A mathematical proof of physics, obtained by formalizing the scientific method

Alexandre Harvey-Tremblay[1]

[1]Independent scientist, aht@protonmail.ch

May 31, 2021

Abstract

It is normally expected that the laws of physics are the general end-product of the scientific process. In this paper, consistently with said expectation, I produce a model of science using mathematics, then I use it to derive the laws of physics by applying the (formalized) scientific method to the model. Specifically, the laws of physics are derived as the probability measure that maximizes the quantity of information produced by the scientific method as the observer traces a path in the space of all possible experiments. In this space said probability measure describes a general linear ensemble of programs which is a foundation sufficient to express all known physics. Since the definitions are purely mathematical and contain no physical baggage of any kind, yet are nonetheless able to derive the laws of physics from first principles, then it follows that the present derivation of said laws, as it is ultimately the product of the (formalized) scientific method, is the minimal foundation of physics as well as its philosophical less controversial formulation. We end with applications of the model to open problems of physics and produce testable predictions.

Contents

1 Introduction **3**
- 1.1 Universal Facts . . . 3
- 1.2 Axiomatic foundation of finite theories . . . 4
  - 1.2.1 Connection to finitely axiomatic systems . . . 6
  - 1.2.2 Axiomatic information . . . 6
- 1.3 Philosophy of facts . . . 7
  - 1.3.1 State of affairs . . . 8
  - 1.3.2 Inconsistencies in the state of affairs...? . . . 9
  - 1.3.3 Intuition: an experimental system that is purely mathematical . . . 10

2 Formal Science **11**
2.1 Axiomatic foundation of science 11
2.1.1 The Fundamental Theorem of Science 13
2.2 Axiomatic foundation of reality 13
2.3 Axiomatic foundation of physics 15

3 Main Result (Physics) **17**
3.1 Introductory investigation of key computing concepts 17
3.1.1 Halting probability of computer science 17
3.1.2 Quantum computing 18
3.1.3 Manifest-based computing... ? 19
3.2 Derivation 19
3.2.1 General linear computing ensemble 19
3.2.2 Prior 21
3.2.3 Representation 21
3.3 Geometric Ensemble 23
3.3.1 Manifests are tensor products of states 24
3.3.2 General linear wave-function 24
3.4 Geometric Hilbert space 26
3.4.1 H is a vector space 26
3.4.2 H is a Hilbert space 27
3.4.3 Inner product — Example in $\mathbb{G}(2, \mathbb{R})$ 27
3.4.4 Observable 28
3.4.5 Observable — Example in $\mathbb{G}(2, \mathbb{R})$ 28
3.4.6 Observable — Geometric spectral theorem 29

4 Applications **30**
4.1 The complex Hilbert space is a special case 30
4.2 Spinor group wave-function 30
4.3 The probability density of an event 31
4.4 The collapse 32
4.5 Quantum gravity (probability measure support) 32
4.6 Quantum gravity + Standard model (probability measure support) 33
4.7 Gauge-theoretical completeness 34
4.8 Lagrangian 34
4.9 Testable Prediction 35
4.9.1 Geometric interference 35

A Notation **38**

B Lagrange equation **39**
B.1 Multiple constraints 40
B.2 Multiple constraints - general case 40

C Quackery (intermission) **41**

1 Introduction

In classical philosophy an axiom is a statement which is self-evidently true such that it is accepted without controversy or question. But this definition has been retired in modern usage. Any so-called "self-evident" axiom can also be posited to be false and either choice of its truth-value yields a different model; the archetypal example being the parallel line postulate of Euclid, allowing for hyperbolic/spherical geometry when it is false. Consequently, in modern logic an axiom is simply a starting point for a premise, and in mathematics an axiom is a sentence of a language that is held to be true by definition.

A long standing goal of philosophy has been to find necessarily true principles that could be used as the basis of knowledge. For instance, the universal doubt method of Descartes had such a goal in mind. The 'justified true belief' theory of epistemology is another attempt with a similar goal. But, so far, all such attempts have exploits; the elimination of which is assumed, at best, to reduce the theory to a handful of statements, rendering it undesirable as a foundation to all knowledge.

In epistemology, the Gettier problem[1] is a well known objection to the belief that knowledge is that which is both true and justified, relating to a family of counter-examples. All such counter-examples rely on the same exploit: if the justification is not 'air-tight' then there exists a case where one is right by pure luck, even if the claim were true and believed to be justified. For instance, if one glances at a field and sees a shape in the form of a dog, one might think he or she is justified in the belief that there is a dog in the field. Now suppose there is a dog elsewhere in the field, but hidden from view. The belief "there is a dog in the field" is justified and true, but it is not knowledge because it is only true by pure luck.

Richard Kirkham[2] proposed to add the criteria of infallibility to the justification. This eliminates the exploit, but it is an unpopular solution because adding it is assumed to reduce epistemology to radical skepticism in which almost nothing is knowledge.

I propose the concept of the universal fact as a new candidate to serve as the foundation to knowledge. Due to their construction universal facts are sufficiently strong to be infallible, yet are sufficiently expressive to form a Turing complete theory. Universal facts be will the primary subject matter of our model of science and they are revealed and verified by the scientific method.

1.1 Universal Facts

Many philosophies discuss facts, but it appears they all missed the mark on what a fact actually is (in terms of a precise exploit-free mathematical definition). The archetypal example of a fact is given in many philosophical textbooks: "$1 + 1 = 2$", is in fact not a fact. Here I will provide what I believe to be the correct definition of a fact, and since the definition is formal (and is universal in the computer-theoretical sense) I will use the term *universal fact* to distinguish the concept from other definitions.

Specifically, the sentence "$1 + 1 = 2$" halts on some Turing machine, but not on others and thus is not a universal fact. Instead consider the sentence $\text{PA} \vdash [1 + 1 = 2]$ to be read as "Peano's axioms prove that $1 + 1 = 2$". Such a statement embeds as a prefix the set of axioms in which it is provable. One can deny that $1 + 1 = 2$ (for example, a trickster could claim binary numbers, in which case $1+1 = 10$), but if one specifies the exact axiomatic basis in which the claim is provable, a trickster would find it harder to find an exploit. Nonetheless, even with this improvement, an exploit always exists because in the general case a trickster can provide a Turing machine for which $\text{PA} \vdash [1 + 1 = 2]$ does not halt.

If we use the tools of theoretical computer science and observe an equivalence between facts and programs, we can cure the concept of a fact of all of its exploits:

Definition 1 (Universal Fact)**.** *Let* $\mathbb{L}$ *be the set of all sentences with alphabet* Σ*. A universal fact is a pair* (TM, p) *of sentences from* $\mathbb{L}$ *such that a universal Turing machine* UTM *halts for it:*

$$\textit{iff } \text{UTM}(\text{TM}, p) \textit{ halts, then } (\text{TM}, p) \textit{ is a universal fact} \tag{1}$$

A universal Turing machine UTM which takes a Turing machine TM and a sentence p as inputs, will halt iff p halts on TM. Thus the fact that p halts on TM is indeed a universal fact because it is verifiable on all universal Turing machines. This definition cures the concept of its exploits: "I believe $\{(\text{TM}_1, p_1), (\text{TM}_2, p_2), \ldots, (\text{TM}_n, p_n)\}$ are facts, why? Answer: I verified that they halt on a universal Turing machine. Anyone else can too, by definition, therefore I am infallibly justified in my belief that these facts hold."

The second objection is that infallible justified true beliefs collapse epistemology to radical skepticism, where at best only a handful of statements constitute knowledge. However, the set of all universal facts constitute the entire domain of the universal Turing machine, and thus the expressiveness of universal facts must be on par with any Turing complete language. There exists no greater expressivity for a formal language than Turing completeness.

1.2 Axiomatic foundation of finite theories

We can use universal facts to define a new, significantly more flexible foundation to mathematics, especially when it comes to formal theories that contain finitely many theorems or sentences. When it comes to formulating a scientific theory of a finite physical system one can intuit why that would be a desirable feature. Working with finite theories using the typical tools of mathematics is mostly ineffective, because all such theories are decidable and thus completely solvable in principle. Furthermore, even tools such as complexity theory require the size of the input to be n, allowing for arbitrarily large sizes of input to produce an effective classification system. Instead of defining a mathematical theory as a finite deductive system of axioms, which typically entails infinitely many

theorems, why not define it as a finite (or in some cases even infinite) set of universal facts?

To distinguish our definition of a formal theory with that of the literature, we will call our definition a *manifest theory.*

Definition 2 (Manifest Theory)**.** *A manifest theory* $\mathbb{T}$ *is defined as a set of universal facts:*

$$\mathbb{T} := \{(\mathrm{TM}_1, p_1), (\mathrm{TM}_2, p_2), \dots\} \tag{2}$$

The set can be either finite or infinite, and it can be either decidable or non-decidable.

For a manifest theory, universal facts replace the normal role of both axioms and theorems and instead form a single verifiable atomic concept constituting a unit of epistemological knowledge. Let me explicitly point out the difference between the literature definition of a formal theory and ours: for the former, its theorems are a subset of the sentences of $\mathbb{L}$ — whereas for a manifest theory, its elements are pairs of $\mathbb{L} \times \mathbb{L}$ which halts on a UTM.

Theorem 1 (Incompleteness Theorem)**.** *If* $\mathbb{T} = \mathrm{Dom}(\mathrm{UTM})$*, then* $\mathbb{M}$ *is recursively enumerable (and non-decidable). The proof follows from the domain of a universal Turing machine being non-computable.*

Note on the upcoming notation: we will designate f as elements of $\mathbb{T}$, and $\pi_1(f)$ and $\pi_2(f)$ designate the first and second projection of the tuple f, respectively. Thus $\pi_1(f)$ is the TM associated with f, and $\pi_2(f)$ is the p associated with f. If applied to a set of tuples, then $\pi_1(\mathbb{T})$ return the set of all p in $\mathbb{T}$ and $\pi_2(\mathbb{T})$ returns the set of all TM in $\mathbb{T}$.

Definition 3 (Valid Sentences)**.** *The valid sentences of* $\mathbb{T}$ *are defined as the set of all* p *in* $\mathbb{T}$*:*

$$\mathbb{V} := \{p : p \in \pi_2(\mathbb{T})\} \tag{3}$$

Definition 4 (Atomic Solver)**.** *The atomic solvers of* $\mathbb{T}$ *are defined as the set of all* TM *in* $\mathbb{T}$*:*

$$\mathbb{A} := \{\mathrm{TM} : \mathrm{TM} \in \pi_1(\mathbb{T})\} \tag{4}$$

Definition 5 (Composite Solver)**.** *Let* r *be the output of* $\mathrm{UTM}(\mathrm{TM}, p)$*. If* $r \in \mathbb{V}$*, then the tuple* $(\mathrm{TM}, \mathrm{TM}_r)$*, where* $(\mathrm{TM}_r, r) \in \mathbb{T}$*, is a composite solver. The size of a tuple of composite solvers can extend to any finite number of elements.*

Definition 6 (Spread (of a sentence))**.** *The set of all atomic solvers in* $\mathbb{T}$ *in which a sentence is repeated is called the spread of the sentence:*

$$\forall f \in \mathbb{T}[\pi_2(f) = p \implies \pi_1(f) \in \mathrm{spread}(p, \mathbb{T})] \tag{5}$$

For instance if $\mathbb{T} = \{(\mathrm{TM}_1, p_1), (\mathrm{TM}_2, p_1)\}$*, then the spread of* p_1 *is* $\{\mathrm{TM}_1, \mathrm{TM}_2\}$*.*

Definition 7 (Scope (of a solver))**.** *The set of valid sentences in* $\mathbb{T}$ *in which a solver is repeated is called the scope of the solver:*

$$\forall f \in \mathbb{T}[\pi_1(f) = \mathrm{TM} \implies \pi_2(f) \in \mathrm{scope}(\mathrm{TM}, \mathbb{T})] \tag{6}$$

for instance if $\mathbb{T} = \{(\mathrm{TM}_1, p_1), (\mathrm{TM}_1, p_2)\}$*, then the scope of* TM_1 *is* $\{p_1, p_2\}$*.*

1.2.1 Connection to finitely axiomatic systems

We can, of course, connect our construction to a finitely axiomatic system:

Definition 8 (Finitely Axiomatic Representation)**.** *Let* FAS *be a finitely axiomatic system, let* $\mathbb{T}$ *be a manifest theory and let* $\mathrm{solver}_{\mathrm{FAS}}$ *be a function which recursively enumerates the theorems of* FAS*. Then* FAS *is a finitely axiomatic representation of* $\mathbb{T}$ *iff:*

$$\forall (s_1, s_2) \in \mathbb{L} \times \mathbb{L} \quad [\mathrm{solver}_{\mathrm{FAS}}(s_1, s_2) \textit{ halts iff } (s_1, s_2) \in \mathbb{T}] \tag{7}$$

Definition 9 (Domain (of FAS))**.** *Let* FAS *be a finitely axiomatic system, let* $\mathbb{T}$ *be a manifest theory and let* $\mathrm{solver}_{\mathrm{FAS}}$ *be a function which recursively enumerates the theorems of* FAS*. Then the domain of* FAS*, denoted as* Dom(FAS)*, is the set of all pairs* $(s_1, s_2) \in \mathbb{L} \times \mathbb{L}$ *which halts for* $\mathrm{solver}_{\mathrm{FAS}}$*.*

Definition 10 (de-facto-isomorphism)**.** *Two finitely axiomatic representations* FAS_1 *and* FAS_2 *are de-facto-isomorphic iff* $\mathrm{Dom}(\mathrm{FAS}_1) = \mathrm{Dom}(\mathrm{FAS}_2)$*.*

Theorem 2 (Principle of Computational Equivalence[3])**.** *If* $\mathbb{T} = \mathrm{Dom}(\mathrm{UTM})$ *then all Turing complete finitely axiomatic systems are de-facto-isomorhic representations of* $\mathbb{T}$*. Furthermore, their solver function is a universal Turing machine. The proof follows because* Dom(UTM) *includes all universal facts.*

1.2.2 Axiomatic information

Although we *can* connect the formulation of a manifest theory to a finitely axiomatic representation, we will find that it is more advantageous for the purposes of constructing a model of science to study manifest theories using the formalism of universal facts we have introduced. We can understand the elements of any particular manifest theory as having been 'picked', in some sense, from the set of all possible universal facts. If the pick is random and described as a probability measure ρ, we can quantify the quantity of information of the pick using the entropy, and thus we can associate the notion of information to the manifest theory.

Definition 11 (Axiomatic Information)**.** *Let* $\mathbb{D}$ *be the domain of a universal Turing machine* $\mathbb{D} = \mathrm{Dom}(\mathrm{UTM})$ *(full theory) or of a subset thereof* $\mathbb{D} \subset \mathrm{Dom}(\mathrm{UTM})$ *(toy theory). Then, let* $\rho : \mathbb{D} \to [0, 1]$ *be a probability measure over* $\mathbb{D}$*. Finally, let* $\mathbb{T}$ *be a manifest theory subset of* $\mathbb{D}$*. The axiomatic information of a single element of* $\mathbb{T}$ *is quantified as the entropy of* ρ*:*

$$S = -\sum_{q \in \mathbb{D}} \rho(q) \ln \rho(q) \tag{8}$$

For instance, a well-known (non-computable) probability measure regarding a sum of prefix-free programs is the Halting probability[4] of computer science:

$$\Omega = \sum_{p \in \mathrm{Dom(UTM)}} 2^{-|p|} \implies \rho(p) = 2^{-|p|} \tag{9}$$

The quantity of axiomatic information of a given manifest theory (and especially its maximization), rather than any particular set of axioms, will be the primary quantity of interest for the production of a maximally informative theory in this framework.

1.3 Philosophy of facts

Now and before we enter section 2, we will try to ease the transition to a full blown purely mathematical model of science, by attempting to set and perhaps even 'reset' the intuition to be in line with the requirements of the model.

Most are likely to fall in either one of two camps: the world is made of things or is made of facts. The first option is definitely the mainstream belief, but the second one nonetheless has a sizeable following. For instance, in the *Tractatus Logico-philosophicus*[5], Wittgenstein produces these two statements: 1) "The world is everything that is the case" and 2) "The world is the totality of facts, not of things". For him, the world is clearly made of facts. We also note John A. Wheeler in *Information, physics, quantum: The search for links*[6] stating "[...] every physical quantity, every it, derives its ultimate significance from bits, binary yes-or-no indications, a conclusion which we epitomize in the phrase, it from bit". So which is it: facts or things?

I submit that there is a fundamental problem with considering that the world is made of things rather than facts, and I believe the problem is insurmountable. I have tried to explain it before using a different formulation, however using universal facts now makes it incredibly difficult (I believe) for anyone not to see and understand it.

1. Universal facts constitute the set of all infallible-justified-true-beliefs.
2. Person A claims to hold a conception of the world which is not supported by universal facts alone (e.g. ice cream is a thing, not a fact).
3. It thus follows, necessarily, that person A's conception of the world cannot be an infallible-justified-true-belief.
4. Person A's conception of the world must therefore be either fallible, unjustified or false.

Believing that the world itself is made of things rather than facts is a version of the Gettier problem (perhaps even its most general version). Indeed, if person A is aware of some information which transforms according to a certain type of rule (e.g. produces a certain shape or structure — say the shape of a dog in a field), but then believes there is a 'thing' that exists in addition to said information, then said belief cannot constitute knowledge in the epistemological sense because A is not infallibly justified in believing the existence of the 'additional thing'. The existence of universal facts as a Turing complete theory, since it can express any concept that *can* be expressed, precludes any and all belief in said 'additional thing' to be infallible.

Consequently, I believe it to be simpler to reproduce all human knowledge in terms of universal facts than it is to overcome this problem. Doing so also cures quite a lot of problems in philosophy, science and physics. To get the program started, I have thus produced here a formal theory of science which is able to recover the laws of physics by using universal facts as the starting point. It took me a preliminary period of ten-ish years juggling with the concept of universal facts and similar notions, and a subsequent five year period to build the overall model, but it appears to be well worth the effort.

I, for one, will gladly trade in ice cream *as a thing* for ice cream *as a fact*, if the latter is infallible and the former isn't.

1.3.1 State of affairs

There is another conceptual trap to avoid regarding how a fact of the world can be a universal fact. For instance, one might say "If I define the store next door as having ice cream, it might not have it by the time I get there. Therefore, even something as innocuous as the store having ice cream cannot be a universal fact because there is always a possibility that the fact be violated in the future". As this could apply to any state of affairs, what, if anything, could truly be a universal fact in nature?

The trap is that this is the incorrect way to use universal facts to define reality. Instead of thinking about a situation as 'past implies present' or 'present implies future', one ought to think about it as 'present implies past'. Think of oneself as a forensic scientist: one walks into a crime scene, and to know what happened one has to reconstruct a model of the past based on the evidence available now. Thus, instead of thinking of ice cream as a thing they may or may not have at the store and this determines whether one will or will not have some later; instead one ought to axiomatically define his or her mouth as having ice cream in it, then figure out a plausible origin story for said ice cream.

The pick of the elements of the state of affairs from the set of all possible elements is axiomatic information. Thus, the statement "there is ice cream in my mouth right now" along with its origin story are, as an element of the state of affairs, axiomatic information. The corresponding universal fact would resemble this: ("Store had ice cream" $\wedge$ "I bought the ice cream" $\wedge$ (all other necessary conditions), "I have ice cream in my mouth"). The state of having ice cream in ones mouth, is verified by a logically-deducible causal history which

'explains' why it's in one's mouth now - and such causal history is sufficiently detailed for "I have ice cream in my mouth" to be provable from said history.

1.3.2 Inconsistencies in the state of affairs...?

A fatal inconsistency for this framework would be a proof that a given universal fact both [HALT] and [NOT HALT] on a UTM. By definition of a UTM, this cannot happen lest the machine was not a UTM to begin with. Thus, we should be safe from contradictions at this level. But what about the other "levels"?

What if the state of affairs contains both "I have ice cream in my mouth" and "I do not have ice cream in my mouth" at the same time — how much of a problem is it for the framework? For the record, let me state that I can construct a computer program that halts for both "I have ice cream in my mouth" and "I do not have ice cream in my mouth" and the universe will not implode:

1. If (p="I have ice cream in my mouth" or p="I do not have ice cream in my mouth") then
2. return 1;
3. else (loop())

A-priori, the words "I have ice cream in my mouth" are just symbols. The tendency of the human mind to attribute a special meaning or a special exclusionary requirement to these symbols is a step that occurs later in the framework, after the selection of the finitely axiomatic representation. The appearance of an irreconcilable contradiction is the result of attributing said special meaning too soon and without 'respecting' the steps of the framework. Let's see with an example.

Suppose one has a sizeable manifest theory which may contain a plurality of valid sentences $p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8 \dots$ such that the negation of some, but not all, are also present as valid sentences: for instance, the sentences p_6 and p_7 might be $p_6 = \neg p_4, p_7 = \neg p_5$. In this case, a finitely axiomatic representation of this theory could take the form of a para-consistent theory which would contain some contradictory elements, yet would not support the principle of explosion (a contradiction in a para-consistent theory is contained and does **not** necessarily entail that everything else in the theory is also contradicted).

Now, let us call the $p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8 \dots$ the various flavours or recipes of ice cream. It could be that the Italians define ice cream in a certain way, and the British define it in a slightly different way. Recall that universal facts are pairs which contain an atomic solver and a valid sentence. The atomic solver of p_i is the 'standard' under which the flavour qualifies as real ice cream. A p_i with a large spread is considered real ice cream by most atomic solvers (i.e. vanilla or chocolate ice cream), and one with a tiny spread would be considered real ice cream by only a few (i.e. tofu-based ice cream). Then, within this example, the state of both [having] and [not having] ice cream in ones mouth, simply implies one is eating tofu-based ice cream; it's "real" ice cream, but then again it's not *real* ice-cream...

This multi-ice-cream-standard interpretation is only one example to reconcile the situation. In principle, there may exist infinitely many alternative ways to do so, and such is dependant on the choice of a de-facto-isomorphic finitely axiomatic representation. Reality is of a complexity such that a one-size-fits-all definition may not fit all concepts, and further competing definitions might exist; a chair may be a chair according to a certain definition, but not according to another. The existence of many definitions for one concept is a part of reality, and the mathematical framework which correctly describes reality to ought be sufficiently flexible to handle this, without itself exploding into a contradiction.

A special case occurs if both p_8 and its negation $\neg p_8$ were to be valid sentences of $\mathbb{T}$ while also having the same atomic solver. In this case, a finitely axiomatic representation of $\mathbb{T}$ which attributes a semantic to the $\neg$ symbol, might qualify said atomic solver as being inconsistent. Is this a problem for the framework? Answer: no, in fact it's a necessary part thereof. Indeed, having both p_8 and $\neg p_8$ with the same atomic solver be elements of $\mathbb{T}$ means one has verified that said atomic solver is inconsistent. Knowing that an atomic solver is inconsistent and also why it is inconsistent *is* valuable knowledge to have and comes at a cost! One has to prove to oneself that a given definition is inconsistent by trying it out against multiple instances of a concept, and those 'trials' are all part of the state of affairs.

1.3.3 Intuition: an experimental system that is purely mathematical

1. Assume person A knows (at least) one universal fact. For instance $(\text{solver}_{\text{PA}}, "1+1=2")$, where $\text{solver}_{\text{PA}}$ is a solver for Peano's axioms (PA).
2. How does A knows it is a universal fact and not just a false claim?
3. Either A verifies it himself or herself, or A is satisfied that it has been verified.
4. How can A verify a universal fact?
5. Answer: A must bring the program to termination on a Turing machine. For relatively simple facts, such as $(\text{solver}_{\text{PA}}, "1+1=2")$, A can likely verify it in his or her mind. But, for more complicated universal facts, it may take considerably longer and could be quite difficult.
6. In any case, A having knowledge of a universal fact implies A is satisfied that a sequence of computing has taken place somewhere somehow to verify that fact.

Comparatively, how much faith would A have in someone claiming to know that an arbitrary program halts, without him or her showing A the proof nor having verified it to completion on a Turing machine. Answer: none. Just like positing axioms implies a collection of theorems, positing a state of affairs made of universal facts, such that one is satisfied that they are indeed universal facts, implies or necessitates a computational verification of these claims.

2 Formal Science

2.1 Axiomatic foundation of science

The fundamental object of study of science is not the electron, the quark or even super-strings, but the reproducible experiment. An experiment represents an 'atom' of verifiable knowledge.

Definition 12 (Experiment)**.** *Let* (TM, p) *be a pair comprising of two sentences of a language* $\mathbb{L}$*. The first sentence,* TM*, is called the protocol. The second sentence,* p*, is called the hypothesis. Let* UTM *be a universal Turing machine. If* $\mathrm{UTM}(\mathrm{TM}, p)$ *halts then the pair* (TM, p) *is said to be an experiment. In this case, we say that the protocol verifies the hypothesis. If* $\mathrm{UTM}(\mathrm{TM}, p)$ *does not halt, we say that the pair fails to verify the hypothesis.*

$$\mathrm{UTM}(\mathrm{TM}, p) \begin{cases} \mathrm{HALT} & \implies \textit{the experiment verifies } p \\ \neg\,\mathrm{HALT} & \implies \textit{the pair fails verification} \end{cases} \tag{10}$$

Of course, in the general case, as per the halting problem there exists no decidable function which can determine which pair is an experiment and which pair fails verification.

An experiment, so defined, is formally reproducible. I can transmit, via fax or other telecommunication medium, the pair (TM, p) to another experimentalist, and I would know with absolute certainty that he or she has everything required to reproduce the experiment to perfection.

Theorem 3 (Formal Reproducibility)**.** *Experiments are formally reproducible.*

Proof. Let UTM and UTM' each be a universal Turing machine. For each pair $\mathrm{UTM}(\mathrm{TM}, p)$ which halts on UTM, there exists a computable function, called an encoding function, which maps said pairs as $\mathit{encode}(\mathrm{TM}, p) \to (\mathrm{TM}', p')$ such that (TM', p') halts for UTM'. The existence of such function is guaranteed by (and equivalent to) the statement that any UTM can simulate any other. □

In the peer-reviewed literature, the typical requirement regarding the reproducibility of an experiment is that an expert of the field be able to reproduce the experiment, and this is of course a much lower standard than formal reproducibility which is a mathematically precise definition. Here, for the protocol TM to be a Turing machine, the protocol must specify all steps of the experiment including the complete inner workings of any instrumentation used for the experiment. The protocol must be described as an effective method equivalent to an abstract computer program. Should the protocol fail to verify the hypothesis, the entire experiment (that is the group comprising the hypothesis, the protocol and its complete description of all instrumentation) is rejected. For these reasons and due to the generality of the definition, I conjecture that the above definition is the only (sensible) definition of the experiment that is formally reproducible (as opposed to say "sufficiently reproducible for practical purposes").

Definition 13 (Empirical Evidence)**.** *The set of all pairs whose protocol* TM *verifies* p *is defined as the empirical evidence* Ev *of* p*:*

$$\mathrm{Ev}(p) := \mathrm{Dom}(\mathrm{UTM}, p) \tag{11}$$

where

$$\mathrm{Dom}(\mathrm{UTM}, p) := \{(\mathrm{TM}, p) : \textit{ where } \mathrm{UTM}(\mathrm{TM}, p) \textit{ halts } \} \tag{12}$$

Definition 14 (Scientific method)**.** *An algorithm which recursively enumerates the empirical evidence, or parts thereof, of an hypothesis or a set thereof, is called a scientific method.*

Empirical evidence is thus produced by the application of the scientific method to an hypothesis.

Theorem 4 (Scientific method)**.** *Existence of the scientific method.*

Proof. Consider a dovetail program scheduler which works as follows.

1. Sort all pairs of sentences of $\mathbb{L} \times \mathbb{L}$ in shortlex. Let the ordered pairs (TM_1, p_1), (TM_2, p_1), (TM_1, p_2), (TM_2, p_2), (TM_3, p_1), ... be the elements of the sort.
2. Take the first element of the sort, $\mathrm{UTM}(\mathrm{TM}_1, p_1)$, then run it for one iteration.
3. Take the second element of the sort, $\mathrm{UTM}(\mathrm{TM}_2, p_1)$, then run it for one iteration.
4. Go back to the first element, then run it for one more iteration.
5. Take the third element of the sort, $\mathrm{UTM}(\mathrm{TM}_1, p_2)$, then run it for one iteration.
6. Continue with the pattern, performing iterations one by one, with each cycle adding a new element of the sort.
7. Make note of any pair (TM_i, p_j) which halts.

This scheduling strategy is called dovetailing and allows one to enumerate the domain of a universal Turing machine recursively, without getting stuck by any singular program that may not halt. Progress will eventually be made on all programs... thus producing a recursive enumeration. □

Dovetailing is of course a simple/non-creative approach to the scientific method. The point here was only to show existence of such an algorithm, not to find the optimal one.

Definition 15 (Scientific theory)**.** *Let* ST *be a finitely axiomatic representation. If* $\mathrm{Ev}(p) = \mathrm{Dom}(\mathrm{ST})$*, then* ST *is a scientific theory of* p*.*

2.1.1 The Fundamental Theorem of Science

With these definitions we can prove, from first principle, that the possibility of falsification is a necessary consequence of the scientific method.

Definition 16 (Predictive theory)**.** *Let* $\mathbb{D}$*, called the 'collected scientific data' or just 'the data', be a subset of the empirical evidence of* p*:*

$$\mathbb{D} \subset \mathrm{Ev}(p) \tag{13}$$

A finitely axiomatized representation is called a predictive theory PT *of* p *if* $\mathbb{D} \subset \mathrm{Dom}(\mathrm{PT})$*. The set* $\mathbb{P}$*, called the predictions of* PT*, is defined as:*

$$\mathbb{P} := \mathrm{Dom}(\mathrm{PT}) \setminus \mathbb{D} \tag{14}$$

Predictive theories are thus supported by the data, but may diverge outside of this support.

Theorem 5 (The Fundamental Theorem of Science)**.** *If the empirical evidence of* p *is recursively enumerable, but not decidable, then the empirical evidence of* p *has measure* 0 *over the set of all possible predictive theories of* p*.*

Proof. The empirical evidence of p is unique, yet —excluding atypical cases $\mathrm{Ev}(p)$ where it is decidable— there exists countably infinitely many predictive theories of p, for any set of data $\mathbb{D}$. Finally, the measure of one element of a countably infinite set is 0. □

Consequently, the fundamental theorem of science leads to the concept of falsification, as commonly understood in the philosophy of science and as given in the sense of Popper. It is (almost) certain that a non-decidable predictive theory will eventually be falsified.

2.2 Axiomatic foundation of reality

Definition 17 (Domain of science)**.** *We note* $\mathbb{S}$ *as the domain* (Dom) *of science. We can define* $\mathbb{S}$ *in reference to a universal Turing machine* UTM *as follows:*

$$\mathbb{S} := \mathrm{Dom}(\mathrm{UTM}) \tag{15}$$

Thus, for all pairs of sentences (TM, h)*, if* $\mathrm{UTM}(\mathrm{TM}, h)$ *halts, then* $(\mathrm{TM}, h) \in \mathbb{S}$*. It follows that all experiments are elements of the domain of science.*

Definition 18 (Manifest)**.** *A manifest* $\mathbf{m}$ *is a n-tuple constructed from elements of the domain of science.*

$$\mathbf{m} := \mathbb{S}^n \tag{16}$$

A manifest is therefore a tuple of experiments:

$$\mathbf{m} := ((\mathrm{TM}_1, p_1), \ldots, (\mathrm{TM}_n, p_n)) \tag{17}$$

We note that since a manifest may contain repetitions (experiments are formally reproducible) we have elected not to define $\mathbf{m}$ *as a set, but instead as a n-tuple to allow said repetitions. Quite remarkably, this tuple vs set (manifest vs manifest-theory) definition is the primary difference between formal theories in math versus those in science — we will investigate the consequences of this difference in great detail in the main results section.*

For a given manifest, the possibility exists that some hypotheses or, likewise, some protocols be repeated in the other tuples of the manifest. For instance it could be the case that within a manifest $p_2 = p_5$, or that $\mathrm{TM}_1 = \mathrm{TM}_6 = \mathrm{TM}_{21}$, etc. The set of all hypotheses for a given protocol is called the *scope* (of the protocol), and the set of all protocols for a given hypothesis is called the *spread* (of the hypothesis).

Definition 19 (The Fundamental Assumption of Science)**.** *Experiments are complete with respect to the state of affairs of reality.*

- *The state of affairs of reality is describable as a set of experiments. Therefore, the state of affairs is describable as a manifest.*
- *To each state of affairs corresponds a manifest.*
- *The manifest is a complete description of the state of affairs.*
- *A manifest is a "brute" description of reality in terms of experiments.*
- *If the assumption of science would be false, it would mean that there are elements of the instantaneous state of nature that are outside the domain of science... (intervention by an oracle?)*

Definition 20 (Experimental Space)**.** *Experimental space* $\mathbb{E}$ *is the "powertuple" of the domain of science:*

$$\mathbb{E} := \bigcup_{i=0}^{\infty} \mathbb{S}^i \tag{18}$$

All elements of experimental space are manifests, and all manifests are elements of experimental space.

Definition 21 (Toy Model)**.** *A subset of experimental space is called a 'toy model'. Note: some toy models may be decidable.*

2.3 Axiomatic foundation of physics

Recall that earlier we used a dovetailing algorithm in Theorem 4 as an implementation of the scientific method, and we claimed that although it was a possible strategy, it was not necessarily the optimal one. So what then is the optimal implementation of the scientific method applicable to a tuple of experiments? Well, I suppose it depends on what we mean by optimal. One might be tempted to search along the lines of an efficient algorithm, perhaps the most elegant one, or the one that uses the least amount of memory, etc., but thinking in those terms would be a trap — we must think a bit more abstractly than postulating or arguing for a specific implementation. Potentially, every manifest could in principle have its own best strategy. Every manifest is also finite, thus any application of the scientific method in experimental space, which continually acquires empirical evidence must follow a path in it. It is therefore more strategic overall to identify a condition applicable to all cases and allow for transformations between cases, which produces the implementation as a maximization problem.

The best strategy will be to *maximize the information* gained from the scientific method; and this means in the technical sense to maximize the entropy of a probability measure on experimental space. To embed the notion of information into the scientific method it must be connected to the mathematical theory of information, and this involves the notion of a random selection of an element from a set:

Definition 22 (The Fundamental Assumption of Physics)**.** *Let $\rho(\mathbf{m})$ be a probability measure on experimental space (full model) or a subset thereof (toy model). An observer, denoted as $\mathcal{O}$, is a point randomly selected from experimental space. Thus, with a probability measure that spawns the set $\mathbb{E}$, every point in experimental space qualifies as an observer. Specifically, an observer is:*

$$\mathcal{O}_i := (\mathbf{m}_i, \rho : \mathbb{E} \to [0, 1]) \tag{19}$$

The definition of the observer is a specialization of the definition of the manifest in the sense that a manifest is a point in experimental space, and the observer is a *randomly selected* point in experimental space (and thus the notion of information is associated to it). Note that even in typical physics, the observer (which is not mathematically integrated into the formalism... leading to a family of open problems regarding the 'observer effect') is associated to a random selection of an element from a set of possible observations. This 'effect' will eventually be revealed to be a consequence of the present definition. Here, the observer 'has knowledge' of a randomly selected state from the set of all possible experimental states, hence he or she is an observer in an information gathering sense, and consequently has the opportunity to implement an information producing version of the scientific method so as to realize its optimal implementation:

Definition 23 (Theory of Everything)**.** *The theory of everything (ToE) is the probability measure that maximizes the entropy of a path in* $\mathbb{E}$ *traced by an observer, for all possible observers.*

The laws of physics are thus formally defined as the *information-theoretical maximum* of the scientific method for all possible observers. As we will see in the main result, this will involve a sum of programs. Obviously, the path of an observer in $\mathbb{E}$ is constrained to experimental space:

Definition 24 (The Fundamental Assumption of Nature)**.** *The set of all constraints*[1] *on the entropy of the probability measure of the observer necessary to keep and transform the path in experimental space defines the universe (of the observer).*

Let us now discuss the philosophy of these definitions in more detail. Let us consider two conceptually similar cases in use in modern physics.

In a gauge theory the equivalent of the 'constraints on entropy', although seldom described in those terms, is a consequence that the Lagrangian of a wave-function remains invariant under the action of a group. In this case, said wave-function must remain within the 'space' generated by this group in order for the Lagrangian to retain its invariance. One may then gauge the local action of said invariant transformation to get a gauge field. The idea that said Lagrangian be invariant with respect to a certain group of transformations is the equivalent of the concept that a probability measure is bounded to a certain 'space', and the gauge field is what emerges as an element that prevents or compensates so that the wave-function remains within the designated 'space'. In our framework, the constraint that an observer remains in experimental space is to the universe what a group-invariant transformation of the Lagrangian is to a gauge field.

A similar concept also finds its way in statistical physics regarding the role that the constraints on entropy play in it and their meaning. Specifically, when one uses the method of the Lagrange multipliers to solve for $\rho(q)$ such that it maximizes the entropy subject to a constraint on the energy $\overline{E} = \sum_{q\in\mathbb{Q}} \rho(q)E(q)$, one will obtain the Gibbs measure $\rho(q) = \frac{1}{Z}\exp -\beta E(q)$. The standard interpretation of the constraint is that it represents the set of all possible measurements of this observable on the system such that it bounds its macroscopic description to an average energy $\overline{E}$. Additional constraints, such as a volume constraint $\overline{V} = \sum_{q\in\mathbb{Q}} \rho(q)V(q)$ or a particle number constraint $\overline{N} = \sum_{q\in\mathbb{Q}} \rho(q)N(q)$, can also be added. In the case of a gas in a box at thermal equilibrium, the constraint is realized by a corresponding instrument such as an energy-meter and a volume-meter, and in the case of an observer bounded by the laws of physics to remain in experimental space, the "instrument" which acts upon the measure to constrain it to its domain is the largest possible instrument: the universe.

[1] The exact definition of the constraints requires setup and is provided as (Equation 35) on page 20. Intuitively, it is simply the set of linear transformations of $\mathbb{E}$ onto itself.

3 Main Result (Physics)

Let us now use these definitions to derive the laws of physics from first principle.

3.1 Introductory investigation of key computing concepts

3.1.1 Halting probability of computer science

Let us start by maximizing the entropy of the random selection of p from Dom(UTM):

$$S = -\sum_{p \in \text{Dom(UTM)}} \rho(p) \log_2 \rho(p) \tag{20}$$

subject to these constraints:

$$\sum_{p \in \text{Dom(UTM)}} \rho(p) = 1 \tag{21}$$

$$\sum_{p \in \text{Dom(UTM)}} \rho(p)|p| = \overline{|p|} \tag{22}$$

Using the method of the Lagrange multipliers, the result is the Gibbs measure (where D is a Lagrange multiplier):

$$\rho(p) = \frac{1}{Z} 2^{-D|p|}, \qquad \text{where } Z = \sum_{p \in \text{Dom(UTM)}} 2^{-D|p|} \tag{23}$$

This is the statistical-physics definition of a halting probability. Here, it is $\rho(s)$ (and NOT Z) that is the halting probability. We note that it is not necessarily all choices of D which causes Z to be non-computable (for instance if $D = 0$ then Z is very much so computable; it is in fact infinite). To recover Ω, the Halting probability[4] of computer science, we would pose the Lagrange multiplier D to 1, then take the encoding of the program to be prefix-free and therefore, via the Kraft-inequality, Z becomes itself a probability measure:

$$\Omega = \sum_{p \in \text{Dom(UTM)}} 2^{-|p|} \tag{24}$$

We further note the work of Tadaki[7] which identifies an 'algorithmic-thermodynamics[8]' definition of Ω by adding D called a 'decompression-term' as follows:

$$\sum_{p \in \text{Dom(UTM)}} = 2^{-D|p|} \tag{25}$$

However, in each of these cases, with the exception of [8], the connection to entropy is lost because the expression of Z is reduced such that it, rather that ρ, acquires the role of the probability measure. So what do we gain by retaining the connection to an entropy maximum? The answer is in information theory. Knowing a message from a set of possible messages according to a probability measure that maximizes the entropy, makes knowing said message maximally informative. Likewise, in the case of the statistical physics version of the Halting probability, the probability measure that maximizes the entropy for this system makes our knowledge of a program that halts maximally informative.

3.1.2 Quantum computing

Let us now investigate how a quantum computation works (with a focus on special, but rarely discussed features: Most text would focus on the superposition or entanglement features — and although these are neat features, there is another much more important one for our goal). For a quantum computation, one starts with a state vector:

$$|\psi_a\rangle = \begin{pmatrix} 0 \\ \vdots \\ n \end{pmatrix} \tag{26}$$

Which evolves unitarily to a final state:

$$|\psi_b\rangle = U_0 U_1 \dots U_m \, |\psi_a\rangle \tag{27}$$

Clever use of the unitary transformations, often arranged as simple 'gates', allows one to execute a program. The input to the program is the state $|\psi_a\rangle$ and the output is the state $|\psi_b\rangle$. One would note that, so defined and if the sequence of unitary transformation is finite, such a program must always halt, and thus its complexity must be bounded. One can however get out of this predicament by taking the final state $|\psi_b\rangle$ to instead be an intermediary state, and then to throw more gates at the state in order continue with a computation:

$$\text{step 1} \qquad |\psi_b\rangle = U_0 U_1 \dots U_p \, |\psi_a\rangle \tag{28}$$

$$\text{step 2} \qquad |\psi_c\rangle = U'_0 U'_1 \dots U'_q \, |\psi_b\rangle \tag{29}$$

$$\vdots \tag{30}$$

$$\text{step k} \qquad |\psi_{k'}\rangle = U'_0 U'_1 \dots U'_v \, |\psi_k\rangle \tag{31}$$

$$\vdots \tag{32}$$

For a quantum computation to simulate a universal Turing machine it must be able to add more steps until a halting state is reached (or continue to add

steps indefinitely if the program never halts). Consequently, we note a feature of quantum computing that makes it quite special: all intermediary steps of the computation are computations in and of themselves. Programs in quantum computations are sequences of halting programs applied end-to-end. For instance, $|\psi_a\rangle \to |\psi_b\rangle$ and $|\psi_b\rangle \to |\psi_c\rangle$ are both a computation that has halted, but so is $|\psi_a\rangle \to |\psi_c\rangle$.

Let us attempt the same, but using the usual concepts of theoretical computer science. One can chain the outputs of a halting program $\mathrm{TM}_1(r_0) = r_1$ to another program $\mathrm{TM}_2(r_1) = r_2$ and so on — then a program that takes r_0 as input and produce r_2 as output exists. What's the difference between this and the chaining of quantum computing? The implementation of the TM's inner working is opaque in theoretical computer science, but in the quantum mechanical case, it is guaranteed that one can subdivide the computing steps at the per gate or unitary-transformation level (in some cases including the infinitesimal level). Quantum computing machines are a special design of a Turing machine that need not reference inner states that are not themselves pure states of computation. Compare this to a Turing machine comprised of auxiliary concepts such as a head and a tape... neither of these items are programs themselves.

3.1.3 Manifest-based computing... ?

We now activate the first special feature of our definitions. Notice that we have defined a state of affairs of reality as a manifest comprised of a set of experiments; that is, a set of pairs (TM, p) that halt. Why did we not include any programs that don't halt? This requirement will necessitate that any state of affairs of reality be in a completed state of computation. This imposes a restriction upon the construction of any Turing machine (the workings of its inner state) supported by these definitions. One will not be able to embed a universal Turing machine within a manifest, unless one chains future transformations of manifests until some halting state is reached... notice the similarity of this feature to that of quantum computation which also chains its computations? My definition recovers a generalized and abstract realization of this feature in a manner that releases it from its implementation. Any path by an observer in experimental space is guaranteed to only encounter steps that are formulated as completed computations.

3.2 Derivation

3.2.1 General linear computing ensemble

Let us start with a sum of programs (i.e. manifests that are comprised of a single element). A probability measure would assign a real number to each program of the sum. Now, recall that a manifest is a tuple of experiments. Consequently, it follows that the assignment of said real number to each element of said tuples transforms experimental space into a **vector space**, where the programs comprise the basis elements of the space and the probabilities comprise the values of the elements of its vectors.

Finally, recall another feature of quantum computation is that the inputs can be a combination of multiple programs (via the tensor product of states). This feature finds its way into our definitions in that any combination of manifests is also a manifest and that each experiment individually forms a manifest. Thus, the **tensor product** of elements of said vector space takes the product of the probabilities and expands the basis elements, previously comprised of a single program each, to manifests of multiple programs.

Consequently, it follows that the probability measure which constrains $\mathcal{O}$ to remain in experimental space, such that the probability measure is able to add, remove or combine manifests invariantly with respect both to single programs to their tensor products, is the set of a linear transformations on this space.

Let us therefore maximize the entropy:

$$S = -\sum_{s\in\mathbb{S}} \rho(s) \ln \rho(s) \tag{33}$$

subject to these constraints:

$$\sum_{s\in\mathbb{S}} \rho(s) = 1 \tag{34}$$

$$\sum_{s\in\mathbb{S}} \rho(s) \operatorname{tr} \mathbf{M}(s) = \operatorname{tr} \overline{\mathbf{M}} \tag{35}$$

where $\mathbf{M}(q)$ is a matrix-valued map[2] from $\mathbb{S}$ to $\mathbb{F}^{n\times n}$, where $\overline{\mathbf{M}}$ is a element-by-element average matrix in $\mathbb{F}^{n\times n}$ and where $\mathbb{F}$ is a field. We use the Lagrange multiplier method to derive the expression for ρ that maximizes the entropy, subject to the above mentioned constraints. Maximizing the following equation with respect to ρ yields the answer:

$$\mathcal{L} = -k_B \sum_{s\in\mathbb{S}} \rho(s) \ln(s) + \alpha \left(1 - \sum_{s\in\mathbb{S}} \rho(s)\right) + \tau \left(\operatorname{tr} \overline{\mathbf{M}} - \sum_{s\in\mathbb{S}} \rho(s) \operatorname{tr} \mathbf{M}(s)\right) \tag{36}$$

where α and τ are the Lagrange multipliers. The explicit derivation is made available in Annex B. The result of the maximization process is:

[2]There is a possibility of greater generality by considering maps $\mathbb{S} \to \mathbb{F}^{n\times m}$. In quantum mechanics this is part of the subject matter of 'quantum operations' which includes quantum channels. This mapping from differently-size vector spaces would be required in the general case to account for all possible paths of the observer in experimental space, and would likely come out as a general linear equivalent to a quantum channel transmitting information between manifests as their sizes change. This is likely interesting, but, as we will see, we will not be running out of applications for the general linear ensemble as it is, and thus we have elected to limit the scope to maps in $\mathbb{F}^{n\times n}$ in line with the typical formulations of quantum mechanics.

$$\rho(q) = \frac{1}{Z} \det \exp -\tau \mathbf{M}(q) \tag{37}$$

where

$$Z = \sum_{q \in \mathbb{S}} \det \exp -\tau \mathbf{M}(q) \tag{38}$$

As we will see, this probability measure supports the largest possible group of **linear** probability amplitudes, and thus offers the highest possible flexibility to describe physical systems consistently with the tenets of quantum mechanics. This probability measure not only includes all of quantum mechanics, but also extends it from the complex group to the general linear group. It is therefore able support any geometry of nature, including general relativity, the standard model, as well as a combination of the two. It further embeds the notion of a 'geometric collapse' of the wave-function and, finally, settles the interpretation of quantum mechanics to that of the ensemble interpretation[9].

3.2.2 Prior

No good probability measure is complete without a prior. The prior, which accounts for an arbitrary preparation of the ensemble, ought to be —for purposes of preserving the scope of the theory— of the same kind as the elements of the probability measure. Let us thus introduce the prior as the map $\mathbf{M}_p : \mathbb{S} \to \mathbb{F}^{n \times n}$ and inject it into the probability measure as well as into the partition function:

$$\rho(s) = \frac{1}{Z} \det \exp \big(\mathbf{M}_p(s)\big) \det \exp \big(-\tau \mathbf{M}(s)\big) \tag{39}$$

where

$$Z = \sum_{s \in \mathbb{S}} \det \exp \big(\mathbf{M}_p(s)\big) \det \exp \big(-\tau \mathbf{M}(s)\big) \tag{40}$$

3.2.3 Representation

We will be well-served (naturally, in terms of clutter-reduction and intuition-improvements but also because it will provide a clear sequence of simplifications from general linear amplitude to complex amplitude as well as intermediaries), to represent $\mathbf{M}_p(s)$, $\mathbf{M}(s)$ and $\overline{\mathbf{M}}$ with geometric algebra (GA). Let us now introduce the desired GA-representation for 2×2 and 4×4 matrices. We can write a general multi-vector of $\mathbb{G}(2, \mathbb{R})$ as follows:

$$\mathbf{m}_2 = a + \mathbf{x} + \mathbf{b} \tag{41}$$

where a is a scalar, $\mathbf{x}$ is a vector and $\mathbf{b}$ is a pseudo-scalar. And we can write a general multi-vector of $\mathbb{G}(4,\mathbb{R})$ as follows:

$$\mathbf{m}_4 = a + \mathbf{x} + \mathbf{f} + \mathbf{v} + \mathbf{b} \tag{42}$$

where a is a scalar, $\mathbf{x}$ is a vector, $\mathbf{f}$ is a bivector, $\mathbf{b}$ is a pseudo-vector and $\mathbf{b}$ is a pseudo-scalar. Each of these constructions admit a structure-preserving (addition/multiplication) matrix representation. Explicitly, the multi-vectors of $\mathbb{G}(2,\mathbb{R})$ are represented as follows:

$$A + X\hat{\mathbf{x}} + Y\hat{\mathbf{y}} + B\hat{\mathbf{x}} \wedge \hat{\mathbf{y}} \cong \begin{pmatrix} A+X & -B+Y \\ B+Y & A-X \end{pmatrix} \tag{43}$$

and those of $\mathbb{G}(4,\mathbb{R})$ are represented as follows:

$$\begin{aligned} &A + T\gamma_0 + X\gamma_1 + Y\gamma_2 + Z\gamma_3 \\ &\quad + F_{01}\gamma_0 \wedge \gamma_1 + F_{02}\gamma_0 \wedge \gamma_2 + F_{03}\gamma_0 \wedge \gamma_3 + F_{23}\gamma_2 \wedge \gamma_3 + F_{13}\gamma_1 \wedge \gamma_3 + F_{12}\gamma_1 \wedge \gamma_2 \\ &\quad + V_t\gamma_1 \wedge \gamma_2 \wedge \gamma_3 + V_x\gamma_0 \wedge \gamma_2 \wedge \gamma_3 + V_y\gamma_0 \wedge \gamma_1 \wedge \gamma_3 + V_z\gamma_0 \wedge \gamma_1 \wedge \gamma_2 \\ &\quad + B\gamma_0 \wedge \gamma_1 \wedge \gamma_2 \wedge \gamma_3 \\ &\qquad \cong \begin{pmatrix} A + X_0 - iF_{12} - iV_3 & F_{13} - iF_{23} + V_2 - iV_1 & -iB + X_3 + F_{03} - iV_0 & X_1 - iX_2 + F_{01} - iF_{02} \\ -F_{13} - iF_{23} - V_2 - iV_1 & A + X_0 + iF_{12} + iV_3 & X_1 + iX_2 + F_{01} + iF_{02} & -iB - X_3 - F_{03} - iV_0 \\ -iB - X_3 + F_{03} + iV_0 & -X_1 + iX_2 + F_{01} - iF_{02} & A - X_0 - iF_{12} + iV_3 & F_{13} - iF_{23} - V_2 + iV_1 \\ -X_1 - iX_2 + F_{01} + iF_{02} & -iB + X_3 - F_{03} + iV_0 & -F_{13} - iF_{23} + V_2 + iV_1 & A - X_0 + iF_{12} - iV_3 \end{pmatrix} \end{aligned} \tag{44}$$

Finally, I note that one can use curvilinear coordinates on said multi-vectors by using a basis defined as follows:

$$\mathbf{e}_0 := t_0\gamma_0 + x_0\gamma_1 + y_0\gamma_2 + z_0\gamma_3 \tag{45}$$
$$\mathbf{e}_1 := t_1\gamma_0 + x_1\gamma_1 + y_1\gamma_2 + z_1\gamma_3 \tag{46}$$
$$\mathbf{e}_2 := t_2\gamma_0 + x_2\gamma_1 + y_2\gamma_2 + z_2\gamma_3 \tag{47}$$
$$\mathbf{e}_3 := t_3\gamma_0 + x_3\gamma_1 + y_3\gamma_2 + z_3\gamma_3 \tag{48}$$

It is now interesting to note that the determinant of the matrix representation of those multi-vectors always produces a real number, even if the matrices themselves may contain complex entries. The determinant can thus be defined as the norm of said multi-vector. Said determinant can also be defined solely using constructs of geometric algebra[10]. We define the Clifford conjugate as:

$$\mathbf{m}^\ddagger := \langle \mathbf{m} \rangle_0 - \langle \mathbf{m} \rangle_1 - \langle \mathbf{m} \rangle_2 + \langle \mathbf{m} \rangle_3 + \langle \mathbf{m} \rangle_4 \tag{49}$$

and $\lfloor \mathbf{m} \rfloor_{\{3,4\}}$ as the blade-conjugate of degree 3 and 4 (flipping the plus sign to a minus sign for blade 3 and blade 4):

$$\lfloor \mathbf{m} \rfloor_{\{3,4\}} := \langle \mathbf{m} \rangle_0 + \langle \mathbf{m} \rangle_1 + \langle \mathbf{m} \rangle_2 - \langle \mathbf{m} \rangle_3 - \langle \mathbf{m} \rangle_4 \tag{50}$$

The determinant of $\mathbb{G}(2, \mathbb{R})$ is:

$$\begin{array}{rccc} \det & : & \mathbb{G}(2, \mathbb{F}) & \longrightarrow & \mathbb{R} \\ & & \mathbf{m} & \longmapsto & \mathbf{m}^\ddagger \mathbf{m} \end{array} \tag{51}$$

whereas the determinant of $\mathbb{G}(4, \mathbb{R})$ is:

$$\begin{array}{rccc} \det & : & \mathbb{G}(4, \mathbb{F}) & \longrightarrow & \mathbb{R} \\ & & \mathbf{m} & \longmapsto & \lfloor \mathbf{m}^\ddagger \mathbf{m} \rfloor_{\{3,4\}} \mathbf{m}^\ddagger \mathbf{m} \end{array} \tag{52}$$

Using GA, the probability measure and its companion partition function are thus as follows:

$$\rho(s) = \frac{1}{Z} \det \exp \big(\mathbf{m}_p(s) \big) \det \exp \big(-\tau \mathbf{m}(s) \big) \tag{53}$$

where

$$Z = \sum_{s \in \mathbb{S}} \det \exp \big(\mathbf{m}_p(s) \big) \det \exp \big(-\tau \mathbf{m}(s) \big) \tag{54}$$

3.3 Geometric Ensemble

Ensemble in the sense that an element of the sample space is randomly picked upon a measurement, and geometric in the sense that the probability measure is invariant with respect to a group of geometric transformations determined by the choice of constraints on the entropy, up to the general linear group.

For all geometric ensembles the opportunity is available to 'split' the probability measure into a first step, where the sample space is re-defined as a vector space admitting a linear evolution operator on a probability amplitude, and a final step, known as a 'probability rule' which connects the probability amplitude to a probability via the determinant. Specifically, one can write the probability amplitudes as vectors:

$$\psi(s) = \exp\big(\mathbf{m}_p(s) \big) \exp\big(-\tau \mathbf{m}(s) \big) \tag{55}$$

Then one connects this vector to the probability measure via a 'probability rule':

$$\rho(s) = \frac{1}{Z} \det \psi(s) \tag{56}$$

This split is of course a 'trick', in the sense that the frequency of all observations in the physical world are obtained with respect to the probability measure, not the amplitude, but, nonetheless, it provides the convenience of linear operations to transform the vector space representing the sample space one level down.

3.3.1 Manifests are tensor products of states

We can now create a sum over all over experimental space (finite-sized tuples). A tensor product of wave-functions $\psi_1 \otimes \psi_2$ induces a replacement of the domain of the sum from $\mathbb{S} \to \mathbb{S} \times \mathbb{S}$. And a sequence of tensor products $\psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_n$ produces a sum over $\mathbb{S}^n$.

3.3.2 General linear wave-function

The set of all complex $n \times n$ matrices maps, via the exponential, to the general linear group in $\mathbb{C}$:

$$\exp : \mathbb{M}(n, \mathbb{C}) \to \mathrm{GL}(n, \mathbb{C}) \tag{57}$$

The map is also possible for the reals, but in this case the general linear group is reduced to the orientation-preserving general linear group, because the left-hand side of the map cannot produce a matrix with a negative determinant and thus is not surjective in the general case:

$$\exp : \mathbb{M}(n, \mathbb{R}) \to \mathrm{GL}^+(n, \mathbb{R}) \tag{58}$$

The entropy maximization procedure we have used produced a probability measure which embeds the exponential map over matrices, thus connects the arbitrary linear transformation of $\mathbb{M}(n, \mathbb{R})$ to the orientation-preserving linear group $\mathrm{GL}^+(n, \mathbb{R})$. This relationship also holds for geometric algebras (of even dimensions), because their multi-vectors map to the set of matrices with real determinant (even if such matrices may be a subset of the complex matrices), and thus admit a group isomorphism with $\mathrm{GL}^+(n, \mathbb{R})$:

$$\exp : \mathbb{G}(2n, \mathbb{R}) \cong \mathrm{GL}^+(2n, \mathbb{R}) \tag{59}$$

The group of the transformations of wave-function of GP is thus the general linear group, reduced to orientation-preserving transformations (positive determinant). Explicitly, and using geometric algebra as the representation, the general linear probability measure in four dimensions is given as:

$$\rho(s) = \frac{1}{Z} \det \exp\big(a_p + \mathbf{x}_p + \mathbf{f}_p + \mathbf{v}_p + \mathbf{b}_p\big) \exp -\tau(a + \mathbf{x} + \mathbf{f} + \mathbf{v} + \mathbf{b}) \tag{60}$$

and the wave-function is:

$$\psi(s) = \exp\big(a_p + \mathbf{x}_p + \mathbf{f}_p + \mathbf{v}_p + \mathbf{b}_p\big) \exp -\tau(a + \mathbf{x} + \mathbf{f} + \mathbf{v} + \mathbf{b}) \tag{61}$$

If we consider that $\psi(s)$ is a general linear amplitude associated to an element of the state vector Ψ of H, then we can write ρ as a map from the vector Ψ to $\mathbb{R}$:

$$\begin{array}{rccl} \rho & : & H \times \Psi & \longrightarrow & \mathbb{R} \\ & & (\Psi, s) & \longmapsto & \big(\det \Psi(s)\big) \,/\, \Big(\sum_{\varphi \in \Psi} \det \varphi\Big) \end{array} \tag{62}$$

Then, if we multiply Ψ by a 'global' general linear transformation G to each element of Ψ, then clearly $\rho(G \odot \Psi, s) = \rho(\Psi, s)$ since $\det G\Psi = \det G \det \Psi$, thereby cancelling $\det G$ as it is both in the numerator and in the denominator:

$$\rho(G \odot \Psi, s) = (\det G\Psi(s))/ \left(\sum_{\varphi \in \Psi} \det G\varphi \right) = (\det \Psi(s))/ \left(\sum_{\varphi \in \Psi} \det \varphi \right) \tag{63}$$

where $\odot$ is the element-wise application of G to the elements of Ψ. The general linear probability measure is invariant with respect to a 'global' general linear transformation $(G \odot \Psi)$ on the probability amplitude.

Finally, we note that the Lagrange multiplier of the evolution part of the probability measure generates the group of said evolution part. The general relation $\exp : \mathfrak{g} \to G$ (where $\mathfrak{g}$ is the algebra of the group G) is, in the probability measure, of this form:

$$\tau \to \exp(\tau \mathbf{M}) \in \mathrm{GL}^+, \tau \in \mathbb{R} \tag{64}$$

where τ is a real number.

We note that using a single geometric constraint produces a probability measure whose evolution is in the form of one-parameter subgroup of G. A multi-parameter probability measure generating the full group is also possible. Consider the following constraints:

$$\sum_{s\in\Omega}\rho(s)=1 \tag{65}$$

$$\sum_{s\in\Omega}\rho(s)\,\mathrm{tr}\begin{pmatrix}a(s) & 0\\ 0 & 0\end{pmatrix}=\mathrm{tr}\begin{pmatrix}\overline{a} & 0\\ 0 & 0\end{pmatrix} \tag{66}$$

$$\sum_{s\in\Omega}\rho(s)\,\mathrm{tr}\begin{pmatrix}0 & b(s)\\ 0 & 0\end{pmatrix}=\mathrm{tr}\begin{pmatrix}0 & \overline{b}\\ 0 & 0\end{pmatrix} \tag{67}$$

$$\sum_{s\in\Omega}\rho(s)\,\mathrm{tr}\begin{pmatrix}0 & 0\\ c(s) & 0\end{pmatrix}=\mathrm{tr}\begin{pmatrix}0 & 0\\ \overline{c} & 0\end{pmatrix} \tag{68}$$

$$\sum_{s\in\Omega}\rho(s)\,\mathrm{tr}\begin{pmatrix}0 & 0\\ 0 & d(s)\end{pmatrix}=\mathrm{tr}\begin{pmatrix}0 & 0\\ 0 & \overline{d}\end{pmatrix} \tag{69}$$

Then the probability measure which maximizes the entropy subject to these constraints would be:

$$\rho(s)=\frac{1}{Z}\det\exp\left(-\tau_{00}\begin{pmatrix}a(s) & 0\\ 0 & 0\end{pmatrix}-\tau_{01}\begin{pmatrix}0 & b(s)\\ 0 & 0\end{pmatrix}-\tau_{10}\begin{pmatrix}0 & 0\\ c(s) & 0\end{pmatrix}-\tau_{11}\begin{pmatrix}0 & 0\\ 0 & d(s)\end{pmatrix}\right) \tag{70}$$

$$=\frac{1}{Z}\det\exp\left(\boldsymbol{\tau}\odot\mathbf{M}(s)\right) \tag{71}$$

where $\odot$ is the element-wise matrix multiplication (or Hadamard product). Here, we have obtained the lie algebra $\mathfrak{gl}(2,\mathbb{R})$ including the 4 free parameters required to generate the full group $\mathrm{GL}(2,\mathbb{R})$ via the exponential map.

3.4 Geometric Hilbert space

Let $H(m,\mathrm{G}(n,\mathbb{F}))$ be a vector space of m dimensions over $\mathrm{G}(n,\mathbb{F})$, and let $\mathrm{G}(n,\mathbb{F})$ be a n dimensional group over a field $\mathbb{F}$, representable as $n\times n$ matrices.

3.4.1 H is a vector space

To prove that H is a vector space, we need to prove the following axioms.

1. Associativity of addition
2. Commutativity of addition
3. Identity element of addition
4. Inverse elements of addition
5. Compatibility of scalar multiplication with field multiplication

6. Identity element of scalar multiplication

7. Distributivity of scalar multiplication with respect to vector addition

8. Distributivity of scalar multiplication with respect to field addition

All 8 axioms follow from the fact that tuples of matrices are vector spaces in addition and scalar multiplication, which itself follows from the fact that $n \times n$ matrices form a vector space.

3.4.2 H is a Hilbert space

We need to show that H admits an inner product, then show that H is complete. For any vector $\mathbf{v} \in H$, such as $\mathbf{v} := (g_1, g_2, \ldots, g_n)$ we define a norm as a sum over of the determinant of each of its elements:

$$\|\mathbf{v}\|^2 := \sum_{g \in \mathbf{v}} \det g \tag{72}$$

Then the inner product is given by the polarization identity:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{4} \left(\|\mathbf{u} + \mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2 \right) \tag{73}$$

Finally, in the case where n is finite, said vector space is necessarily complete. Since we have a vector space that contains an inner product and is complete, it then follows that it is a Hilbert space.

3.4.3 Inner product — Example in $\mathbb{G}(2, \mathbb{R})$

As an example, let us calculate the inner product of two multi-vectors of $\mathbb{G}(2, \mathbb{R})$. The norm of $H(n, \mathbb{G}(2, \mathbb{R}))$ is given as:

$$\|\mathbf{v}\|^2 := \mathbf{v}^\ddagger \mathbf{v} \tag{74}$$

where we define $\mathbf{v}^\ddagger$ as the Clifford transpose of a vector $\mathbf{v}$ as:

$$\begin{pmatrix} g_1 \\ \vdots \\ g_n \end{pmatrix}^\ddagger := \begin{pmatrix} g_1^\ddagger & \ldots & g_n^\ddagger \end{pmatrix} \tag{75}$$

Therefore, its inner product is:

$$
\begin{aligned}
4\langle \mathbf{u}, \mathbf{v} \rangle :&= (\mathbf{u}+\mathbf{v})^{\ddagger}(\mathbf{u}+\mathbf{v}) - (\mathbf{u}-\mathbf{v})^{\ddagger}(\mathbf{u}-\mathbf{v}) && (76)\\
&= \mathbf{u}^{\ddagger}\mathbf{u} + \mathbf{u}^{\ddagger}\mathbf{v} + \mathbf{v}^{\ddagger}\mathbf{u} + \mathbf{v}^{\ddagger}\mathbf{v} - (\mathbf{u}^{\ddagger}\mathbf{u} - \mathbf{u}^{\ddagger}\mathbf{v} - \mathbf{v}^{\ddagger}\mathbf{u} + \mathbf{v}^{\ddagger}\mathbf{v}) && (77)\\
&= \mathbf{u}^{\ddagger}\mathbf{u} + \mathbf{u}^{\ddagger}\mathbf{v} + \mathbf{v}^{\ddagger}\mathbf{u} + \mathbf{v}^{\ddagger}\mathbf{v} - \mathbf{u}^{\ddagger}\mathbf{u} + \mathbf{u}^{\ddagger}\mathbf{v} + \mathbf{v}^{\ddagger}\mathbf{u} - \mathbf{v}^{\ddagger}\mathbf{v} && (78)\\
&= 2(\mathbf{u}^{\ddagger}\mathbf{v} + \mathbf{v}^{\ddagger}\mathbf{u}) && (79)\\
\implies \langle \mathbf{u}, \mathbf{v} \rangle &= \frac{1}{2}(\mathbf{u}^{\ddagger}\mathbf{v} + \mathbf{v}^{\ddagger}\mathbf{u}) && (80)\\
& && (81)
\end{aligned}
$$

The inner product of two multi-vectors of $\mathbb{G}(2, \mathbb{R})$:

$$
\begin{aligned}
\mathbf{m}_1 &= a_1 + x_1 e_1 + y_1 e_2 + b_1 e_{12} && (82)\\
\mathbf{m}_2 &= a_2 + x_2 e_1 + y_2 e_2 + b_2 e_{12} && (83)
\end{aligned}
$$

is:

$$
\implies \frac{1}{2}(\mathbf{m}_1^{\ddagger}\mathbf{m}_2 + \mathbf{m}_2^{\ddagger}\mathbf{m}_1) = a_1 a_2 - x_1 x_2 - y_1 y_2 + b_1 b_2 \tag{84}
$$

3.4.4 Observable

A linear operator A acting on the geometric Hilbert space is an observable iff it is a self-adjoint operator; defined as:

$$
\langle A\mathbf{v}, \mathbf{u} \rangle = \langle \mathbf{v}, A\mathbf{u} \rangle \tag{85}
$$

3.4.5 Observable — Example in $\mathbb{G}(2, \mathbb{R})$

Let us now introduce an observable $A = \begin{pmatrix} A_{00} & A_{01} \\ A_{10} & A_{11} \end{pmatrix}$, and two two-state vectors $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ where $u_1, u_2, v_1, v_2, A_{00}, A_{01}, A_{10}, A_{11}$ are multi-vectors of $\mathbb{G}(2, \mathbb{R})$. Let us now calculate $\langle A\mathbf{u}, \mathbf{v} \rangle$:

$$
\begin{aligned}
2\langle A\mathbf{u}, \mathbf{v} \rangle &= (A_{00}u_1 + A_{01}u_2)^{\ddagger}v_1 + v_1^{\ddagger}(A_{00}u_1 + A_{01}u_2) \\
&\quad + (A_{10}u_1 + A_{11}u_2)^{\ddagger}v_2 + v_2^{\ddagger}(A_{10}u_1 + A_{11}u_2) && (86)\\
&= u_1^{\ddagger}A_{00}^{\ddagger}v_1 + u_2^{\ddagger}A_{01}^{\ddagger}v_1 + v_1^{\ddagger}A_{00}u_1 + v_1^{\ddagger}A_{01}u_2 \\
&\quad + u_1^{\ddagger}A_{10}^{\ddagger}v_2 + u_2^{\ddagger}A_{11}^{\ddagger}v_2 + v_2^{\ddagger}A_{10}u_1 + v_2^{\ddagger}A_{11}u_2 && (87)
\end{aligned}
$$

Now, $\langle \mathbf{u}, A\mathbf{v} \rangle$:

$$
\begin{aligned}
2\langle \mathbf{u}, A\mathbf{v}\rangle &= u_1^\ddagger (A_{00}v_1 + A_{01}v_2) + (A_{00}v_1 + A_{01}v_2)^\ddagger u_1 \\
&\quad + u_2^\ddagger (A_{10}v_1 + A_{11}v_2) + (A_{10}v_1 + A_{11}v_2)^\ddagger u_1 \qquad (88)\\
&= u_1^\ddagger A_{00}v_1 + u_1^\ddagger A_{01}v_2 + v_1^\ddagger A_{00}^\ddagger u_1 + v_2^\ddagger A_{01}^\ddagger u_1 \\
&\quad + u_2^\ddagger A_{10}v_1 + u_2^\ddagger A_{11}v_2 + v_1^\ddagger A_{10}^\ddagger u_1 + v_2^\ddagger A_{11}^\ddagger u_1 \qquad (89)
\end{aligned}
$$

For $\langle A\mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{u}, A\mathbf{v}\rangle$ to be realized, it follows that these relations must hold:

$$A_{00}^\ddagger = A_{00} \tag{90}$$
$$A_{01}^\ddagger = A_{10} \tag{91}$$
$$A_{10}^\ddagger = A_{01} \tag{92}$$
$$A_{11}^\ddagger = A_{11} \tag{93}$$

Therefore, it follows that it must be the case that A must be equal to its own Clifford transpose. Thus, A is an observable iff:

$$A^\ddagger = A \tag{94}$$

which is the geometric Hilbert space equivalent of the Hermitian operator $A^H = A$ of the complex Hilbert space. All geometric Hilbert spaces have self-adjoint operators but the expression for them may differ; $A^\ddagger = A$ is only applicable for $\mathbb{G}(2, \mathbb{R})$, just like $A^H = A$ is only applicable to complex Hilbert spaces.

3.4.6 Observable — Geometric spectral theorem

Let us show how the spectral theorem applies to $A^\ddagger = A$, such that its eigenvalues are real. Consider:

$$A = \begin{pmatrix} a_{00} & a - xe_1 - ye_2 - be_{12} \\ a + xe_1 + ye_2 + be_{12} & a_{11} \end{pmatrix} \tag{95}$$

In this case, it follows that $A^\ddagger = A$:

$$A^\ddagger = \begin{pmatrix} a_{00} & a - xe_1 - ye_2 - be_{12} \\ a + xe_1 + ye_2 + be_{12} & a_{11} \end{pmatrix} \tag{96}$$

This example is the most general 2×2 matrix A such that $A^\ddagger = A$. The eigenvalues are obtained as follows:

$$0 = \det(A - \lambda I) = \det \begin{pmatrix} a_{00} - \lambda & a - xe_1 - ye_2 - be_{12} \\ a + xe_1 + ye_2 + be_{12} & a_{11} - \lambda \end{pmatrix} \tag{97}$$

implies:

$$0 = (a_{00} - \lambda)(a_{11} - \lambda) - (a - xe_1 - ye_2 - be_{12})(a + xe_1 + ye_2 + be_{12} + a_{11}) \tag{98}$$

$$0 = (a_{00} - \lambda)(a_{11} - \lambda) - (a^2 - x^2 - y^2 + b^2) \tag{99}$$

finally:

$$\lambda = \{\frac{1}{2}\left(a_{00} + a_{11} - \sqrt{(a_{00} - a_{11})^2 + 4(a^2 - x^2 - y^2 + b^2)}\right), \tag{100}$$

$$\frac{1}{2}\left(a_{00} + a_{11} + \sqrt{(a_{00} - a_{11})^2 + 4(a^2 - x^2 - y^2 + b^2)}\right)\} \tag{101}$$

We note that in the case where $a_{00} - a_{11} = 0$, the roots would be complex iff $a^2 - x^2 - y^2 + b^2 < 0$, but we already stated that the manifold must be orientable — therefore it is the case that $a^2 - x^2 - y^2 + b^2 \geq 0$, as this expression is the determinant of the multi-vector. Consequently, $A^{\ddagger} = A$ implies, for an orientable[3] geometric Hilbert space, that its roots are real-valued, and thus constitute a 'geometric' observable in the traditional sense of an observable in a Hilbert space.

4 Applications

4.1 The complex Hilbert space is a special case

The geometric Hilbert space $\mathbb{G}(2, \mathbb{R})$ reduces to the complex Hilbert space under the following elimination:

$$(A + \mathbf{X} + \mathbf{B}|_{\mathbf{X} \to 0} = A + \mathbf{B} \tag{102}$$

The observables become Hermitian operator and the evolution is unitary.

4.2 Spinor group wave-function

This representation of the general linear wave-function in geometric algebra admits multiple reductions to simpler systems. An important reduction was the one done in the previous section, from the general linear group in two dimensions to the complex, yielding the familiar quantum mechanics by annulling the

[3]We note the exception that a geometric observable may have real eigenvalues even in the case of a transformation that reverses the orientation if the elements $a_{00} - a_{11}$ are not zero and up to a certain magnitude, whereas transformations in the natural orientation are not bounded by a magnitude — thus creating an orientation-based asymmetry.

geometric components of the multi-vector. Another reduction is from the general linear group to the spinor group, yielding a geometric formulation of the relativistic wave-function, by annulling both the $\mathbf{x}$ and the $\mathbf{v}$ component of the general linear group. Let us do this reduction right now.

Posing:

$$(a + \mathbf{x} + \mathbf{f} + \mathbf{v} + \mathbf{b}|_{\mathbf{x}\to 0, \mathbf{v}\to 0} = a + \mathbf{f} + \mathbf{b} \tag{103}$$

the wave-function becomes:

$$\psi(s) = \exp\bigl(a_p + \mathbf{f}_p + \mathbf{b}_p\bigr) \exp -\tau(\mathbf{f} + \mathbf{b}) \tag{104}$$

Application of the 'probability rule' via the determinant produces:

$$\rho(s) = \frac{1}{Z} \exp 4a_p \tag{105}$$

where Ψ is the sample space (which is also a vector space).

We note that the prior of this wave-function is of the same form as that identified by David Hestenes as the geometric algebra formulation of the relativistic wave-function[11], which he defines as $\psi = \rho e^{iB} R$, where $R = e^{-\mathbf{F}/2}$ is a rotor.

4.3 The probability density of an event

So far we have considered Ω to be a discrete and finite vector space. However, it is preferable to select a parametrization in the same category as the domain of the transformations of the theory. In this case, as the transformations are general linear (and thus geometric), consequently the parametrization will also be over the geometry of the system, thereby creating a normalization condition self-contained to the scope of the theory. One selects the parametrization to be continuous and uncountable:

$$Z = \int_M \det \psi(\mathbf{x}, \tau) \sqrt{|g|}\, \mathrm{d}\mathbf{x} \tag{106}$$

where $\mathbf{x}$ is taken to be a n-vector, and where n is of the same value as that of the $n \times n$ matrix $\mathbf{M}(q)$, and where M is the domain of integration. The probability density would then be:

$$\rho(m) = \frac{\int_m \det \psi(\mathbf{x}, \tau) \sqrt{|g|}\, \mathrm{d}\mathbf{x}}{\int_M \det \psi(\mathbf{x}, \tau) \sqrt{|g|}\, \mathrm{d}\mathbf{x}} \tag{107}$$

where $m \subset M$. We note that for this parametrization, one needs a metric g (and also for M to be orientable). This gives the probability density that a geometrically-extended instruments (for instance the screen at the end of a

double-slit experiment) 'clicks' at event $(x_0, x_1, \ldots, x_n)$, and this probability density further remains invariant with respect to any geometric transformation available to the observer including arbitrary change of coordinates (general linear transformations).

4.4 The collapse

So what happens during a measurement that causes the wave-function to collapse to a point in spacetime? The wave-function collapse has been a tough pill to swallow because, before these results, we did not know that the origin of the wave-function was in entropy under geometric constraint — we, at best, believed it was a measure over a postulated unitary sample space, and any geometric properties it may have (space-time normalization/Lorentz invariance) were strapped on as a secondary set of axioms tied in to a normalization condition in space-time, then we noticed to our surprise that such geometric properties ought to be represented in the same unitary space as the rest of the theory. In this framework, the collapse is simply the result of selecting an element of the geometric sample space. All geometric features of the wave-function are part of the sample space, and thus its geometry behaves in the same manner as any other observables under measurement. From this, it is then obvious (even trivial) that the wave-function ought to 'collapse' to a single point in space-time upon a measurement of its sample space. Just like measuring a coin-toss causes the system to pick [HEAD] or [TAILS], a measurement of a geometric sample space causes the system to pick (t_0, x_1, x_2, x_3), (t'_0, x'_1, x'_2, x'_3), etc., from the sample space.

4.5 Quantum gravity (probability measure support)

Since all finite dimensional groups have matrix representations, it then follows that our framework —as it works with any matrices— is able to tackle any such groups; including, of course, those groups resulting from the direct product of groups such as the affine group: $A(n, \mathbb{R}) = T(n, \mathbb{R}) \times \mathrm{GL}(n, \mathbb{R})$. Let us now consider the metric-affine theory of gravity[12], which is obtained as the result of gauging the affine group $A(n, \mathbb{R})$. We will now create a linear geometric amplitude over said affine group and a corresponding probability rule. Let $\mathfrak{a}$ be the lie algebra of $A(n, \mathbb{R})$, and consider the following set of constraints:

$$\sum_{s \in \Omega} \rho(s) = 1 \tag{108}$$

$$\sum_{s \in \Omega} \rho(s) \operatorname{tr} \mathfrak{a}(s) = \operatorname{tr} \overline{\mathfrak{a}} \tag{109}$$

After we solve the Lagrange equation, we get:

$$\rho(s) = \frac{1}{Z} \det \exp\left(-\tau(\mathfrak{a}(s)\right) \tag{110}$$

where

$$Z = \sum_{s\in\Omega} \det \exp\left(-\tau(\mathfrak{a}(s)\right) \tag{111}$$

We note that $\exp -\tau\mathfrak{a}(s)$ is the exponential map of the algebra to the group $A(n,\mathbb{R})$. Consequently, it follows that $\rho(s)$ is invariant with respect to a global transformation of the affine group. A gauge-invariant derivative is obtained by gauging a 'local' application of the group $A \to A(s)$ and identifying a compensating field, called a gauge-field, which maintains the invariance. This produces the metric-affine theory of gravitation[12] or, if the gauge group is the Poincaré group $T(n,\mathbb{R}) \times \mathrm{SO}(n,\mathbb{R})$, produces the Einstein–Cartan gravity theory[13].

4.6 Quantum gravity + Standard model (probability measure support)

The flexibility provided by our methodology allows us to combine the gravity (for instance the metric-affine gauge theory or the Einstein–Cartan theory of gravity resulting from gauging the Poincaré group) to the standard model using a unifying probability measure specifically for this purpose, which adheres to the rules of geometric probabilities — themselves a superset of unitary QM. For instance and as a curiosity, we will create the geometric probability measure associated to this group:

$$SM := T(4) \times SO(4) \times U(1) \times SU(2) \times SU(3) \tag{112}$$

where we take $T(4) \times \mathrm{SO}(4)$ to be the Poincaré group (alternatively we could have picked the affine group here) and $U(1) \times SU(2) \times SU(3)$ to be the (gauge group of the) standard model. Let $\mathfrak{sm}$ be the lie algebra of the above group. Then, consider the following constraints:

$$\sum_{s\in\Omega} \rho(s) = 1 \tag{113}$$

$$\sum_{s\in\Omega} \rho(s) \operatorname{tr} \mathfrak{sm}(s) = \operatorname{tr} \overline{\mathfrak{sm}(s)} \tag{114}$$

Solving the Lagrange equation, we obtain:

$$\rho(s) = \frac{1}{Z} \det \exp -\tau(\mathfrak{sm}(s)) \tag{115}$$

and the wave-function as:

$$\psi(s) = \exp -\tau(\mathfrak{sm}(s)) \tag{116}$$

Here, as before, the exponential map generates the group associated with the algebra. The probability amplitude is linear within the associated geometric Hilbert space, and said amplitude connects to the probability via the determinant. Consequently, the map from $\psi(s)$ to $\rho(s)$ is invariant with respect to a global transformation of said group. Then, as per the methodology, producing a gauge-invariant derivative for this group $SM \to SM(s)$ induces a number of compensating gauge fields associated to these groups.

A 'physically-correct' representation of the standard-model-group × general-relativity-group likely depends upon the choice of lie algebra such that it correctly maps the algebra to the observed particle population, and this is a different problem than merely showing that geometric Hilbert spaces, unlike complex Hilbert spaces, are sufficiently flexible to support any choice (good or bad) consistently with the core tenets of quantum mechanics.

4.7 Gauge-theoretical completeness

The general linear amplitude is able to support any gauge theory or any gauge transformation, provided that said gauge transformation be expressible as a matrix at the lie algebra level.

4.8 Lagrangian

A typical Lagrangian density relies upon the existence of a measure of the momentum:

$$\overline{P} = \frac{1}{Z} \int_M P(\mathbf{x}) \psi(\mathbf{x})^* \psi(\mathbf{x}) \, \mathrm{d}\mathbf{x} \tag{117}$$

Now, iff there exists a $\hat{p}$ and a basis of $\psi(\mathbf{x})$ such that $P(\mathbf{x})\psi(\mathbf{x}) = \hat{p}\psi(\mathbf{x})$, then $\hat{p}$ is called the momentum operator. In relativistic quantum mechanics, $\hat{p} = \gamma^0 \hbar \not{D}$. Then one can support a Lagrangian density as a measure of the energy:

$$\mathcal{L}(\mathbf{x}) = \psi^*(\mathbf{x}) \hat{p} \psi(\mathbf{x}) - \psi^*(\mathbf{x}) \gamma_0 m c^2 \psi(\mathbf{x}) \tag{118}$$

A similar probability measure can be constructed for a geometric Hilbert space, using the determinant instead of the complex norm:

$$\overline{P} = \frac{1}{Z} \int_M P(\mathbf{x}) \det \psi(\mathbf{x}, \tau) \, \mathrm{d}\mathbf{x} \tag{119}$$

And the potential energy as:

$$\overline{V} = \frac{1}{Z} \int_M V(\mathbf{x}) \det \psi(\mathbf{x}, \tau) \, \mathrm{d}\mathbf{x} \tag{120}$$

resulting in the Lagrangian density:

$$\mathcal{L}(\mathbf{x}) = (P(\mathbf{x}) - V(\mathbf{x})) \det \psi(\mathbf{x}, \tau) \tag{121}$$

4.9 Testable Prediction

We are all familiar with the probabilities of a coin toss, whose sample space is {[HEAD], [TAILS]}. Sampling the space returns an element [HEAD] or [TAILS] with a probability of 50%/50%. But this system ignores the geometry of actually tossing a coin in the air, which is a luxury we do not necessarily have in the physical universe. If we do account for the geometry and the state of relative observers, then said probability ought to preserve the probabilities for any geometric transformations available to said observers. For instance, what if the coin is tossed in a spaceship, traveling away from Earth at near the speed of light; what if it is accelerated; what if it is tossed in the presence of a strong gravitational field? Will all observers report the same probability? They should, but for that we need a general linear invariant probability measure and a geometric Hilbert space.

We can tackle such cases with a general linear probability amplitude and a probability rule that together describe the probability density of a (coin landing) event in space-time, then additionally attribute one of two possible values to this landing event. Measuring the sample space triggers a pick of [HEAD] or [TAILS], of course, but also a pick of an event in space-time where the coin lands — thus also triggers a collapse of the wave-function. To tackle this case or other similar cases, we will consider the ideal case of a 'general linear two-state system', which one may call a 'general linear coin-toss' or perhaps even a 'general linear qubit'.

4.9.1 Geometric interference

In this section we propose a falsifiable experiment that could be used, in the lab, to confirm/infirm this theory. Specifically, for a two-state system geometric interference is possible as a new predicted form of interference. Suppose a two-state system comprised of a linear combination of ψ_1 and ψ_2:

$$\psi_1 = \exp(A_1 + \mathbf{X}_1 + \mathbf{F}_1 + \mathbf{V}_1 + B_1\mathbf{I}) \tag{122}$$
$$\psi_2 = \exp(A_2 + \mathbf{X}_2 + \mathbf{F}_2 + \mathbf{V}_2 + B_2\mathbf{I}) \tag{123}$$

Such a system can be obtained by using a transformation $\mathbf{T}$ to transform the wave-function:

$$Z = \det \psi_1 + \det \psi_2 \tag{124}$$
$$\text{transformation by } \mathbf{T} \rightarrow \det(\psi_1 + \psi_2) \tag{125}$$
$$\rightarrow (e^{A_1})^4 + (e^{A_2})^4 + \text{interference-pattern...} \tag{126}$$

To find an explicit expression for the interference pattern it is more convenient to express the determinant entirely in the language of geometric algebra:

$$\det \mathbf{m} := \lfloor \mathbf{m}^\ddagger \mathbf{m} \rfloor_{3,4} \mathbf{m}^\ddagger \mathbf{m} \tag{127}$$

We are now ready to produce the expression for the interference pattern. With straightforward algebraic manipulations, the probability addition rules for a two-state system reduce to a sum of two states $\lfloor \psi_1^\ddagger \psi_1 \rfloor_{3,4} (\psi_1^\ddagger \psi_1) = (e^{A_1})^4$ and $\lfloor \psi_2^\ddagger \psi_2 \rfloor_{3,4} (\psi_2^\ddagger \psi_2) = (e^{A_2})^4$, along with an accompanying interference pattern. We note that both the geometric conjugate and the 3,4 blade conjugate are distributive.

$$Z = \det(\psi_1 + \psi_2) \tag{128}$$

$$= \lfloor (\psi_1 + \psi_2)^\ddagger (\psi_1 + \psi_2) \rfloor_{3,4} (\psi_1 + \psi_2)^\ddagger (\psi_1 + \psi_2) \tag{129}$$

$$= \underbrace{(e^{A_1})^4 + (e^{A_2})^4}_{\text{sum of two states}} + \underbrace{\begin{array}{c} \lfloor \psi_1^\ddagger \psi_1 \rfloor_{3,4} (\psi_1^\ddagger \psi_2 + \psi_2^\ddagger \psi_1 + \psi_2^\ddagger \psi_2) + \lfloor \psi_2^\ddagger \psi_2 \rfloor_{3,4} (\psi_1^\ddagger \psi_1 + \psi_1^\ddagger \psi_2 + \psi_2^\ddagger \psi_1) \\ + \lfloor \psi_1^\ddagger \psi_2 + \psi_2^\ddagger \psi_1 \rfloor_{3,4} (\psi_1^\ddagger \psi_1 + \psi_1^\ddagger \psi_2 + \psi_2^\ddagger \psi_1 + \psi_2^\ddagger \psi_2) \end{array}}_{\text{general linear interference pattern}} \tag{130}$$

In this case the interference pattern is much more complicated than the simple cosine of the standard Born rule, but that is to be expected as it comprises the full general linear group and not just the unitary group. It accounts for the group of all geometric transformations which preserves the probability distribution ρ for a two-state general linear system.

Let us note that as elements of the algebra are nullified, then the interference pattern correspondingly reduces in complexity to eventually recover ordinary complex interference. Therefore, the general linear interference can be understood as a generalization of complex interference. Specifically, when all elements of the odd-sub-algebra are eliminated (by posing $\mathbf{X} \to 0$, $\mathbf{V} \to 0$), then the wave-function reduces to the geometric algebra form of the relativistic wave-function identified by David Hestenes, in terms of a spinor field. Of course, setting a full sector of the multi-vector to 0 is a very brute restriction — more subtle conditions can be imposed on the relations of the elements of the multivectors without necessarily posing them equal to zero, and such restrictions would entail a slightly different interference pattern.

These reductions produce a series of interference patterns of decreasing complexity, and as such they provide a method to experimentally identify which group of geometric transformations the world obeys, using interference experiments on space-time. We note that interference experiments have paid off substantial dividends in the history of physics and are usually easier to construct than alternatives. Identification of the general linear interference pattern $(A, \mathbf{X}, \mathbf{F}, \mathbf{V}, \mathbf{B})$ in a lab experiment would suggest the world obeys the metric-affine gravitational theory; whereas identification of a reduced interference pattern $(A, \mathbf{F}, \mathbf{B})$ and subsequent showing a failure to observe the full general linear interference, would suggest at most Poincaré gauge theory and so on.

In any such case, a general experimental setup would send a particle into two distinct paths. Then, either: a) one of the paths undergoes a general linear transformation, while the other doesn't or b) both paths undergo a different general linear transformation. Then, the paths are recombined to produce an interference pattern on a screen. Depending on the nature of the transformation, a deformation of the interference pattern based on the geometry of the setup should be observed.

In a possibly even better construction of such an experiment, one would exploit the non-commutativity of the general linear transformations to identify only the difference between complex-interference and general linear interference. One would apply the same general linear transformations to each path, but would reverse the order in which the transformations are applied. The resulting interference pattern would then be compared to a case where both paths are transformed in the same order. Then, complex-interference, as it is fully commutative, would predict the same interference pattern irrespective of the order the transformations are applied in — whereas, with general linear interference, as it is non-commutative, would predict different interference patterns.

To achieve this, it may be necessary to use a three-dimensional detector, whose idealized example is a homogeneous bath of impurities (allowing photons to 'click' anywhere within the volume of the detector), instead of a two-dimensional screen, since the opportunity for non-commutative behaviour often kicks in at three dimensions or higher. In a real experiment, it is probably easier to use a 2d x-y screen, and stepping it along an orthogonal z-axis, capturing the 2d interference pattern at each step, then numerically reconstructing the volumetric interference pattern out of the steps.

To my knowledge, such experiments have not been carried out to date. Finally, I note the work of B. I. Lev.[14] which suggest (theoretically) the possibility of an extended interference pattern associated with the David Hestenes form of the relativistic wave-function and for the subset of rotors.

References

[1] Edmund L Gettier. Is justified true belief knowledge? *analysis*, 23(6):121–123, 1963.

[2] Richard L Kirkham. Does the gettier problem rest on a mistake? *Mind*, 93(372):501–513, 1984.

[3] Stephen Wolfram. *A new kind of science*, volume 5. Wolfram media Champaign, IL, 2002.

[4] Gregory J. Chaitin. A theory of program size formally identical to information theory. *J. ACM*, 22(3):329–340, July 1975.

[5] Ludwig Wittgenstein. *Tractatus logico-philosophicus*. Routledge, 2013.

[6] John A Wheeler. Information, physics, quantum: The search for links. *Complexity, entropy, and the physics of information*, 8, 1990.

[7] Kohtaro Tadaki. A generalization of chaitin's halting probability omega and halting self-similar sets. *Hokkaido Math. J.*, 31(1):219–253, 02 2002.

[8] John Baez and Mike Stay. Algorithmic thermodynamics. *Mathematical. Structures in Comp. Sci.*, 22(5):771–787, September 2012.

[9] Leslie E Ballentine. The statistical interpretation of quantum mechanics. *Reviews of Modern Physics*, 42(4):358, 1970.

[10] Douglas Lundholm and Lars Svensson. Clifford algebra, geometric algebra, and applications. *arXiv preprint arXiv:0907.5356*, 2009.

[11] David Hestenes. Spacetime physics with geometric algebra. *American Journal of Physics*, 71(7):691–714, 2003.

[12] Friedrich W Hehl, J Dermott McCrea, Eckehard W Mielke, and Yuval Ne'eman. Metric-affine gauge theory of gravity: field equations, noether identities, world spinors, and breaking of dilation invariance. *Physics Reports*, 258(1-2):1–171, 1995.

[13] K Hayashi and T Shirafuji. Gravity from the poincaré gauge theory of fundamental interactions, i. general formulation, ii. equations of motion for test bodies and various limits, iii. weak field approximation, iv. mass and energy of particle spectrum. *Prog. Theor. Phys*, 64:866–883, 1980.

[14] Bohdan I Lev. Wave function as geometric entity. *Journal of Modern Physics*, 3:709–713, 2012.

A Notation

S will denote the entropy, $\mathcal{A}$ the action, L the Lagrangian, and $\mathcal{L}$ the Lagrangian density. Sets, unless a prior convention assigns it another symbol, will be written using the blackboard bold typography (ex: $\mathbb{L}, \mathbb{W}, \mathbb{Q}$, etc.). Matrices will be in bold upper case (ex: $\mathbf{A}, \mathbf{B}$), whereas vectors and multi-vectors will be in bold lower case (ex: $\mathbf{u}, \mathbf{v}, \mathbf{g}$) and most other constructions (ex.: scalars, functions) will have plain typography (ex. a, A). The identity matrix is I, the unit pseudo-scalar (of geometric algebra) is $\mathbf{I}$ and the imaginary number is i. The Dirac gamma matrices are $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ and the Pauli matrices are $\sigma_x, \sigma_y, \sigma_z$. The basis elements of an arbitrary curvilinear geometric basis will be denoted $\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ (such that $\mathbf{e}_\nu \cdot \mathbf{e}_\mu = g_{\mu\nu}$) and if they are orthonormal as $\hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_n$ (such that $\hat{\mathbf{x}}_\mu \cdot \hat{\mathbf{x}}_\nu = \eta_{\mu\nu}$). The asterisk z^* denotes the complex conjugate of z, and the dagger $\mathbf{A}^\dagger$ denotes the conjugate transpose of $\mathbf{A}$. A geometric algebra of n dimensions over a field $\mathbb{F}$ is noted as $\mathcal{G}_{n,\mathbb{F}}$. We note the matrix representation of a multi-vector $\mathbf{g}$ as $M[\mathbf{g}]$, defined as a map

$M : \mathcal{G}_{n,\mathbb{F}} \to \mathbb{M}_{n,\mathbb{F}}$ which preserves the geometric product via the matrix product, and thus benefits from group isomorphism. The grades of a multi-vector will be denoted as $\langle \mathbf{v} \rangle_k$. Specifically, $\langle \mathbf{v} \rangle_0$ is a scalar, $\langle \mathbf{v} \rangle_1$ is a vector, $\langle \mathbf{v} \rangle_2$ is a bi-vector, $\langle \mathbf{v} \rangle_{n-1}$ is a pseudo-vector and $\langle \mathbf{v} \rangle_n$ is a pseudo-scalar. Furthermore, a scalar and a vector $\langle \mathbf{v} \rangle_0 + \langle \mathbf{v} \rangle_1$ is a para-vector, and a combination of even grades ($\langle \mathbf{v} \rangle_0 + \langle \mathbf{v} \rangle_2 + \langle \mathbf{v} \rangle_4 + \dots$) or odd grades ($\langle \mathbf{v} \rangle_1 + \langle \mathbf{v} \rangle_3 + \dots$) are even-multi-vectors or odd-multi-vectors, respectively. The commutator is defined as $[\mathbf{A}, \mathbf{B}] := \mathbf{AB} - \mathbf{BA}$ and the anti-commutator as $\{\mathbf{A}, \mathbf{B}\} := \mathbf{AB} + \mathbf{BA}$. We use the symbol $\cong$ to relate two sets that are related by a group isomorphism (ex: $\mathcal{G}_{4,\mathbb{C}} \cong \mathbb{M}_{4,\mathbb{C}}$). We denote the Hadamard product, or element-wise multiplication, of two matrices using $\odot$, and is written for instance as $\mathbf{M} \odot \mathbf{P}$, and for a multivector as $\mathbf{u} \odot \mathbf{v}$; for instance: $(a_0 + x_0\hat{\mathbf{x}} + y_0\hat{\mathbf{y}} + b_0\hat{\mathbf{x}} \wedge \hat{\mathbf{y}}) \odot (a_1 + x_1\hat{\mathbf{x}} + y_1\hat{\mathbf{y}} + b_0 1\hat{\mathbf{x}} \wedge \hat{\mathbf{y}})$ would equal $a_0 a_1 + x_0 x_1 \hat{\mathbf{x}} + y_0 y_1 \hat{\mathbf{y}} + b_0 b_1 \hat{\mathbf{x}} \wedge \hat{\mathbf{y}}$.

B Lagrange equation

The Lagrangian equation to maximize is:

$$\mathcal{L}(\rho, \alpha, \tau) = -k_B \sum_{q \in \mathbb{Q}} \rho(q) \ln \rho(q) + \alpha \left(1 - \sum_{q \in \mathbb{Q}} \rho(q) \right) + \tau \operatorname{tr} \left(\overline{\mathbf{M}} - \sum_{q \in \mathbb{Q}} \rho(q) \mathbf{M}(q) \right) \tag{131}$$

where α and τ are the Lagrange multipliers. We note the usage of the trace operator for the geometric constraint such that a scalar-valued equation is maximized. Maximizing this equation for ρ by posing $\frac{\partial \mathcal{L}}{\partial \rho(p)} = 0$, where $p \in \mathbb{Q}$, we obtain:

$$\begin{aligned}
\tfrac{\partial \mathcal{L}}{\partial \rho(p)} &= -k_B \ln \rho(p) - k_B - \alpha - \tau \operatorname{tr} \mathbf{M}(p) && (132)\\
0 &= k_B \ln \rho(p) + k_B + \alpha + \tau \operatorname{tr} \mathbf{M}(p) && (133)\\
\implies \ln \rho(p) &= \frac{1}{k_B} \left(-k_B - \alpha - \tau \operatorname{tr} \mathbf{M}(p) \right) && (134)\\
\implies \rho(p) &= \exp\left(\frac{-k_B - \alpha}{k_B} \right) \exp\left(-\frac{\tau}{k_B} \operatorname{tr} \mathbf{M}(p) \right) && (135)\\
&= \frac{1}{Z} \det \exp\left(-\frac{\tau}{k_B} \mathbf{M}(p) \right) && (136)
\end{aligned}$$

where Z is obtained as follows:

$$1 = \sum_{q\in\mathbb{Q}} \exp\left(\frac{-k_B-\alpha}{k_B}\right)\exp\left(-\frac{\tau}{k_B}\operatorname{tr}\mathbf{M}(q)\right) \quad (137)$$

$$\implies \left(\exp\left(\frac{-k_B-\alpha}{k_B}\right)\right)^{-1} = \sum_{q\in\mathbb{Q}} \exp\left(-\frac{\tau}{k_B}\operatorname{tr}\mathbf{M}(q)\right) \quad (138)$$

$$Z := \sum_{q\in\mathbb{Q}} \det\exp\left(-\frac{\tau}{k_B}\mathbf{M}(q)\right) \quad (139)$$

We note that the Trace in the exponential drops down to a determinant, via the relation $\det\exp A \equiv \exp\operatorname{tr} A$.

B.1 Multiple constraints

Consider a set of constraints:

$$\overline{\mathbf{M}}_1 = \sum_{q\in\mathbb{Q}} \rho(q)\mathbf{M}_1(q) \quad (140)$$

$$\vdots \quad (141)$$

$$\overline{\mathbf{M}}_n = \sum_{q\in\mathbb{Q}} \rho(q)\mathbf{M}_n(q) \quad (142)$$

Then the Lagrange equation becomes:

$$\mathcal{L} = -k_B\sum_{q\in\mathbb{Q}}\rho(q)\ln\rho(q) + \alpha\left(1-\sum_{q\in\mathbb{Q}}\rho(q)\right) + \tau_1\operatorname{tr}\left(\overline{\mathbf{M}}_1 - \sum_{q\in\mathbb{Q}}\rho(q)\mathbf{M}_1(q)\right) + \ldots + \tau_n\operatorname{tr}\left(\overline{\mathbf{M}}_n - \sum_{q\in\mathbb{Q}}\rho(q)\mathbf{M}_n(q)\right) \quad (143)$$

and the measure references all n constraints:

$$\rho(q) = \frac{1}{Z}\det\exp\left(-\frac{\tau_1}{k_B}\mathbf{M}_1(q) - \cdots - \frac{\tau_n}{k_B}\mathbf{M}_n(q)\right) \quad (144)$$

B.2 Multiple constraints - general case

In the general case of a multi-constraint system, each entry of the matrix corresponds to a constraint:

$$\overline{M}_{00}\begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} = \sum_{q\in\mathbb{Q}} \rho(q) M_{00}(q) \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} \tag{145}$$

$$\vdots \tag{146}$$

$$\overline{M}_{01}\begin{pmatrix} 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} = \sum_{q\in\mathbb{Q}} \rho(q) M_{01}(q) \begin{pmatrix} 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \tag{147}$$

$$\vdots \tag{148}$$

$$\overline{M}_{nn}\begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} = \sum_{q\in\mathbb{Q}} \rho(q) M_{nn}(q) \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} \tag{149}$$

For a $n \times n$ matrix, there are n^2 constraints.

The probability measure which maximizes the entropy is as follows:

$$\rho(q) = \frac{1}{Z} \det \exp\left(-\frac{1}{k_B}\boldsymbol{\tau} \odot \mathbf{M}(q)\right) \tag{150}$$

where $\boldsymbol{\tau}$ is a matrix of Lagrange multipliers, and $\odot$, the element-wise multiplication, assigns the corresponding Lagrange multiplier to each constraint.

C Quackery (intermission)

Definition 25 (Quack)**.** *A function of p which does NOT recursively enumerate the empirical evidence of p is a quack function for p.*

We do note that it is the case that for some specific functions one can prove there are scientific methods, such as the dovetail scheduler, or are quacks, such as a function that works on a single non-halting program indefinitely. As an example, consider one who attempts to square the circle using only compass and straightedge by continuously trying new permutations of such. Since this person will be trying forever, then a function, which models this behaviour, constitutes a quack function.

Theorem 6 (The Fundamental Theorem of Quackery)**.** *In the general case, one cannot prove if an arbitrary function of p is a quack or a scientific method. The proof follows directly from the halting problem.*