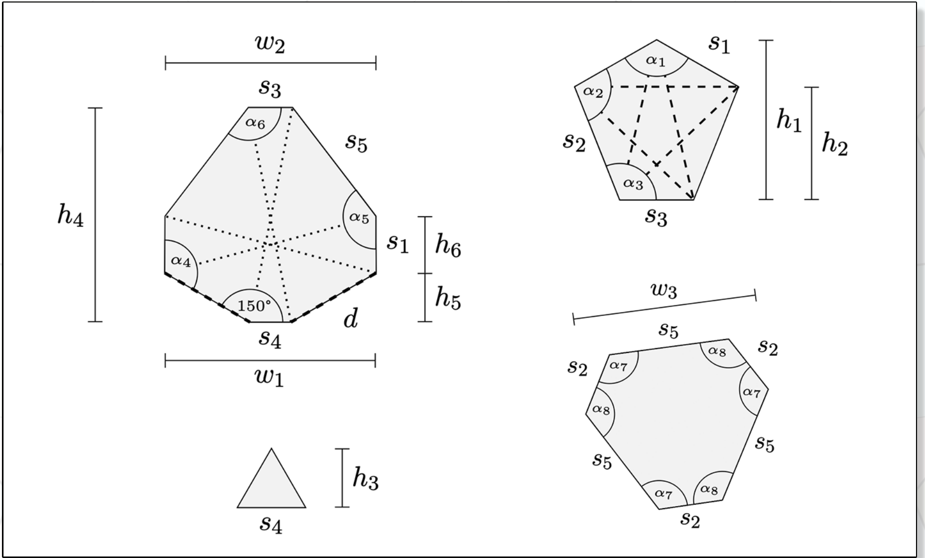


Second coloring: $0.418 \leq d \leq 0.657$



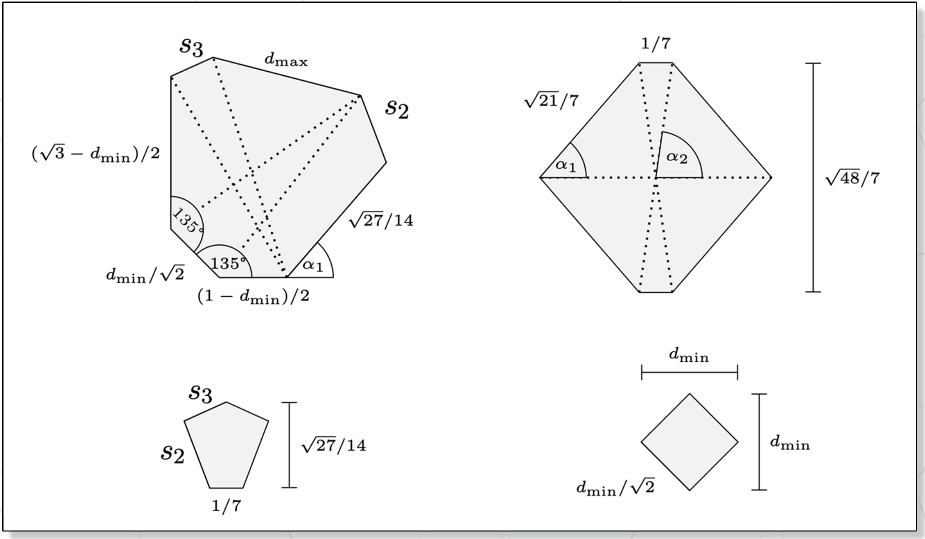
First coloring: exact components

Neural Networks as Colorings



Second coloring: exact components

Neural Networks as Colorings



Final Remarks

1. AI can be used in various ways in modern mathematical research workflows (actual discovery, verification, etc); beyond simple black-box prompting
2. Fully-automatic discovery of new mathematics might be possible in the future but relies on *strong* verification approaches
3. The agentic harness seems to be key: how does the agent receive feedback on its work, how is it guided, and which tools are available?
4. Empirically: the human-in-the-loop is crucial to guide the search

The promises of AI4MATH are great but need to go beyond simple black-box prompting "your favorite Erdős problem" (which then turns out having a solution that is already known)...

Thank you!

Mundinger, K., Pokutta, S., Spiegel, C., and Zimmer, M. (2024). Extending the Continuum of Six-Colorings. Geombinatorics Quarterly. Available at https://arxiv.org/abs/2404.05509.

Mundinger, K., Zimmer, M., Kiem, A., Spiegel, C., and Pokutta, S. (2025). Neural Discovery in Mathematics: Do Machines Dream of Colored Planes? Proceedings of the 42nd International Conference on Machine Learning (ICML), 267, 45236-45255. Available at https://arxiv.org/abs/2501.18527.

Zimmer, M., Pelleriti, N., Roux, C., and Pokutta, S. (2026). The Agentic Researcher: A Practical Guide to AI-Assisted Research in Mathematics and Machine Learning. Preprint. Available at https://arxiv.org/abs/2603.15914.

Pokutta, S. (2026). Frank-Wolfe Beyond 1/t Convergence. Preprint. Available at https://arxiv.org/abs/2604.28006.

References I

K. I. Appel and W. Haken. Every planar map is four-colorable. *Illinois Journal of Mathematics*, 21(3):429–490, 1977. Computer-assisted proof of the four color theorem (first major such proof).

N. d. Bruijn and P. Erdos. A colour problem for infinite graphs and a problem in the theory of relations. *Indigationes Mathematicae*, 13:371–373, 1951.

A. D. De Grey. The chromatic number of the plane is at least 5. *arXiv preprint arXiv:1804.02385*, 2018.

G. Exoo and D. Ismailescu. The chromatic number of the plane is at least 5: a new proof. *Discrete & Computational Geometry*, 64(1):216–226, 2020.

J. Haase and S. Pokutta. Human-AI Cocreativity: Exploring synergies across levels of creative collaboration. In J. C. Kaufman and M. Worwood, editors, *Generative Artificial Intelligence and Creativity*, chapter 16, pages 205–221. 1 2026. doi: 10.1016/B978-0-443-34073-4.00009-5.

T. C. Hales, M. Adams, G. Bauer, D. T. Dang, J. Harrison, H. L. Truong, C. Kaliszyk, V. Magron, S. McLaughlin, N. T. Thang, N. Q. Truong, T. Nipkow, S. Obua, J. Pleso, J. Rute, A. Solovyev, et al. A formal proof of the kepler conjecture. *Forum of Mathematics, Pi*, 5:e2, 2017. doi: 10.1017/fmp.2017.1. Flyspeck project: complete formal verification of the Kepler conjecture.

M. J. H. Heule, O. Kullmann, and V. W. Marek. Solving and verifying the boolean pythagorean triples problem via cube-and-conquer. In *Theory and Applications of Satisfiability Testing – SAT 2016*, volume 9710 of *Lecture Notes in Computer Science*, pages 228–245. Springer, 2016. Computer-intensive proof generating a multi-terabyte certificate.

I. Hoffman and A. Soifer. Almost chromatic number of the plane. *Geombinatorics*, 3(2):38–40, 1993.

I. Hoffman and A. Soifer. Another six-coloring of the plane. *Discrete Mathematics*, 150(1-3):427–429, 1996.

D. Mixon. Polymath16, seventeenth thread: Declaring victory. *Polymath16*, February 1 2021. Retrieved 16 August 2021.

L. Moser and M. Moser. Solution to problem 10. *Canadian Mathematical Bulletin*, 4:187–189, 1961.

K. Mundinger, S. Pokutta, C. Spiegel, and M. Zimmer. Extending the Continuum of Six-Colorings. *Geombinatorics Quarterly*, 5 2024.

K. Mundinger, M. Zimmer, A. Kiem, C. Spiegel, and S. Pokutta. Neural Discovery in Mathematics: Do Machines Dream of Colored Planes? *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, 267:45236–45255, 5 2025. ISSN 2640-3498. doi: 10.48550/arXiv.2501.18527.

J. F. Nash and M. T. Rassias. *Open problems in mathematics*. Springer, 2016.

J. Parts. What percent of the plane can be properly 5-and 6-colored? *arXiv preprint arXiv:2010.12668*, 2020.

S. Pokutta. Frank-Wolfe Beyond 1/t Convergence. *preprint*, 4 2026. doi: 10.48550/arXiv.2604.28006.

D. Pritikin. All unit-distance graphs of order 6197 are 6-colorable. *Journal of Combinatorial Theory, Series B*, 73(2):159–163, 1998.

N. Robertson, D. P. Sanders, P. Seymour, and R. Thomas. The four colour theorem. *Journal of Combinatorial Theory, Series B*, 70(1):2–44, 1997. Shorter and fully computer-verified proof reducing the number of configurations.

A. Soifer. A six-coloring of the plane. *Journal of Combinatorial Theory, Series A*, 61(2):292–294, 1992.

A. Soifer. Six-realizable set x_6. *Geombinatorics*, III(4):140–145, 1994a.

A. Soifer. An infinite class of six-colorings of the plane. *Congressus Numerantium*, pages 83–86, 1994b.

A. Soifer. *The mathematical coloring book: Mathematics of coloring and the colorful life of its creators*. Springer, 2009.

A. Soifer. *The New Mathematical Coloring Book*. Springer, 2024.

M. Zimmer, N. Pelleriti, C. Roux, and S. Pokutta. The Agentic Researcher: A Practical Guide to AI-Assisted Research in Mathematics and Machine Learning. *preprint*, 3 2026. doi: 10.48550/arXiv.2603.15914.