The Archimedean trap: Why traditional reinforcement learning will probably not yield AGI

Samuel Allen Alexander   SAMUELALLENALEXANDER@GMAIL.COM

Quantitative Research Analyst
The U.S. Securities and Exchange Commission
New York Regional Office

Editor: TBD

Abstract

After generalizing the Archimedean property of real numbers in such a way as to make it adaptable to non-numeric structures, we demonstrate that the real numbers cannot be used to accurately measure non-Archimedean structures. We argue that, since an agent with Artificial General Intelligence (AGI) should have no problem engaging in tasks that inherently involve non-Archimedean rewards, and since traditional reinforcement learning rewards are real numbers, therefore traditional reinforcement learning cannot lead to AGI. We indicate two possible ways traditional reinforcement learning could be altered to remove this roadblock.

1. Introduction

Whenever we measure anything using a particular number system, the corresponding measurements will be constrained by the structure of that number system. If the number system has a different structure than the things we are measuring with it, then our measurements will suffer accordingly, just as if we were trying to force square pegs into round holes.

For example, the natural numbers make lousy candidates for measuring lengths in a physics laboratory. Lengths in the lab have properties such as, for example, the fact that for any two distinct lengths, there is an intermediate length strictly between them. The natural numbers lack this property. Imagine the poor physicist, brought up in a world of only natural numbers, scratching his or her head upon encountering a rod with length strictly between two rods of length 1 and 2.

It is tempting to think of the real numbers $\mathbb{R}$—i.e., the unique complete ordered field—as a generic number system with whatever structure suits our needs. But the real numbers do have their own specific structure. That structure is flexible enough to accomodate many needs, but we shouldn't just take that for granted. One particular property constraining the real numbers is the following.

Lemma 1 *(The Archimedean Property[1]) Let $r > 0$ be any positive real number. For every real number y, there is some natural number n such that $nr > y$.*

---

1. The Archimedean property is named after Archimedes of Syracuse. A similar property appears as the fifth axiom in his *On the Sphere and Cylinder* (Archimedes, 1897):

Rather than directly prove Lemma 1, we will prove a generalized result which, we will argue, is more adaptable to other structures.

Lemma 2 *(The Generalized Archimedean Property) Let $r > 0$ be any positive real number. For any $x, y \in \mathbb{R}$, say that x is* significantly less *than y if $x \leq y - r$. If $x_0, x_1, x_2, \ldots$ is any infinite sequence of real numbers, where each x_i is significantly less than x_{i+1}, then for every real number y, there exists some i such that y is significantly less than x_i.*

Proof If not, then there is some y such that $y + r > x_i$ for all i. Thus, $X = \{x_0, x_1, x_2, \ldots\}$ has an upper bound. By the completeness of $\mathbb{R}$, X must have a least upper bound $z \in \mathbb{R}$. Since z is the least upper bound for X, $z - r$ is not an upper bound for X, so there is some i such that $x_i > z - r$. By assumption, $x_i \leq x_{i+1} - r$, so $x_{i+1} > z$, contradicting the choice of z. ∎

Lemma 1 follows from Lemma 2 by letting $x_i = ir$.

The above property is automatically inherited by subsystems of the reals, such as the rational numbers $\mathbb{Q}$, the natural numbers $\mathbb{N}$, the integers $\mathbb{Z}$, or the algebraic numbers. All inherit the Generalized Archimedean Property in obvious ways.

Lemma 2 allows us to adapt the notion of Archimedeanness to other things than real numbers, even to things for which there is no notion of arithmetic at all (Lemma 1 would not adapt to such things). All we need is a notion of "significantly less than". For any set of things, some of which are "significantly less than" others, we can ask whether or not the property in Lemma 2 holds. We will make this formal in Section 2.

Example 1 *(Fuzzy widgets) Suppose we have some fuzzy widgets, and we observe that certain widgets are fuzzier than others. Naturally, we are inclined to quantify the fuzziness of the widgets, assigning them numerical fuzziness measures from some number system. Nine times out of ten, we choose to use the real numbers, or a subsystem thereof, often without a second thought. But suppose among these widgets, there happen to be widgets $w_1, w_2, \ldots$ such that each w_i is significantly less fuzzy than w_{i+1}, and another widget w_∞ such that all the w_i's are significantly less fuzzy than w_∞. Suddenly, our decision to use real numbers puts us in a bind. It is impossible to assign real number fuzziness measures to*

> Further, of unequal lines, unequal surfaces, and unequal solids, the greater exceeds the less by
> such a magnitude as, when added to itself, can be made to exceed any assigned magnitude
> among those which are comparable with [it and with] one another.

Note that Archimedes specifically restricts his statement to lengths, surface areas and volumes, in fact going out of his way to limit the magnitudes to which said length/area/volume can be made to exceed (he could have saved some words by stopping his sentence at "...can be made to exceed any assigned magnitude", if that were his intention).

The Archimedean property is also closely related to Definition 4 of Book V of Euclid's *Elements* (Euclid, 2007):

> (Those) magnitudes are said to have a ratio with respect to one another which, being multiplied,
> are capable of exceeding one another.

Proposition 1 of Book X is also relevant. Many math historians speak of the modern-day Archimedean property, Archimedes' 5th axiom, and Euclid's properties as being identical, but in fact they are all subtly different from one another, see Bair et al. (2013).

our widgets in such a way that significantly less fuzzy widgets get significantly smaller real number measures. That would contradict Lemma 2.

Note that the above example does not require us to have any notion of multiplying fuzziness by a natural number n (as we would need to have if we wanted to adapt Lemma 1). This illustrates the enhanced adaptability of Lemma 2.

The structure of this paper is as follows.

- In Section 2 we formally adapt Lemma 2 to obtain a notion of Archimedeanness for non-numerical structures, and demonstrate that non-Archimedean such structures cannot accurately be measured using the real numbers.

- In Section 3 we argue that traditional reinforcement learning will not lead to AGI because its rewards are overly constrained.

- In Section 4 we discuss non-traditional variations of reinforcement learning that avoid the problem of overly constrained rewards.

- In Section 5 we summarize and make concluding remarks.

2. Generalized Archimedean Structures

The real numbers possess the Archimedean property, but other structures may or may not. To make this more precise, we introduce the following formalism, adapting from Lemma 2.

Definition 3 *A* significantly-ordered structure *is a collection X with an ordering $\ll$. For $x_1, x_2 \in X$, we say x_1 is* significantly less *than x_2 if $x_1 \ll x_2$. A significantly-ordered structure is* Archimedean *if it has the following property: for every X-sequence $x_0 \ll x_1 \ll x_2 \ll \cdots$, for every $x_\infty \in X$, there is some i such that $x_\infty \ll x_i$.*

For any real number $r > 0$, a prototypical example of an Archimedean significantly-ordered structure is the real numbers with $\ll$ defined such that $x \ll y$ if and only if $x \leq y - r$.

Definition 4 *Suppose $(X, \ll)$ is a significantly-ordered structure. A function $f : X \to \mathbb{R}$ is said to* accurately measure *$(X, \ll)$ if there is some real $r > 0$ such that the following requirement holds:*

- *For all $x, y \in X$, $x \ll y$ if and only if $f(x) \leq f(y) - r$.*

The following proposition formalizes the dilemma we illustrated in Example 1 (think of X as a set of things we want to measure).

Proposition 5 *(Inadequacy of the reals for non-Archimedean structures) Suppose $(X, \ll)$ is a significantly-ordered structure. If X is non-Archimedean, then no function $f : X \to \mathbb{R}$ accurately measures $(X, \ll)$.*

3

Proof Assume, for sake of a contradiction, that some $f : X \to \mathbb{R}$ exists which accurately measures $(X, \ll)$. Thus there is some real $r > 0$ such that for all $x, y \in X$, $x \ll y$ if and only if $f(x) \leq f(y) - r$. Since X is non-Archimedean, there is some X-sequence $x_0 \ll x_1 \ll x_2 \ll \cdots$ and some $x_\infty \in X$ such that there is no i such that $x_\infty \ll x_i$. By choice of r, each $f(x_i) \leq f(x_{i+1}) - r$ and there is no i such that $f(x_\infty) \leq f(x_i) - r$. This contradicts Lemma 2. ∎

Proposition 5 tells us that we cannot accurately measure non-Archimedean structures using real numbers[2]. Any attempt to do so will necessarily be misleading, because ordering relationships among the non-Archimedean structures will fail to be reflected by the real-number measurements given to them. We will inevitably end up like the puzzled physicist brought up in a world of only natural numbers, confronted by a rod of length 1.5.

Remark 6 *(Spearman's Law of Diminishing Returns) Suppose $(X, \ll)$ is a non-Archimedean significantly-ordered structure with elements $x_0, x_1, \ldots$ and x_∞ such that $x_0 \ll x_1 \ll \cdots$ and each $x_i \ll x_\infty$. Suppose $f : X \to \mathbb{R}$ has the property that $f(x_0) < f(x_1) < \cdots$ and each $f(x_i) < f(x_\infty)$. Then the monotone convergence theorem implies that $\lim_{i \to \infty} f(x_i)$ converges, which in turn implies that $\lim_{i \to \infty}(f(x_{i+1}) - f(x_i)) = 0$. This suggests a general law of diminishing returns: any time a non-Archimedean significantly-ordered structure $(X, \ll)$ is measured using real numbers, if the measurement does not blatantly violate $\ll$ (in other words, if there are no $x \ll y$ such that x is given a larger real-number measurement than y), then there will inevitably be elements $x_0 \ll x_1 \ll \cdots$ exhibiting* diminishing returns, *in the sense that the measurements of x_i and x_j are approximately equal for large i, j. If human intelligence is non-Archimedean, this could potentially shed light on a psychometrical phenomenon called* Spearman's Law of Diminishing Returns *(Spearman, 1927; Blum and Holling, 2017; Hernández-Orallo, 2019), the empirical tendency of cognitive ability tests to be less correlated in high-intelligence populations. Even tiny measurement errors would eventually dominate the test result differences as the true results plateau.*

Example 2 *(Examples of non-Archimedean structures)*

- *(Sets) Say that set U is significantly smaller than set V if there is an injective function from U into V but there is no bijective function from U onto V. It is easy to show there are sets $U_0, U_1, \ldots$, with each U_i significantly smaller than U_{i+1}, and each U_i is significantly smaller than $U_\infty = \cup_{i=0}^\infty U_i$. Thus, sets are non-Archimedean. In the field of* set theory, *mathematicians measure the size of sets using Georg Cantor's famous non-Archimedean number system, the cardinal numbers.*

- *(Logical theories) It is not difficult to come up with (for example) true theories $T_0, T_1, \ldots$ (in the language of arithmetic) such that each T_{i+1} proves the consistency of T_i, and an additional true theory T_∞ (in the language of arithmetic) which proves the consistency of $\cup_{i=0}^\infty T_i$. In a sense, then, each T_i is significantly weaker than T_{i+1}*

---

2. There is an area of research known as *measurement theory*, which, traditionally, "takes the real numbers as a pre-given numerical domain" (Niederée, 1992). Some work has been done to generalize measurement theory away from this assumption (Narens, 1974; Skala, 1975; Rizza, 2016). We would submit this paper as further motivation in that direction.

(see Gödel's incompleteness theorems), and each T_i is significantly weaker than T_∞. In this sense, logical theories are non-Archimedean. In the field of proof theory *(Pohlers, 2008; Rathjen, 2006), logicians measure the logical strength of theories using computable ordinal numbers, another non-Archimedean number system.*

- *(Asymptotic runtime complexities) Suppose $A_0, A_1, \ldots$ are algorithms such that each A_i has runtime complexity $\Theta(n^i)$, and suppose A_∞ is an algorithm with runtime complexity $\Theta(2^n)$. Then in a certain sense, each A_i has significantly lower asymptotic runtime complexity than A_{i+1}, and each A_i has significantly lower asymptotic runtime complexity than A_∞. In this sense, asymptotic runtime complexity is non-Archimedean. In computer science, these runtime complexities are usually measured using big-O, big-Θ, or similar notation systems.*

Example 3 *(Speculative examples of potentially non-Archimedean structures) Certain structures might plausibly be non-Archimedean, but it is a difficult question to say whether they truly are or not. The reader could come up with such examples in great abundance.*

- *(Musical beauty) Assuming there is such a thing as objective musical beauty (not contingent on features of the human condition, etc.), then it is plausible that music might be non-Archimedean, in the following sense: there might be songs $S_0, S_1, \ldots$ such that each S_i is significantly less beautiful than S_{i+1}, and another song S_∞ such that each S_i is significantly less beautiful than S_∞.*

- *(Ethical utility) Early utilitarian Jeremy Bentham suggested a hedonistic calculus in which pleasure measurements would be assigned to actions, to help adjudicate ethical dilemmas. His successor, John Stuart Mill, objected that some actions are incomparably better than others: "If one of [two pleasures] is, by those who are competently acquainted with both, placed so far above the other that they prefer it ... and would not resign it for any quantity of the other pleasure which their nature is capable of, we are justified in ascribing to the preferred enjoyment a superiority in quality, so far outweighing quantity as to render it, in comparison, of small account" (Mill, 2016). This suggests that Bentham's pleasures are non-Archimedean.*

- *(AGI) It is plausible that there are[3] AGIs $A_0, A_1, \ldots$ such that each A_i is significantly less intelligent than A_{i+1}, and another AGI A_∞ such that each A_i is significantly less intelligent than A_∞. We first pointed this out in (Alexander, 2019b), where we propose measuring the intelligence of mechanical knowing agents using computable ordinals, the same non-Archimedean number system which proof theorists use to measure logical strength of mathematical theories. Incidentally, if AGI intelligence is non-Archimedean, then Proposition 5 shows it is impossible to measure machine intelligence using real numbers without some of those measurements being misleading[4].*

---

3. As hinted by Protagoras, assuming Protagoras's own intelligence stays constant and remains higher than the intelligence of his student and that they live forever and that *better* means *significantly better*: "The very day you start, you will go home a better man, and the same thing will happen the day after. Every day, day after day, you will get better and better" (Plato, 1997).

4. This would solve an open problem implicitly stated by Legg and Hutter (2007) when they said of their real-number universal intelligence measure: "...none of these people have been able to communicate why

- *(Nonstandard cosmologies) Some authors (Al-Dhalimy and Geyer, 2016; Andréka et al., 2012; Reeder, 2012; Rosinger, 2007; Chen, forthcoming) have even speculated about the nature of non-Archimedean space and/or time.*

3. Reinforcement learning

In reinforcement learning (RL), an agent interacts with an environment, taking actions from a fixed set of possible actions. With every action the agent takes, the environment responds with a new observation and with a reward. In traditional RL, these rewards are real numbers (many authors further constrain them to be rational numbers).

By restricting rewards to be real (or rational) numbers, we unconsciously constrain RL to only be applicable toward tasks of an inherently Archimedean nature. For example, Wirth et al. (2017) point out that in tasks related to cancer treatment (Zhao, Kosorok, and Zeng, 2009), "the death of a patient should be avoided at any cost. However, an infinitely negative reward breaks classic reinforcement learning algorithms and arbitrary, finite values have to be selected." This problem could be avoided if instead of real numbers, rewards were drawn from a suitable non-Archimedean number system containing negative infinities. Doing so would be a departure from traditional RL.

Example 4 *To give an intuitive example, assume that musical beauty is non-Archimedean, as in Example 3. We can imagine environments where the RL agent is tasked with composing songs. For example, the possible actions the agent is allowed to take might include one action for each piano key, plus an additional "stand and bow" action to signal that a song is finished. Whenever the agent stands and bows, the agent is rewarded with applause based on the beauty of the song the agent composed[5]. Assuming musical beauty is non-Archimedean, such an environment falls outside the possibility of traditional RL. By Lemma 5, there is no way to assign real number rewards to songs without misleading the agent. If $S_0, S_1, S_2, \ldots$ are songs where each S_i is significantly less beautiful than S_{i+1}, and all the S_i are significantly less beautiful than another song S_∞, then there is no way to assign real-valued rewards to these songs such that each S_i gets significantly less reward than S_{i+1} and significantly less reward than S_∞.*

Or, to re-use the cancer example, assume there are certain bad procedures the robotic surgeon could take, each one significantly worse than the previous, but all still significantly better than killing the patient. There is no way to assign real-valued rewards to these actions, and to killing the patient, in such a way that each bad action gets punished significantly harsher than the previous, but still significantly more forgiving than the punishment for killing the patient.

The reader might object by challenging the non-Archimedeanness of music and of medical procedures. But we only used those to make the examples more intuitive. If the reader insists, we can resort to mathematical tasks.

---

the work [on measuring universal intelligence using real numbers] is so obviously flawed in any concrete way ... If anyone would like to properly explain their position to us in the future, we promise not to chase you down the street!"

5. To quote Wang and Hammer (2015): "Decision makings often do not happen at the level of basic operations, but at the level of composed actions, where there are usually infinite possibilities."

Example 5 *Imagine that the agent is tasked with typing up mathematical theories, and when the agent stands and bows, the agent is rewarded with applause based on the proof-theoretical strength of the theory (or hit with tomatoes if the theory is inconsistent). In Example 2 we noted that proof-theoretical strength of theories is non-Archimedean. There exist theories $T_0, T_1, \ldots$, each significantly proof-theoretically weaker than the next, and another theory T_∞, significantly proof-theoretically stronger than all the T_i's. We cannot possibly assign real-valued rewards to these theories without misleading the agent.*

The reader might object to Example 5 on the grounds that judging the proof-theoretical strength of a theory is inherently non-computable anyway. The example could be modified so that instead of typing up mathematical theories, the agent has to type up mathematical subtheories in (say) the language of Peano arithmetic, accompanied by consistency proofs in (say) ZFC. It can be shown that the proof-theoretical strength of mathematical theories is still non-Archimedean, even when restricted to subtheories of arithmetic whose consistency can be proven in ZFC[6].

The reader might object that the above theories-with-proofs example is contrived. But an AGI with human or better intelligence should have no problem at least comprehending and attempting such a task (regardless of whether or not the AGI is able to perform well at it). When we prove that the Halting Problem is unsolvable, we do so by considering contrived programs that we could write if the Halting Problem were solvable. The contrivedness of those programs does not invalidate the proof of the unsolvability of the Halting Problem. Again, when we prove that C++ templates are Turing complete (Veldhuizen, 2003), we do so by considering extremely bizarre C++ templates that would never arise naturally in a software studio. This does not invalidate the proof that C++ templates are Turing complete.

Finally, the reader might object that approximating infinite rewards with arbitrary large finite rewards is good enough. Who cares (the argument might go) whether pushing a button gives the agent infinite pleasure or only a million units of pleasure? Either way (the argument goes) the agent is going to learn to prefer that button over a button that gives only .1 units of pleasure. The following example shows that this logic breaks down in non-Markov environments.

Example 6 *(Delayed gratification) Consider an environment with a red button and a blue button. Pushing the red button always grants $+1$ reward. As for the blue button, suppose the agent presses the blue button for the ith time. If $i = 2^j$ for some integer j, then the agent shall receive a reward of ω (the smallest infinite ordinal), but otherwise, the agent shall receive 0 reward. If we approximate ω with a real-value of, say, $1,000,000$, then after a long enough time spent in the environment, an AGI will be misled into thinking that it isn't worth the longer and longer wait-times between blue-button rewards: eventually, it will take more than $1,000,000$ blue-button-presses to get rewarded, and the AGI will consider it more worthwhile to get the guaranteed $+1$ reward from the red button.*

---

6. For example, let T_0 be the theory of Peano arithmetic, and for each i, let T_{i+1} be T_i together with $CON(T_i)$, a canonical axiom encoding the consistency of T_i. Let T_∞ be the theory of Peano arithmetic along with $CON(\cup_i T_i)$. ZFC is certainly adequate to prove the consistency of each of these theories. In the sense of Example 2, each T_i is significantly weaker than T_{i+1} and significantly weaker than T_∞.

Our critic could respond to Example 6 by making the approximation dynamic, say, making the 2^jth press of the blue button grant $1000000 \cdot 2^j$ reward, but at this point, the critic is clearly just hard-coding the correct actions into the reward function, something which is only possible in Example 6 because the environment is simple enough that we can completely understand it ourselves. For the kinds of non-trivial environments where AGI would actually be useful, such carefully engineered reward approximations would quickly become intractible.

Reinforcement Learning is useful for many practical tasks, but at least in its traditional flavor, it is too constrained (by its arbitrary choice of number system for its rewards) to apply to certain non-Archimedean tasks[7], which, however contrived they are, could certainly be attempted by an AGI. Traditional reinforcement learning will not lead to AGI.

4. Non-traditional reinforcement learning

We have argued that traditional RL cannot lead to AGI, because an AGI is capable of attempting non-Archimedean tasks whose rewards are too rich to express using real numbers. There are at least two potential ways to change RL so as to make it applicable to such tasks and, thus, at least potentially capable of leading to AGI. Of course, there is no guarantee that removing the roadblock in this paper will cause RL to lead to AGI. There might be other roadblocks besides the inadequate reward number system[8].

4.1 Preference-based reinforcement learning

A lot of exciting research has been done on non-traditional variations of RL where, instead of giving the agent numerical rewards for taking actions, one instead informs the agent about the relative preference of various actions or action-sequences. See (Wirth et al., 2017) for a survey. This very nicely side-steps the problems from this paper.

4.2 Reinforcement learning with other number systems

The most obvious way to modify RL to avoid the problems presented in this paper is to change which number system is used[9]. As far as this author is aware, the choice to use real (or rational) numbers for rewards was not made based on any fundamental criteria[10]. The real (or rational) numbers are currently a useful pragmatic choice because they are easy to compute with using 21st century software and 21st century school curricula, but that's hardly relevant in the field of genuine AGI. One might say the real numbers were a good choice because they are familiar, but even that is arguable: in general, students are usually

---

7. Perhaps explaining why "despite almost two decades of RL research, there has been little solid evidence of RL systems that may one day lead to [AGI]" (Livingston, Garvey, and Elhanany, 2008).
8. For example, many RL authors consider non-deterministic environments where rewards and observations include an element of randomness. The probabilities involved are, traditionally, assumed to be real numbers. Perhaps some recent work (Benci, Horsten, and Wenmackers, 2013) on non-Archimedean probability could be relevant against that roadblock.
9. Anticipated by Rizza (2016).
10. Niederée (1992) points out that there are no deeper reasons to assume that the number system should necessarily even have the same cardinality as $\mathbb{R}$. And Rizza (2016) says: "No particular feature of the space of informational states suggests that such a codomain [as $\mathbb{R}$] should be selected".

not taught what the real numbers *actually are*, unless they major in pure mathematics at the university level. Anyway, the familiarity argument is totally irrelevant in the field of AGI.

Various non-Archimedean number systems exist. Number systems can be discrete or continuous; the nature of reinforcement learning clearly suggests a continuous number system. We will consider three possible number systems: formal Laurent series; hyperreal numbers; and surreal numbers.

4.2.1 FORMAL LAURENT SERIES

David Tall (1980) described the following real-number-extending number system (which he called the "superreals", but that vocabulary does not seem to have caught on).

Definition 7 *A* formal Laurent series *is a formal expression of the following form (where m can be any integer and $a_{-m}, a_{-(m-1)}, \ldots, a_0, a_1, a_2, \ldots$ are real numbers, $a_{-m} \neq 0$):*

$$\sum_{i=-m}^{\infty} a_i \epsilon^i.$$

Suppose $A = \sum_{i=-m}^{\infty} a_i \epsilon^i$ and $B = \sum_{i=-n}^{\infty} b_i \epsilon^i$ are two distinct formal Laurent series. We declare $A < B$ if and only if $a_i < b_i$ where i is the smallest index such that $a_i \neq b_i$ (where we consider a_i to be 0 for all $i < -m$ and we consider b_i to be 0 for all $i < -n$).

We can consider the real numbers $\mathbb{R}$ to be embedded in the formal Laurent series by way of the embedding $r \mapsto r\epsilon^0$. Having done so, the intuition is that, for example, $1\epsilon^1$ is what we might call a "first-order infinitesimal number", smaller than every positive real; $1\epsilon^2$ is what we might call a "second-order infinitesimal number", smaller than every positive first-order infinitesimal number; and so on. Likewise, $1\epsilon^{-1}$ is what we might call a "first-order infinite number", bigger than every real; $1\epsilon^{-2}$ is what we might call a "second-order infinite number", bigger than every first-order infinite number; and so on. Thus, the formal Laurent series are quite adequate to address the specific problem described by Wirth et al. (2017) in which an infinite negative reward is required when the RL agent kills the cancer patient.

There are natural ways to define arithmetic on formal Laurent series, but we will avoid those details here. The advantage of the formal Laurent series number system is that it is relatively concrete, compared to the more abstract hyperreal or surreal numbers discussed below.

Example 7 *(Examples of formal Laurent series comparisons)*

1. *Consider $A = 5\epsilon^{-1} - 2\epsilon^0 + 3\epsilon^1 + 4\epsilon^2$ and $B = 5\epsilon^{-1} - 2\epsilon^0 + 1\epsilon^1 + 4\epsilon^2 + 5\epsilon^6$. The ϵ^{-1}- and ϵ^0-coefficients of A and B are equal, so we compare their ϵ^1-coefficients. A has an ϵ^1-coefficient of 3 and B has an ϵ^1-coefficient of 1, and $3 > 1$, so $A > B$.*

2. *Consider $A = 999999\epsilon^5$ and $B = 0.00001\epsilon^4$. We consider A to have an ϵ^4-coefficient of 0, which is smaller than B's ϵ^4-coefficient (0.00001), so $A < B$.*

There is a natural way to consider formal Laurent series as a significantly-ordered structure, generalizing the notion of "significantly greater than" from Lemma 2.

Definition 8   *1. For every Laurent series $A = \sum_{i=-m}^{\infty} a_i \epsilon^i$, let $o(A) = -m$ (call this the order of A), and let $LC(A)$ be the $\epsilon^{o(A)}$-coefficient of A (call this the* leading coefficient *of A).*

2. Let $r > 0$ be any positive real number. For any formal Laurent series $A = \sum_{i=-m}^{\infty} a_i \epsilon^i$ and $B = \sum_{i=-n}^{\infty} b_i \epsilon^i$, we say $A \ll_r B$ if one of the following conditions holds:

- *$o(A) > o(B)$ and $LC(B) > 0$; or*
- *$o(A) < o(B)$ and $LC(A) < 0$; or*
- *$o(A) = o(B)$ and $LC(A) \leq LC(B) - r$.*

Lemma 9 *For any real $r > 0$, the formal Laurent series, considered as a significantly-ordered structure according to $\ll_r$, are non-Archimedean.*

Proof Let $x_0 = 0\epsilon^1$, $x_1 = r\epsilon^1$, $x_2 = 2r\epsilon^1$, and in general let $x_i = ir\epsilon^1$. Let $x_\infty = 1\epsilon^0$. Then each $x_i \ll_r x_{i+1}$, and yet each $x_i \ll_r x_\infty$. ∎

Unfortunately, although the formal Laurent series contain infinities and infinitesimals, in a sense we will make formal, they still do not contain "enough" infinities and infinitesimals to accomodate genuine AGI. To make this formal, we introduce a weaker notion of Archimedeanness.

Definition 10 *Suppose $(X, \ll)$ is a significantly-ordered structure. We define a new order $\ll'$ on X as follows. For any $x, y \in X$, we declare $x \ll' y$ if and only if there is a sequence $x_0, x_1, \ldots$ such that the following conditions hold:*

1. $x_0 = x$.

2. Each $x_i \ll x_{i+1}$.

3. Each $x_i \ll y$.

We say $(X, \ll)$ is 2-Archimedean if $(X, \ll')$ is Archimedean.

Proposition 11 *For any real $r > 0$, the formal Laurent series are 2-Archimedean when considered as a significantly-ordered structure as in Definition 8.*

We omit the proof of Proposition 11 because it is somewhat tedious and we do not need it anywhere later in the paper. It merely serves to explain why formal Laurent series are still not good enough to avoid the AGI roadblock from this paper. Just as we argued that a genuine AGI should be capable of engaging in tasks that involve inherently non-Archimedean rewards, by similar reasoning, an AGI should be capable of engaging in tasks that involve inherently non-2-Archimedean rewards. It is not hard to show that all the structures in Example 2 are non-2-Archimedean. Thus, replacing real number rewards by formal Laurent series rewards is not enough to remove the roadblock, but it would at least expand the types of rewards which are possible.

4.2.2 HYPERREAL NUMBERS

The field of mathematics where the calculus is formalized with infinite and infinitesimal quantities is called *nonstandard analysis* (Robinson, 1974). The numbers most commonly associated with this field are the so-called *hyperreal numbers*.

The hyperreal numbers can be introduced axiomatically or by means of a semi-constructive method which depends on usage of a certain black box, a device known as a *free ultrafilter*. Logicians have proven that free ultrafilters exist but that, unfortunately, it is impossible to concretely exhibit one. This severely limits (if not completely ruins) the practical usefulness of reinforcement learning with hyperreal rewards.

Nevertheless, the hyperreals might be useful for proving abstract structural properties about AGI[11]. It can be shown that the hyperreals are not 2-Archimedean, and indeed, not α-Archimedean for any ordinal α, where "α-Archimedean" refers to a certain natural weakening of 2-Archimedeanness. Thus, for the purpose of proving abstract theorems about RL agents with generalized rewards, the hyperreals would be more appropriate than the formal Laurent series.

4.2.3 SURREAL NUMBERS

All of the well-known non-Archimedean extensions of $\mathbb{R}$ (including formal Laurent series and hyperreals) are subsystems of the so-called *surreal numbers* (Conway, 2000; Knuth, 1974; Ehrlich, 2012). The surreal numbers were initially discovered during John Conway's attempts to study two-player combinatorial games like Go and Chess, so it would not be surprising if they turn out to be important in the eventual development of AGI.

Unlike the hyperreals, the construction of the surreal numbers does not depend on any non-constructive black boxes such as free ultrafilters. They are constructed as the union of a hierarchy S_α of subsystems where α ranges over the ordinal numbers. Assuming that agents with AGI are implemented using computers with no additional power beyond the Church-Turing Thesis, then for the purposes of AGI, it would be appropriate to restrict our attention to some computable subset of the surreal numbers, which would presumably be the union of some hierarchy C_α where α ranges over the computable ordinal numbers. For any particular level C_α in this hierarchy, we could consider the sub-universe E_α of surreal-reward RL environments with rewards restricted to C_α.

Assuming AGI agents are Turing computable, no individual AGI can possibly comprehend codes for all computable ordinals, because the set of codes of computable ordinals is badly non-computably-enumerable. This is profound, because it seems to suggest that any particular AGI can only comprehend RL environments in E_α if that AGI can comprehend α. In other words, for any particular RL environment e with computable surreal number rewards, there must be some minimal computable ordinal α such that e has rewards from E_α; if an AGI is not intelligent enough to comprehend α, then it seems like there should be no way for the AGI to comprehend e either[12]. We would submit this state of affairs as

---

11. Similar to the way we use free ultrafilters in (Alexander, 2019a) to obtain comparators of the utility-maximizing ability of traditional deterministic RL agents, and prove structural properties about said comparators. In fact, in that paper, we essentially independently re-invented the free ultrafilter construction of the hyperreals, without realizing it at the time!

12. This situation is highly reminiscent of (Hibbard, 2011).

strong evidence in favor of our thesis (Alexander, 2019b) that a machine's intelligence ought to be measured in terms of the computable ordinals which the machine comprehends.

The above paragraph points at a possible joint path toward AGI incorporating both machine learning and symbolic logic, perhaps a much-needed reconciliation of these two approaches.

4.3 Alternate number systems: tentative verdict

For many simple environments not too far outside of traditional RL, formal Laurent series could probably serve as a fairly practical number system. But formal Laurent series have limitations which suggest that RL with formal Laurent series rewards will not be enough to reach AGI, for the exact same reason that RL with real number rewards will not be enough.

Because of their dependence on free ultrafilters, the hyperreal numbers will probably never be of practical use as RL rewards, but it could conceivably be possible to use them to prove abstract structural results about AGI from a bird's-eye view.

The surreal numbers (or a computable subset thereof) seem like the most promising candidate for RL rewards that could plausibly lead to AGI. We would certainly hesitate to call them "practical", though. To work with any but the most trivial of surreal numbers, one would need to implement sophisticated symbolic-logical machinery, and that's just to get one's foot in the door. This does, however, offer a ray of hope in the sense that doing deep learning techniques with surreal numbers would be a way to combine both symbolic logic and statistical methods into a joint approach.

5. Conclusion

In traditional reinforcement learning, utility-maximizing agents interact with environments, receiving real (or rational) number rewards in response to actions, and using those rewards to update their behavior. We have argued that the decision to limit rewards to real numbers is inappropriate in the context of AGI, because the real numbers have the Archimedean property, which makes it impossible to use them to accurately portray the value of actions when a task involves inherently non-Archimedean rewards. Thus, we argue, traditional RL cannot possibly lead to AGI, because a genuine AGI should have no trouble comprehending and at least attempting tasks that inherently involve non-Archimedean rewards. We suggested two possible ways to modify traditional reinforcement learning to fix this bug: switch to preference-based reinforcement learning, or else generalize reinforcement learning to allow rewards from a non-Archimedean number system.

Acknowledgments

References

Al-Dhalimy, H., and Geyer, C. J. 2016. Surreal Time and Ultratasks. *The Review of Symbolic Logic* 9(4):836–847.

Alexander, S. A. 2019a. Intelligence via ultrafilters: structural properties of some intelligence comparators of deterministic Legg-Hutter agents. *Journal of Artificial General Intelligence* 10:24–45.

Alexander, S. A. 2019b. Measuring the intelligence of an idealized mechanical knowing agent. In *Cognition, Interdisciplinary Foundations, Models, and Applications (CIFMA)*. Springer.

Andréka, H.; Madarász, J. X.; Németi, I.; and Székely, G. 2012. A logic road from special relativity to general relativity. *Synthese* 186(3):633–649.

Archimedes. 1897. On the Sphere and Cylinder. In Heath, T., ed., *The works of Archimedes*. Cambridge University Press.

Bair, J.; Błaszczyk, P.; Ely, R.; Henry, V.; Kanovei, V.; Katz, K. U.; Katz, M. G.; Kutateladze, S. S.; McGaffey, T.; Schaps, D. M.; Sherry, D.; and Shnider, S. 2013. Is mathematical history written by the victors? *Notices of the American Mathematical Society* 60(7):886–904.

Benci, V.; Horsten, L.; and Wenmackers, S. 2013. Non-Archimedean probability. *Milan Journal of Mathematics* 81(1):121–151.

Blum, D., and Holling, H. 2017. Spearman's law of diminishing returns. A meta-analysis. *Intelligence* 65:60–66.

Chen, L. forthcoming. Infinitesimal gunk. *Journal of Philosophical Logic*.

Conway, J. H. 2000. *On Numbers and Games*. CRC Press, 2nd edition.

Ehrlich, P. 2012. The absolute arithmetic continuum and the unification of all numbers great and small. *Bulletin of Symbolic Logic* 18:1–45.

Euclid. 2007. Book V: Theory of Proportion. In Casey, J., ed., *First Six Books of the Elements of Euclid*. Project Gutenburg.

Hernández-Orallo, J. 2019. AI Generality and Spearman's Law of Diminishing Returns. *Journal of Artificial Intelligence Research* 64:529–562.

Hibbard, B. 2011. Measuring agent intelligence via hierarchies of environments. In *International Conference on Artificial General Intelligence*, 303–308. Springer.

Knuth, D. E. 1974. *Surreal numbers: a mathematical novelette*. Addison-Wesley.

Legg, S., and Hutter, M. 2007. Universal intelligence: A definition of machine intelligence. *Minds and machines* 17(4):391–444.

Livingston, S.; Garvey, J.; and Elhanany, I. 2008. On the broad implications of reinforcement learning based AGI. In *International Conference on Artificial General Intelligence*, 478–482.

Mill, J. S. 2016. Utilitarianism. In *Seven masterpieces of philosophy*. Routledge. 337–383.

Narens, L. 1974. Measurement without Archimedean axioms. *Philosophy of Science* 41(4):374–393.

Niederée, R. 1992. What do numbers measure?: A new approach to fundamental measurement. *Mathematical Social Sciences* 24(2-3):237–276.

Plato. 1997. Protagoras. In Cooper, J. M.; Hutchinson, D. S.; et al., eds., *Plato: complete works*. Hackett Publishing.

Pohlers, W. 2008. *Proof theory: The first step into impredicativity*. Springer.

Rathjen, M. 2006. The art of ordinal analysis. In *Proceedings of the International Congress of Mathematicians*, volume 2, 45–69.

Reeder, P. F. 2012. *Infinitesimals for Metaphysics: Consequences for the Ontologies of Space and Time*. Ph.D. Dissertation, The Ohio State University.

Rizza, D. 2016. Divergent Mathematical Treatments in Utility Theory. *Erkenntnis* 81(6):1287–1303.

Robinson, A. 1974. *Non-standard analysis*. Princeton University Press.

Rosinger, E. E. 2007. Cosmic Contact: To Be, or Not To Be Archimedean? *arXiv preprint physics/0702206*.

Skala, H. J. 1975. *Non-Archimedean utility theory*. D. Reidel Publishing.

Spearman, C. 1927. *The abilities of man*. Macmillan.

Tall, D. 1980. Looking at graphs through infinitesimal microscopes, windows and telescopes. *The Mathematical Gazette* 64:22–49.

Veldhuizen, T. L. 2003. C++ Templates are Turing Complete. Technical report, Indiana University.

Wang, P., and Hammer, P. 2015. Assumptions of decision-making models in AGI. In *International Conference on Artificial General Intelligence*, 197–207. Springer.

Wirth, C.; Akrour, R.; Neumann, G.; and Fürnkranz, J. 2017. A survey of preference-based reinforcement learning methods. *The Journal of Machine Learning Research* 18(1):4945–4990.

Zhao, Y.; Kosorok, M. R.; and Zeng, D. 2009. Reinforcement learning design for cancer clinical trials. *Statistics in medicine* 28(26):3294–3315.