The Design of a Formal System of Science able to prove the Laws of Physics from Unhypothetical First Principles

Alexandre Harvey-Tremblay[1]

[1]Independent scientist, aht@protonmail.ch

October 17, 2021

Abstract

A method to derive the laws of physics from unhypothetical first principles is reported. The first part of the proof consists of constructing a formal system of science which is universal in the computer-theoretical sense, and the second part consists of using the model to derive the laws of physics as an exact solution to a maximization problem. Specifically, the laws of physics are derived in the form of an information-theoretical model maximizing the axiomatic information produced by the scientific method. To construct the formal system, modern notions relating to mathematical undecidability are leveraged to create a 'trial and error' foundation to the discovery of mathematical knowledge, such that one is required to run programs to completion —essentially to perform 'mathematical experiments'— to acquire said knowledge, thereby permitting a re-formulation of mathematics conductive to experimental methods. The laws of physics are then derived as the probability measure that maximizes the entropy of a path in the space of all possible arrangements of experiments, leading to a fundamental definition of the observer effortlessly supporting what appears to the author as a theory of quantum gravity. Finally, applications of the system to fundamental open problems of physics as well as testable predictions are proposed.

Contents

1 Introduction **3**
- 1.1 Atomic Facts 4
- 1.2 Knowledge-Bases > Formal Axiomatic Systems 5
  - 1.2.1 Connection to Formal Axiomatic Systems 7
  - 1.2.2 Axiomatic Information 8
- 1.3 Discussion — The Mathematics of Knowledge 8

2 The Formal System of Science **10**
2.1 Axiomatic Foundation of Science 10
2.1.1 Classification of Scientific Theories 13
2.1.2 The Fundamental Theorem of Science 14
2.2 Axiomatic Foundation of Physics 14

3 Main Result **15**
3.1 Prior 17
3.2 Overview 17
3.2.1 Quantum Computing (Recall) 17
3.2.2 Sketch 18
3.2.3 Matrix-Valued Vector and Transformations 19
3.2.4 Sums of Paths 20

4 Foundation of Physics (Standard QM) **21**
4.1 Born Rule 21
4.2 Axioms of QM 22
4.2.1 Dirac–von Neumann Axioms (State vector) 22
4.2.2 Dirac–von Neumann Axioms (Expectation Value) 23
4.2.3 Time-Evolution (Schrödinger equation) 23
4.3 Discussion — Wave-function Collapse, Interpretation of QM . . . 24

5 Foundation of Physics (Quantum Gravity) **25**
5.1 Algebra of Natural States, in 2D 25
5.1.1 Geometric Representation of 2×2 matrices 26
5.1.2 Axiomatic Definition of the Algebra, in 2D 27
5.1.3 Reduction to Complex Hilbert Spaces 28
5.1.4 Observable, in 2D — Self-Adjoint Operator 29
5.1.5 Observable, in 2D — Eigenvalues / Spectral Theorem . . 30
5.2 Algebra of Natural States, in 3D (brief) 31
5.2.1 Axiomatic Definition of the Algebra, in 3D 31
5.2.2 Reduction to Complex Hilbert Spaces 32
5.3 Algebra of Natural States, in 4D 32
5.3.1 Geometric Representation (in 4D) 33
5.3.2 Axiomatic Definition of the Algebra, in 4D 34
5.3.3 Reduction to Complex Hilbert Space 35
5.4 Law of Motion 35
5.4.1 Probability-Preserving Evolution (Left Action in 2D) . . . 35
5.4.2 Probability-Preserving Evolution (Adjoint Action in 2D) . 37
5.4.3 General Linear Schrödinger Equation (Left Action) 38
5.4.4 General Linear Schrödinger Equation (Adjoint Action) . . 38
5.4.5 Conservation of Probability (Left Action in 2D) 39
5.4.6 Conservation of Probability (Adjoint Action in 2D) 40
5.5 Gravity as a Gauge Theory 41
5.5.1 Unitary Gauge (Recap) 41
5.5.2 General Linear Gauge 42

6 Testable Prediction **42**
6.1 Geometric Interference 42
6.1.1 Geometric Algebra Dot Product 43
6.1.2 Geometric Interference (General Form) 44
6.1.3 Complex Interference (Recall) 44
6.1.4 Geometric Interference in 2D 45
6.1.5 Geometric Interference in 4D 45
6.1.6 Geometric Interference in 4D (Shallow Phase Rotation) . 46
6.1.7 Geometric Interference in 4D (Deep Phase Rotation) . . . 46
6.1.8 Geometric Interference in 4D (Deep Spinor Rotation) . . 47
6.1.9 Geometric Interference Experiment (Sketch) 49

7 Discussion — Existence of the Universe **50**

A Notation **51**

B Lagrange equation **52**
B.1 Multiple constraints 52
B.2 Multiple constraints - General Case 53

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 demoted to simply be 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 flaws and loopholes, 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 loophole: 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. What used to be "justified true belief" would now become "infallible true belief". This eliminates the loophole, but it is an unpopular solution because adding it is assumed to reduce epistemology to radical skepticism in which almost nothing is knowledge.

Here, we will propose the concept of the atomic fact as a new candidate to serve as the foundation to knowledge. The term 'atomic fact' has been used by some authors before, notably Bertrand Russell, but our mathematical take on the concept is unrelated to his, except on a connotative level. As we will see in a moment atomic facts, due to their construction, are sufficiently strong to be infallible, yet have sufficient expressive power to form a Turing complete theory thus they resolve the Gettier problem without reducing epistemology. Since they form a Turing complete language, are infallible and are true, we will qualify atomic facts as *epistemologically complete*. Atomic facts are the primary building blocks of our formal system of science and they will be verified by what we will call formal scientific methods. Finally, as they relate to programs and are thus subject to the halting problem, we will be able to use them to import experimental methods into pure mathematics by referencing these features[3].

1.1 Atomic Facts

Let us take the example of a statement that may appear as an obvious true statement such as "$1 + 1 = 2$", but is in fact not infallible. Here, I will provide what I believe to be the correct definition of an infallible statement, but equally important, such that the set of all such statements is Turing complete, thus forming a language of maximum expressive power. I will use the term *atomic fact* to refer to the concept.

Specifically, the sentence "$1 + 1 = 2$" halts on some Turing machine, but not on others and thus is not an atomic fact. Instead consider the sentence $\text{PA} \vdash [s(0)+s(0) = s(s(0))]$ 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 a loophole to fail the claim. Nonetheless, even with this improvement, a trickster can fail the claim by providing a Turing machine for which $\text{PA} \vdash [s(0) + s(0) = s(s(0))]$ does not halt.

If we use the tools of theoretical computer science and observe an equivalence with programs, we can produce statements free of all loopholes, thus ensuring they are infallible:

Definition 1 (Atomic Fact)**.** *Let $\mathbb{L}$ be the set of all sentences with alphabet Σ. An atomic fact f is a pair (TM, p) of sentences from $\mathbb{L} \times \mathbb{L}$ such that a universal*

Turing machine UTM *halts for it:*

$$\textit{iff } \mathrm{UTM}(\mathrm{TM}, p) \textit{ halts, then } f = (\mathrm{TM}, p) \textit{ is an atomic fact} \tag{1}$$

A universal Turing machine UTM which takes a Turing machine TM and a sentence p as inputs, will halt if and only if p halts on TM. Thus a claim that p halts on TM, if true, is an atomic fact because it is verifiable on all universal Turing machines.

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 atomic facts constitutes the entire domain of the universal Turing machine, and thus the expressive power of atomic facts must be on par with any Turing complete language. Since there exists no greater expressive power for a formal language than that of Turing completeness, then no reduction takes place.

1.2 Knowledge-Bases > Formal Axiomatic Systems

We can use atomic facts to redefine the foundations of mathematics in terms of *knowledge-bases*, rather in terms of *formal axiomatic systems*.

Definition 2 (Knowledge base)**.** *Let* $\mathbb{D} = \mathrm{Dom}(\mathrm{UTM})$ *be the set of all atomic facts for a given universal Turing machine. A knowledge base* $\mathbf{kb}$ *of* n *atomic fact is an element of the* n*-fold Cartesian product of* $\mathbb{D}$*:*

$$\mathbf{kb} \in \mathbb{D}^n, \quad \mathbf{kb} := \big((\mathrm{TM}_1, p_1), \ldots, (\mathrm{TM}_n, p_n)\big) \tag{2}$$

The tuple, in principle, can be empty $\mathbf{kb} := ()$*, finite* $n \in \mathbb{N}$ *or countably infinite* $n = \infty$*.*

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

Let us now explore the advantages of using knowledge-bases. Its most notable advantage is that it coincides to a complete *description* of reality of the type we are required to have as the starting point of a formal system of science. This is quite the concept to grasp and absorb and so we will proceed step by step. Comparatively, formal axiomatic systems would be more akin to an *interpretation* of reality based on a preference of some patterns or tools (we like sets, thus ZFC!, or we prefer categories, thus category theory!). The artefacts introduced by picking a preferential formal axiomatic systems, rather than been

solely concerned with describing reality non-preferentially, in turn make it impossible to derive the laws of physics from first principles, as we erased said 'first principle' by picking the preference. Here, said 'first principle' refers to using a non-preferential description of reality as the starting point, as that is what reality gives us to work it. Knowledge-bases, unlike formal axiomatic systems, are logic's only non-preferential starting points. As we explore the concept of knowledge-bases further as well as its power, let us now use an example.

Suppose we wish to represent in real-time, and with live updates, the set of all knowledge produced by a group of 50,000ish (and growing!) mathematicians working in a decentralized manner (perhaps from their offices) over the course of at least many decades, and perhaps even for an indefinite amount of time into the future. Some of the work they produced may build on each others', but it will also be the case that part of their work is incompatible. For instance, some might find contradictions in their assumptions and abandon large segments of their work. As one learns primarily from his or her errors, we may wish to catalogue these contractions for posterity. Finding the 'correct' and singular formal axiomatic system to describe the totality of what they have discovered, including abandoned work and contradictions, will be quite a challenge. One challenge occurs whenever a new contradiction is found, as one would need to further isolate it within a wrapper of para-consistent logic, before inclusion within the all-encompassing formal axiomatic system. Another challenge, this time unsurmountable, occurs when mathematicians invent new, possibly more elegant, axiomatic basis outright. One would constantly need to adjust his or her proposed all-encompassing formal axiomatic system to account for new discoveries as they are made. Such an axiomatic basis would eventually grow to an unmaintainable level, not unlike the spaghetti codes of the early days of software engineering. As mathematicians are a creative bunch, one would never be able to settle on a final system as they could always decide to explore a segment of mathematical space not covered by the current system. Comparatively, with a knowledge-base framework, the task is much easier: One simply need to push each new discovery at the end of the sequence; no adjustment is ever required after insertion, we never run out of space, and atomic facts do not undermine each other even if one represents a known contradiction.

Knowledge-bases are the true starting point of the logical inquiry as they represent an infialible and non-preferential description of reality, whereas axioms, contrary to what was advertised, are in fact a pre-interpreted mid-way "pseudo-starting" point containing a load of prior context in the form of a choice of preferred tool.

We will now explore the concept more rigorously.

Note on the notation: we will designate $f_i = (\mathrm{TM}_i, p_i)$ as an atomic fact element of **kb**, and $\mathrm{proj}_1(f_i)$ and $\mathrm{proj}_2(f_i)$ designate the first and second projection of the pair f_i, respectively. Thus $\mathrm{proj}_1(f_i)$ is the TM_i associated with f_i, and $\mathrm{proj}_2(f_i)$ is the input p_i associated with f_i. If applied to a tuple or set of pairs, then $\mathrm{proj}_1(\mathbf{kb})$ returns the set of all p in **kb** and $\mathrm{proj}_2(\mathbf{kb})$ returns the set of all TM in **kb**.

Theorem 1 (Incompleteness Theorem)**.** *Let* **kb** *be a knowledge base. If* $\mathbf{kb} = \text{Dom(UTM)}$*, then* **kb** *is recursively enumerable (and non-decidable). The proof follows from the domain of a universal Turing machine being non-decidable.*

Definition 3 (Premises)**.** *Let* **kb** *be a knowledge base. The premises in* **kb** *are defined as the set of all* TM *in* **kb***:*

$$\mathbb{P} := \text{proj}_1(\mathbf{kb}) \tag{3}$$

Definition 4 (Theorems)**.** *Let* **kb** *be a knowledge base. The theorems of* **kb** *are defined as the set of all* p *in* $\mathbb{M}$*:*

$$\mathbb{T} := \text{proj}_2(\mathbf{kb}) \tag{4}$$

Definition 5 (Spread (of a theorem))**.** *The set of all premises in* **kb** *in which a theorem is repeated is called the spread of the theorem. For instance if* $\mathbf{kb} = \big(\text{TM}_1, p_1), (\text{TM}_2, p_1)\big)$*, then the spread of* p_1 *is* $\{\text{TM}_1, \text{TM}_2\}$*.*

Definition 6 (Scope (of a premise))**.** *The set of all theorems in* **kb** *in which a premise is repeated is called the scope of the premise. For instance if* $\mathbf{kb} = \{(\text{TM}_1, p_1), (\text{TM}_1, p_2)\}$*, then the scope of* TM_1 *is* $\{p_1, p_2\}$*.*

1.2.1 Connection to Formal Axiomatic Systems

We can, of course, connect our construction to a formulation in terms of a formal axiomatic system (FAS):

Definition 7 (Enumerator (of a FAS))**.** *A function* $\text{enumerator}_{\text{FAS}}$ *is an enumerator for* FAS *if it recursively enumerates the theorems of* FAS*. For instance:*

$$\text{enumerator}_{\text{FAS}}(s) = \begin{cases} 1 & \text{FAS} \vdash s \\ \nexists / \text{ does-not-halt} & \text{otherwise} \end{cases} \tag{5}$$

Definition 8 (Domain (of FAS))**.** *Let* FAS *be a formal axiomatic system, let* **kb** *be a knowledge base and let* $\text{enumerator}_{\text{FAS}}$ *be a function which recursively enumerates the theorems of* FAS*. Then the domain of* FAS*, denoted as* Dom(FAS)*, is the set of all sentences* $s \in \mathbb{L}$ *which halts for* $\text{enumerator}_{\text{FAS}}$*.*

Definition 9 (Formal Axiomatic Representation)**.** *Let* FAS *be a formal axiomatic system, let* **kb** *be a knowledge base and let* $\text{enumerator}_{\text{FAS}}$ *be a function which recursively enumerates the theorems of* FAS*. Then* FAS *is a formal axiomatic representation of* **kb** *iff:*

$$\text{Dom(FAS)} = \text{proj}_2(\mathbf{kb}) \tag{6}$$

Definition 10 (Factual Isomorphism)**.** *Two formal axiomatic systems* FAS_1 *and* FAS_2 *are factually-isomorphic if and only if* $\text{Dom}(\text{FAS}_1) = \text{Dom}(\text{FAS}_2)$*.*

1.2.2 Axiomatic Information

Although we can connect the formulation of a knowledge base to a formal axiomatic representation, we will find that it is more advantageous to stick with knowledge-bases. The first result of interest will be the introduction of axiomatic information. We can understand the elements of any particular knowledge base as having been 'picked', in some sense, from the set of all possible atomic facts. If the pick is random and described as a probability measure ρ, we can quantify the information of the pick using the entropy.

Definition 11 (Axiomatic Information)**.** *Let* $\mathbb{D} = \text{Dom(UTM)}$ *be the domain of a universal Turing machine. Then, let* $\rho : \mathbb{D} \to [0, 1]$ *be a probability measure over* $\mathbb{D}$*. Finally, let* $\mathbf{kb}$ *be a knowledge base subset of* $\mathbb{D}^n$*. The axiomatic information of a single element of* $\mathbf{kb}$ *is quantified as the entropy of* ρ*:*

$$S = -\sum_{q \in \mathbb{D}} \rho(q) \ln \rho(q) \tag{7}$$

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

$$\Omega = \sum_{p \in \text{Dom(UTM)}} 2^{-|p|} \implies \rho(p) = 2^{-|p|} \tag{8}$$

The quantity of axiomatic information of a given knowledge-base (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. A strategy to gather mathematical knowledge which picks atomic facts according to the probability measure which maximizes the entropy will be a maximally informative strategy.

1.3 Discussion — The Mathematics of Knowledge

Each element of a knowledge base is a program-input pair representing an algorithm which is known to produce a specific result. Let us see a few examples.

How does one know how to tie one's shoes? One knows the algorithm required to produce a knot in the laces of the shoe. How does one train for a new job? One learns the internal procedures of the shop, which are known to produce the result expected by management. How does one impress management? One learns additional skills outside of work and applies them at work to produce results that exceed the expectation of management. How does one create a state in which there is milk in the fridge? One ties his shoes, walks to the store, pays for milk using the bonus from his or her job, then brings the milk back home and finally places it in the fridge. How does a baby learn about object permanence? One plays peak-a-boo repeatedly with a baby, until it ceases to amuse the baby — at which point the algorithm which hides the

parent, then shows him or her again, is learned as knowledge. How does one untie his shoes? One simply pulls on the tip of the laces. How does one untie his shoes if, after partial pulling, the knot accidentally tangles itself preventing further pulling? One uses his fingers or nails to untangle the knot, and then tries pulling again.

Knowledge can also be in more abstract form — for instance in the form of a definition that holds for a special case. How does one know that a specific item fits a given definition of a chair? One iterates through all properties referenced by the definition of the chair, each step confirming the item has the given property — then if it does for all properties, it is known to be a chair according to the given definition.

In all cases, knowledge is an algorithm along with an input, such that the algorithm halts for it, lest it is not knowledge. The set of all known pairs forms a knowledge base.

Let us consider a few edge cases. What if a knowledge base contains both "A" and "not A" as theorems? For instance, consider:

$$\mathbf{kb} := \big((\mathrm{TM}_1, A), (\mathrm{TM}_1, \neg A)\big) \tag{9}$$

Does allowing contradictions at the level of the theorems of **kb** create a problem? Should we add a few restrictions to avoid this unfortunate scenario? Let us try an experiment to see what happens — specifically, let me try to introduce $A \wedge \neg A$ into my personal knowledge base, and then we will evaluate the damage I have been subjected to by this insertion. Consider the following program TM_1:

1. If (p="A" or p="not A") then
2. return 1;
3. else (loop())

It thus appears that I can have knowledge that the above program halts for both "A" and "not A" and still survive to tell the tale. A-priori, the sentences "A" and "not A" are just symbols. Our reflex to attribute the law of excluded middle to these sentences requires the adoption of a deductive system. This occurs one step further at the selection of a specific formal axiomatic representation of the knowledge base, and not at the level of the knowledge base itself.

The only inconsistency that would create problems for this framework would be a proof that a given atomic fact both [HALTS] and [NOT HALTS] on a UTM. By definition of a UTM, this cannot happen lest the machine was not a UTM to begin with. Thus, we are expected to be safe from such contradictions.

Now, suppose one has a sizeable knowledge base which may contain a plurality of pairs:

$$\mathbf{kb} := \big((\mathrm{TM}_1, p_1), (\mathrm{TM}_2, \neg p_1), (\mathrm{TM}_1, p_2), (\mathrm{TM}_2, p_1), (\mathrm{TM}_2, \neg p_3)\big) \tag{10}$$

Here, the negation of some, but not all, is also present across the pairs: in this instance, the theorems p_1 and p_3 are negated but for different premises. What interpretation can we give to such elements of a knowledge base? For our example, let us call the sentences p_1, p_2, p_3 the various flavours 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 atomic facts are pairs which contain a premise and a theorem. The premise contains the 'definition' under which the flavour qualifies as real ice cream. A flavour with a large spread is considered real ice cream by most definitions (i.e. vanilla or chocolate ice cream), and one with a tiny spread would be considered real ice cream by only very few definitions (i.e. tofu-based ice cream). Then, within this example, the presence of p_1 and its negation simply means that tofu-based ice cream is ice cream according to one definition, but not according to another.

Reality is of a complexity such that a one-size-fits-all definition does not work for 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 its knowledge base ought to be sufficiently flexible to handle this, without itself exploding into a contradiction.

Even in the case where both A and its negation $\neg A$ were to be theorems of **kb** while also having the same premise, is still knowledge. It means one has verified that said premise is inconsistent. 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 knowledge base.

2 The Formal System of Science

We now assign to our knowledge-based reformulation of mathematics the interpretation of a purely mathematical system of science. This is motivated principally because constructing a knowledge base, as it faces the non-halting problem, makes it an experimental process. Indeed, the halting problem of Alan Turing necessitates that, in the general case, one must try programs on a UTM until termination to identify those that halt. This connects the acquisition of mathematical knowledge to a trial and error approach which cannot be decided by a general algorithm.

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* $f = (\mathrm{TM}, p)$ *be a pair comprised 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.*

$$\mathrm{UTM}(\mathrm{TM}, p) \begin{cases} \mathrm{HALT} & \implies \textit{the experiment verifies } p \\ \neg\,\mathrm{HALT} & \implies \textit{the pair fails verification} \end{cases} \tag{11}$$

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. In the general case, one must try them out to see which one halts — this is why they are called experiments.

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 2 (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 possible definition of the experiment that is formally reproducible (as opposed to say "sufficiently reproducible for practical purposes"), yet universal as a Turing complete language.

Definition 13 (Domain of science)**.** *We note* $\mathbb{D}$ *as the domain* (Dom) *of science. We can define* $\mathbb{D}$ *in reference to a universal Turing machine* UTM *as follows:*

$$\mathbb{D} := \mathrm{Dom}(\mathrm{UTM}) \tag{12}$$

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

Definition 14 (Experimental Space)**.** *Let* **kb** *be a knowledge base comprised of* n *atomic facts, and let* $\mathbb{KB} = \bigcup_{i=1}^{n} \mathrm{proj}_i(\mathbf{kb})$ *be the set comprised of the atomic facts of* **kb***. The experimental space* $\mathbb{E}$ *of* **kb** *is the "powertuple" of* **kb***:*

$$\mathbb{E} := \bigcup_{i=0}^{n} (\mathbb{KB})^i \tag{13}$$

- *Put simply, experimental space is the set of all possible knowledge-bases (including the empty knowledge-base).*
- *Conceptually, a powertuple is similar to a powerset where the notion of the set is replaced by that of the tuple.*
- *All elements of an experimental space are knowledge-bases, and all "sub-tuples" of a knowledge-base are elements of its experimental space.*

Definition 15 (Scientific method)**.** *An algorithm which recursively enumerates experiments is called a scientific method. A scientific method thus produces a knowledge-base.*

Theorem 3 (Scientific method (Existence of the general case))**.** *Existence of the scientific method for the general case of the full domain of science.*

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.

2.1.1 Classification of Scientific Theories

Definition 16 (Scientific Theory)**.** *Let* **kb** *be a knowledge-base and let* ST *be a formal axiomatic system. If*

$$\mathrm{proj}_2(\mathbf{kb}) \cap \mathrm{Dom}(\mathrm{ST}) \neq \varnothing \tag{14}$$

then ST *is a scientific theory of* **kb***.*

Definition 17 (Empirical Theory)**.** *Let* **kb** *be a knowledge-base and let* ST *be a scientific theory. If*

$$\mathrm{proj}_2(\mathbf{kb}) = \mathrm{Dom}(\mathrm{ST}) \tag{15}$$

then ST *is an empirical theory of* **kb***.*

Definition 18 (Scientific Field)**.** *Let* **kb** *be a knowledge-base and let* ST *be a scientific theory. If*

$$\mathrm{Dom}(\mathrm{ST}) \subset \mathrm{proj}_2(\mathbf{kb}) \tag{16}$$

then ST *is a scientific field of* **kb***.*

Definition 19 (Predictive Theory)**.** *Let* **kb** *be a knowledge-base and let* ST *be a scientific theory. If*

$$\mathrm{proj}_2(\mathbf{kb}) \subset \mathrm{Dom}(\mathrm{ST}) \tag{17}$$

then ST *is a predictive theory of* **kb***.*
Specifically, the predictions of ST *are given as follows:*

$$\mathbb{S} := \mathrm{Dom}(\mathrm{ST}) \setminus \mathbf{kb} \tag{18}$$

Scientific theories that are predictive theories are supported by experiments, but may diverge outside of this support.

2.1.2 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.

Theorem 4 (The Fundamental Theorem of Science)**.** *Let* $\mathbf{kb}_1$ *and* $\mathbf{kb}_2$ *be two knowledge-bases, such that the later includes and is larger than the former:* $\mathbf{kb}_1 \subset \mathbf{kb}_2$*. If* ET_2 *is an empirical theory of* $\mathbf{kb}_2$*, then it follows that* ET_2 *is a predictive theory of* $\mathbf{kb}_1$*. Finally, up to factual-isomorphism,* $\mathrm{Dom}(\mathrm{ET}_2)$ *has measure* 0 *over the set of all distinct domains spawned by the predictive theories of* $\mathbf{kb}_2$*.*

Proof. $\mathrm{Dom}(\mathrm{ET}_2)$ is unique. Yet, the number of distinct domains spawned by the set of all possible predictive theories of $\mathbf{kb}_1$ is infinite. Finally, the measure of one element of an 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 predictive scientific theory will eventually be falsified.

2.2 Axiomatic Foundation of Physics

Recall that earlier we used a dovetailing algorithm in Theorem 3 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? 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. It is more strategic overall to identify a condition applicable to all cases, which produces the implementation as a maximization problem.

The best strategy will be to *maximize the axiomatic information* gained from the scientific method as it produces a knowledge-base, and this means, in the technical sense, to maximize the entropy of a probability measure on experimental space.

To have information (in the information-theoretical sense) one must work with a probability measure.

Axiom 1 (Observer)**.** *An observer, denoted as* $\mathcal{O}$*, is an element randomly selected from experimental space* $\mathbb{E}$ *according to a probability measure* ρ*. Specifically, an observer is a triplet (reminiscent of measure theory):*

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

where ρ *is a probability measure.*

The definition of the observer is a specialization of the definition of the knowledge-base in the sense that a knowledge-base is an element of experimental space, and the observer is a *randomly selected element* of experimental space (and thus the notion of information is associated to it). Note that typically in 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, inherited from the requirement of maximizing axiomatic information. Although it is still obscure at this point, this definition will perfectly coincide with what we understand an observer to be in quantum physics, and also to what we understand it to be in general relativity.

Axiom 2 (Laws of Physics)**.** *The laws of physics are the probability measure ρ that maximizes the entropy S of an element $\mathcal{O}$ randomly selected in $\mathbb{E}$.*

The definition of the laws of physics are a specialization of the definition of the observer, in the sense that an observer is a *randomly selected element of experimental space*, and the laws of physics are its *entropy-maximized* version. The laws of physics are thus defined formally as the *information-theoretical maximum* of the scientific method for an observer. The best strategy to maximize ones axiomatic knowledge of reality produces the laws of physics as the solution. We will derive this measure explicitly in the Main Result section.

3 Main Result

Let us now use these definitions to derive the laws of physics from first principle, and show that they are at least the same as the laws of physics, and then investigate extended results and predictions.

Our starting point will be the definition of the observer, then we will maximize the entropy of ρ using the method of the Lagrange multipliers. We recall that our definition of the observer is:

$$\mathcal{O}_i := (\mathbf{kb}_i, \mathbb{E}, \rho : \mathbb{E} \to [0, 1]) \tag{20}$$

where $\mathbf{kb}_i$ is a n-tuple, $\mathbb{E}$ is a "powertuple" and ρ is a (probability) measure over $\mathbb{E}$.

Note the similarity between our definition of the observer to that of a measure space. Comparatively, the definition of a measure space is:

$$M := (X, \Sigma, \mu(X)) \tag{21}$$

where X is a set, Σ is (often) taken to be the powerset of X, and μ is a measure over Σ. The difference is simply that sets have been replaced by tuples. Consequently, we must adapt the definition of a measure space from set to tuples. To do so, we will use the following prescription:

1. We assign a non-negative number to each element of **kb**.
2. We then equip said numbers with the addition operation, converting the construction to a vector space.
3. We maximize the entropy of a single atomic fact under the effect of constraints, by using the method of the Lagrange multipliers.
4. We prescribe that any and all constraints on said entropy must remain invariant with respect to a change of basis of said vector space.
5. Finally, we use the tensor product n-times over said vector space to construct probability measure of n-tuples of atomic facts.

Explicitly, we maximize the entropy:

$$S = -\sum_{f \in \mathbb{KB}} \rho(f) \ln \rho(f) \tag{22}$$

subject to these constraints:

$$\sum_{f \in \mathbb{KB}} \rho(f) = 1 \tag{23}$$

$$\sum_{f \in \mathbb{KB}} \rho(f) \operatorname{tr} \mathbf{M}(f) = \operatorname{tr} \overline{\mathbf{M}} \tag{24}$$

where $\mathbf{M}(f)$ are a matrix-valued maps from $\mathbb{KB}$ to $\mathbb{C}^{n \times n}$ representing the linear transformations of the vector space and where $\overline{\mathbf{M}}$ is a element-by-element average matrix.

Usage of the trace of a matrix as a constraint imposes an invariance with respect to a similarity transformation, accounting for all possible linear reordering of the elements of the tuples of the sum, thus allowing the creation of a measure of a tuple or group of tuples form within a space of tuples, invariantly with respect to the order of the elements of the tuples. Similarity transformation invariance on the trace is the result of this identity:

$$\operatorname{tr} \mathbf{M} = \operatorname{tr} \mathbf{B}\mathbf{M}\mathbf{B}^{-1} \tag{25}$$

We now 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:

$$\mathcal{L} = -k_B \sum_{f \in \mathbb{KB}} \rho(f) \ln(f) + \alpha \left(1 - \sum_{f \in \mathbb{KB}} \rho(f) \right) + \tau \left(\operatorname{tr} \overline{\mathbf{M}} - \sum_{f \in \mathbb{KB}} \rho(f) \operatorname{tr} \mathbf{M}(f) \right) \tag{26}$$

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

$$\rho(f,\tau) = \frac{1}{Z(\tau)} \det \exp -\tau \mathbf{M}(f) \tag{27}$$

where

$$Z(\tau) = \sum_{f \in \mathbb{KB}} \det \exp -\tau \mathbf{M}(f) \tag{28}$$

3.1 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 $\mathbf{P} : \mathbb{Q} \to \mathbb{C}^{n \times n}$ and inject it into the probability measure as well as into the partition function:

$$\rho(f) = \frac{1}{Z} \det \exp \big(\mathbf{P}(f)\big) \det \exp \big(-\tau \mathbf{M}(f)\big) \tag{29}$$

where

$$Z = \sum_{f \in \mathbb{KB}} \det \exp \big(\mathbf{P}(f)\big) \det \exp \big(-\tau \mathbf{M}(f)\big) \tag{30}$$

3.2 Overview

We will begin with a small overview connecting the main result to its origins as a knowledge-base, via a passage into quantum computation, then we will expand the connection to physics into two parts:

Part 1: We will show that a special case of this result is an equivalent representation of standard quantum mechanics, and discuss the implications of this formulation.

Part 2: We will show that this result, in the general case, is a quantum theory of gravity which adheres to both 1) the axioms of quantum mechanics and 2) to the theory of general relativity.

3.2.1 Quantum Computing (Recall)

Let us begin by reviewing the basics of quantum computation. One starts with a state vector:

$$|\psi_a\rangle = \begin{pmatrix} 0 \\ \vdots \\ n \end{pmatrix} \tag{31}$$

Which evolves unitarily to a final state:

$$|\psi_b\rangle = U_0 U_1 \dots U_m \, |\psi_a\rangle \tag{32}$$

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 add more gates in order continue with a computation:

$$\text{step 1} \qquad |\psi_b\rangle = U_0 U_1 \dots U_p \, |\psi_a\rangle \tag{33}$$

$$\text{step 2} \qquad |\psi_c\rangle = U'_0 U'_1 \dots U'_q \, |\psi_b\rangle \tag{34}$$

$$\vdots \tag{35}$$

$$\text{step k} \qquad |\psi_{k'}\rangle = U'_0 U'_1 \dots U'_v \, |\psi_k\rangle \tag{36}$$

$$\vdots \tag{37}$$

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). But note, that each step is itself a completed program, and further it is the case that each step can be infinitely divided.

3.2.2 Sketch

The linear transformations of our main result are here interpreted in the same manner as those used in quantum computations, but extended to the general linear group.

We begin by splitting the probability measure into a first step, which is linear with respect to a 'probability amplitude', and a second which connects the amplitude to the probability. We thus write the probability measure as:

$$\rho(f, \tau) = \frac{1}{Z} \det \psi(f, \tau) \tag{38}$$

where

$$\psi(f,\tau) = \exp\big(\mathbf{P}(f)\big)\exp\big(-\tau\mathbf{M}(f)\big) \tag{39}$$

Here, the determinant is interpreted as a generalization of the Born rule and reduces to exactly it when $\mathbf{M}$ is the matrix representation of the complex numbers. In the general case where $\mathbf{M}$ are arbitrary $n \times n$ matrices, $\psi(f,\tau)$ will be called the *general linear probability amplitude*.

We can write $\psi(f,\tau)$ as a column vector:

$$\boldsymbol{\psi} := |\psi\rangle := \begin{pmatrix} \psi(f_1,\tau) \\ \psi(f_2,\tau) \\ \vdots \\ \psi(f_n,\tau) \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_n \end{pmatrix} \tag{40}$$

Paths will be constructed by chaining transformations on those vectors:

$$|\psi_b\rangle = \underbrace{\mathbf{T}_1\mathbf{T}_2\ldots\mathbf{T}_n}_{\text{computing steps}}|\psi_a\rangle \tag{41}$$

As more transformations are chained, progressively richer manifests are constructed. Paths in experimental space are realized by completing the missing computational steps required for a starting-point manifest to be the end-point manifest.

Comparatively, quantum mechanical computations are simply a special cases when he transformations are unitary:

$$|\psi_b\rangle = \underbrace{\mathbf{U}_1\mathbf{U}_2\ldots\mathbf{U}_n}_{\text{computing steps}}|\psi_a\rangle \tag{42}$$

3.2.3 Matrix-Valued Vector and Transformations

To work with the general linear probability amplitude, we will use vectors whose elements are matrices. An example of such a vector is:

$$\mathbf{v} = \begin{pmatrix} \mathbf{M}_1 \\ \vdots \\ \mathbf{M}_m \end{pmatrix} \tag{43}$$

Likewise a linear transformation of this space will expressed as a matrix of matrices:

$$\mathbf{T} = \begin{pmatrix} \mathbf{M}_{00} & \dots & \mathbf{M}_{0m} \\ \vdots & \ddots & \vdots \\ \mathbf{M}_{m0} & \dots & \mathbf{M}_{mm} \end{pmatrix} \tag{44}$$

Note: The scalar element of the vector space are given as:

$$a\mathbf{v} = \begin{pmatrix} a\mathbf{M}_1 \\ \vdots \\ a\mathbf{M}_m \end{pmatrix} \tag{45}$$

3.2.4 Sums of Paths

So far, the sums of programs we have used were over knowledge-bases comprised of a single program each. How do we extend this to knowledge-bases containing multiple programs? We have to use a Cartesian product on the sets of manifests and a tensor product on the probability amplitudes. For instance, let us consider the following sets of atomic facts:

$$\mathbb{KB}_1 = \{(f_{1a}), (f_{1b})\} \tag{46}$$
$$\mathbb{KB}_2 = \{(f_{2a}), (f_{2b})\} \tag{47}$$

The Cartesian product produces knowledge-bases comprised of two elements:

$$\mathbf{kb} \in \mathbb{KB}_1 \times \mathbb{KB}_2 = \{(f_{1a}, f_{2a}), (f_{1a}, f_{2b}), (f_{1b}, f_{2a}), (f_{1b}, f_{2b})\} \tag{48}$$

At the level of the probability amplitude, the Cartesian product of sets translates to the tensor product. For instance, we start with a wave-function of one program;

$$\boldsymbol{\psi}_1 = \begin{pmatrix} \exp \mathbf{P}(f_{1a}) \\ \exp \mathbf{P}(f_{1b}) \end{pmatrix} \tag{49}$$

Adding a program-step via a linear transformation produces:

$$\mathbf{T}\boldsymbol{\psi}_1 = \begin{pmatrix} T_{00} \exp \mathbf{P}(f_{1a}) + T_{01} \exp \mathbf{P}(f_{1b}) \\ T_{10} \exp \mathbf{P}(f_{1a}) + T_{11} \exp \mathbf{P}(f_{1b}) \end{pmatrix} \tag{50}$$

If we tensor product this wave-function:

$$\boldsymbol{\psi}_2 = \begin{pmatrix} \exp \mathbf{P}(f_{2a}) \\ \exp \mathbf{P}(f_{2b}) \end{pmatrix} \tag{51}$$

along with a program-step:

$$\mathbf{T}'\boldsymbol{\psi}_2 = \begin{pmatrix} T'_{00}\exp\mathbf{P}(f_{2a}) + T'_{01}\exp\mathbf{P}(f_{2b}) \\ T'_{10}\exp\mathbf{P}(f_{2a}) + T'_{11}\exp\mathbf{P}(f_{2b}) \end{pmatrix} \tag{52}$$

Then the tensor product of these states produces the probability measure of a knowledge-base as follows:

$$\mathbf{T}\boldsymbol{\psi}_1 \otimes \mathbf{T}'\boldsymbol{\psi}_2 = \begin{pmatrix} (T_{00}\exp\mathbf{P}(f_{1a}) + T_{01}\exp\mathbf{P}(f_{1b}))(T'_{00}\exp\mathbf{P}(f_{2a}) + T'_{01}\exp\mathbf{P}(f_{2b})) \\ (T_{00}\exp\mathbf{P}(f_{1a}) + T_{01}\exp\mathbf{P}(f_{1b}))(T'_{10}\exp\mathbf{P}(f_{2a}) + T'_{11}\exp\mathbf{P}(f_{2b})) \\ (T_{10}\exp\mathbf{P}(f_{1a}) + T_{11}\exp\mathbf{P}(f_{1b}))(T'_{00}\exp\mathbf{P}(f_{2a}) + T'_{01}\exp\mathbf{P}(f_{2b})) \\ (T_{10}\exp\mathbf{P}(f_{1a}) + T_{11}\exp\mathbf{P}(f_{1b}))(T'_{10}\exp\mathbf{P}(f_{2a}) + T'_{11}\exp\mathbf{P}(f_{2b})) \end{pmatrix} \tag{53}$$

Now, each element of the resulting vector is a knowledge-base of two programs, but its probability is a sum over a path. One can repeat the process n times, and even take to limit to infinity.

4 Foundation of Physics (Standard QM)

Remarkably, the main result, even if it is barely three lines, is able to produce, as theorems, all axioms of quantum mechanics –including the famously elusive wave-function collapse mechanism, and the origin of the Born rule– thus it constitute, necessarily, a more fundamental formalism of QM that all other alternatives requiring axiomatic definitions. Finally, it necessitate, implies and guarantees not only ensemble interpretation of quantum mechanics, but also derives it from maximum entropy principles. It thus offers a "total solution" to the foundations, origins and interpretation of QM.

The study of the main result will be in two parts. In the first part, to recover QM, we will reduce the expressivity of the constraint of the main result to that of the matrix representation of the complex number, as opposed to a general matrix. In this special case, we will recover standard QM without gravity. In the section following this one, we will investigate the unadulterated version of the main result and show that it produces a complete theory of quantum gravity adhering both to 1) all axioms of QM, and also 2) is a complete formulation of general relativity, thus producing what we interpret as a theory of quantum gravity.

4.1 Born Rule

Definition 20 (Born Rule)**.** *The standard definition of the Born rule connects the probability to the complex norm of the wave-function:*

$$\rho = |\psi|^2 \tag{54}$$

The Born rule was postulated by Born in 1926, but attempts to derive it from first principles has been elusive. Notable proposals are those of Andrew M. Gleason[5], Kastner, R. E[6], and Lluis Masanes and Thomas Galley[7], and others.

In our formalism the determinant rule is the result of maximizing the entropy of a measure over a tuple. In the case of a matrix representation of the complex number, the determinant rule becomes the Born rule. Indeed, we first note that a complex number is represented by a matrix as follows:

$$a + ib \cong \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \tag{55}$$

Then, we note that its determinant is the same as the complex norm:

$$\det \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = a^2 + b^2 \tag{56}$$

Finally, the determinant rule reduces to the Born rule forthe complex case:

$$\rho(f, \tau) = \frac{1}{Z} \det \left(\exp -\beta \begin{pmatrix} a(f) & 0 \\ 0 & a(f) \end{pmatrix} \exp -\tau \begin{pmatrix} 0 & -b(f) \\ b(f) & 0 \end{pmatrix} \right) \tag{57}$$

$$= \frac{1}{Z} \exp -\beta 2a(f) \tag{58}$$

where $\exp -\beta 2a(f)$ is the preparation of the ensemble (the prior) and Z is the normalization constants. In the case where $\beta = 1/k_B T$, then it is a thermal preparation. The Born rule is thus revealed to be the probability measure derived from maximizing the entropy of the selection of a tuple from a set of tuples, and in this sense is analogous to the Gibbs measure in statistical physics.

4.2 Axioms of QM

4.2.1 Dirac–von Neumann Axioms (State vector)

The standard Dirac–von Neumann Axioms are recovered as theorems. This becomes apparent if we split the probability measure into a two-step process:

$$\rho(f, \tau) = \det \psi(f, \tau)|_{\psi \in \mathbb{C}} = |\psi(f, \tau)|^2 \tag{59}$$

where

$$\psi(f, \tau) = \frac{1}{\sqrt{Z}} \left(\exp -\beta \begin{pmatrix} a(f) & 0 \\ 0 & a(f) \end{pmatrix} \exp -\tau \begin{pmatrix} 0 & -b(f) \\ b(f) & 0 \end{pmatrix} \right) \tag{60}$$

The formalism in terms of Hilbert space is obtained simply by taking ψ rather than ρ as the object of study. To show this is almost trivial; $\psi \in \mathbb{C}$ and ρ is given as the complex norm, therefore the ψ are the unit vectors of a complete complex vector space.

4.2.2 Dirac–von Neumann Axioms (Expectation Value)

In statistical physics, an observable is simply a real value tied to each element of the probability measure:

$$\overline{O} = \sum_{q \in \mathbb{Q}} \rho(q) O(q) \tag{61}$$

Applied to our matrix-based constraints, this definition becomes that of a self-adjoint operator. And the expectation is a sum as follows:

$$\overline{O} = \sum_{q \in \mathbb{Q}} \det \psi(q) O(q)|_{\psi \in \mathbb{C}} = \sum_{q \in \mathbb{Q}} \psi(q)^* \psi(q) O(q) \tag{62}$$

and is thus the same definition as that of the expectation value of a self-adjoint operator acting on the unit vectors of a complex Hilbert space:

$$\sum_{q \in \mathbb{Q}} \psi(q)^* \psi(q) O(q) = \langle \psi | \, O \, | \psi \rangle \tag{63}$$

4.2.3 Time-Evolution (Schrödinger equation)

The last axiom of QM is usually a statement that the time-evolution of a state vector ψ is given by the Schrödinger equation. To derive it from our framework, that we can write:

$$\psi(f, \tau) = \frac{1}{\sqrt{Z}} \left(\exp -\beta \begin{pmatrix} a(f) & 0 \\ 0 & a(f) \end{pmatrix} \exp -\tau \begin{pmatrix} 0 & -b(f) \\ b(f) & 0 \end{pmatrix} \right) \tag{64}$$

as:

$$\psi(f, \tau) = \frac{1}{\sqrt{Z}} \exp -\tau i b(f) \tag{65}$$

where, to simplify, we have taking the prior to be unity ($a(f) = 0$). This is the familiar form of the quantum mechanical unitary evolution operator, where we rewrite $\tau \to t$ and $b \to H$. The Taylor expansion to the first linear term: $U(\delta t) \approx 1 - i \delta t H$. Then:

$$\big|\psi(t+\delta t)\big\rangle = U(\delta t)\,\big|\psi(t)\big\rangle \approx (1 - i\delta t H)\,|\psi\rangle \tag{66}$$

$$\implies \big|\psi(t+\delta t)\big\rangle - |\psi\rangle \approx -i\delta t H\,|\psi\rangle \tag{67}$$

$$\implies i\frac{\big|\psi(t+\delta t)\big\rangle - |\psi\rangle}{\delta t} \approx H\,|\psi\rangle \tag{68}$$

$$\implies i\frac{\partial\,\big|\psi(t)\big\rangle}{\partial t} = H\,|\psi\rangle \tag{69}$$

which is the time-dependant Schrödinger equation.

4.3 Discussion — Wave-function Collapse, Interpretation of QM

The final piece of the puzzle for standard QM is to provide a mechanism for the wave-function collapse and an interpretation. First, let must detail the interpretation. Here we derive from first principles the probability measure (Born rule) as the fundamental object and the wave-function is simply a convenient construction to work with linear transformations. In fact, we have found the Born rule as a special case of the Gibbs measure and occurs when the entropy is maximized under the constraint of a phase-invariance instrument. Let us see in more details.

In statistical physics, constraints on the entropy are interpreted as instruments acting on the system. For instance, an energy constraint on the entropy:

$$\overline{E} = \sum_{q\in\mathbb{Q}} \rho(q)E(q) \tag{70}$$

is interpreted, physically, as a energy-meter measuring the system and producing a series of energy measurement $E_1, E_2, \ldots$ converging to a average value $\overline{E}$. Another typical constraint is that of the volume:

$$\overline{V} = \sum_{q\in\mathbb{Q}} \rho(q)V(q) \tag{71}$$

associated to a volume-meter acting on the system and converging towards an average volume value $\overline{V}$, also by producing a sequence of measurements of the volume $V_1, V_2, \ldots$. With these two constraints, the typical system of statistical phsycis is obtained, and its Gibbs measure is:

$$\rho(q,\beta,p) = \frac{1}{Z}\exp\big(-\beta(E(q)+pV(q))\big) \tag{72}$$

Comparatively, in the present recovery of QM, the statistical physics interpretation is also adopted. Instead of an energy-meter or a volume-meter, we have a phase-invariant meter, and the constraint is given as follows:

$$\mathrm{tr}\begin{pmatrix} 0 & -\overline{b} \\ \overline{b} & 0 \end{pmatrix} = \sum_{q\in\mathbb{Q}} \rho(q)\,\mathrm{tr}\begin{pmatrix} 0 & -b(q) \\ b(q) & 0 \end{pmatrix} \tag{73}$$

The usage of the trace enforces the phase-invariance of the instrument. Yet, and quite simply, maximizing the entropy under this constraints produces the probability measure of the wave-function including the Born rule. This is the true origin of the Born rule, here reported for the first time. The interpretation simply becomes that of an instrument performing a sequence of measurement on the system such that an average value is obtained, but instead of the simpler scalar instruments used in statistical physics, here we have a phase-invariant instrument. This instrument is responsible for the quantum mechanical behaviour associated with the wave-function. What is an example of such a detector; quite simply a photo-counter would be one. Such an instrument produces a sequence of incidences ('clicks') as photons are detected and "advanced features" such as an interference pattern is a consequence of this phase-invariance.

Now, the correct interpretation of quantum mechanics is simply the consequence of the existence of phase-invariant photon counters producing series of incidence counts which constrains the entropy of a physical system. Finally, maximizing said entropy using the typical tools of statistical physics produces the Born rule as a special case of the Gibbs ensemble. The interpretation of QM and statistical physics are now unified as the same interpretation.

It is not a collapse that occurs but instead a derivation of the wave-function from a sequence of incidences recorded by phase-invariant instruments. This is the natural interpretation; as a system of statistical physics whose entropy is maximized under proper phase-invariant constraint. Consequently, the interpretation does not admit a collapse problem, anymore than a coin landing on head creates a probability collapse problem (from 50-head/50-tail to 100-head after landing) for classical statistics.

5 Foundation of Physics (Quantum Gravity)

In this section we will show that the probability measure here-in derived adheres to 1) the axioms of quantum mechanics, and 2) is simultaneously a gauge-theoretical theory of gravitation.

We will introduce an *algebra of natural states* and we will use it to classify the linear transformations on said amplitude. We will start with the 2D case, then the 4D case. In all cases, the probability amplitude transforms linearly with respect to general linear transformations and the probability measure, obtained from the determinant, is positive-definite.

5.1 Algebra of Natural States, in 2D

The notation of our upcoming definitions will be significantly improved if we use a geometric representation for matrices. Let us therefore introduce a geometric

representation of 2×2 matrices.

5.1.1 Geometric Representation of 2×2 matrices

Let $\mathbb{G}(2, \mathbb{R})$ be the two-dimensional geometric algebra over the reals. We can write a general multi-vector of $\mathbb{G}(2, \mathbb{R})$ as follows:

$$\mathbf{u} = A + \mathbf{X} + \mathbf{B} \tag{74}$$

where A is a scalar, $\mathbf{X}$ is a vector and $\mathbf{B}$ is a pseudo-scalar. Each multi-vector has a structure-preserving (addition/multiplication) matrix representation. Explicitly, the multi-vectors of $\mathbb{G}(2, \mathbb{R})$ are represented as follows:

Definition 21 (Geometric representation of a matrix (2×2))**.**

$$A + X\hat{\mathbf{x}} + Y\hat{\mathbf{y}} + B\hat{\mathbf{x}} \wedge \hat{\mathbf{y}} \cong \begin{pmatrix} A + X & -B + Y \\ B + Y & A - X \end{pmatrix} \tag{75}$$

And the converse is also true, each 2×2 real matrix is represented as a multi-vector of $\mathbb{G}(2, \mathbb{R})$.

We can define the determinant solely using constructs of geometric algebra[8].

Definition 22 (Clifford conjugate (of a $\mathbb{G}(2, \mathbb{R})$ multi-vector))**.**

$$\mathbf{u}^{\ddagger} := \langle \mathbf{u} \rangle_0 - \langle \mathbf{u} \rangle_1 - \langle \mathbf{u} \rangle_2 \tag{76}$$

Then the determinant of $\mathbf{u}$ is:

Definition 23 (Geometric representation of the determinant (of a 2×2 matrix))**.**

$$\begin{array}{rccc} \det & : & \mathbb{G}(2, \mathbb{R}) & \longrightarrow & \mathbb{R} \\ & & \mathbf{u} & \longmapsto & \mathbf{u}^{\ddagger}\mathbf{u} \end{array} \tag{77}$$

For example:

$$\begin{aligned} \det \mathbf{u} &= (A - \mathbf{X} - \mathbf{B})(A + \mathbf{X} + \mathbf{B}) && (78) \\ &= A^2 - X^2 - Y^2 + B^2 && (79) \\ &= \det \begin{pmatrix} A + X & -B + Y \\ B + Y & A - X \end{pmatrix} && (80) \end{aligned}$$

Finally, we define the Clifford transpose:

Definition 24 (Clifford transpose (of a matrix of 2×2 matrix elements))**.** *The Clifford transpose is the geometric analogue to the conjugate transpose. Like the conjugate transpose can be interpreted as a transpose followed by an element-by-element application of the complex conjugate, here the Clifford transpose is*

a transpose, followed by an element-by-element application of the Clifford conjugate:

$$\begin{pmatrix} \mathbf{u}_{00} & \dots & \mathbf{u}_{0n} \\ \vdots & \ddots & \vdots \\ \mathbf{u}_{m0} & \dots & \mathbf{u}_{mn} \end{pmatrix}^{\ddagger} = \begin{pmatrix} \mathbf{u}_{00}^{\ddagger} & \dots & \mathbf{u}_{m0}^{\ddagger} \\ \vdots & \ddots & \vdots \\ \mathbf{u}_{m0} & \dots & \mathbf{u}_{nm}^{\ddagger} \end{pmatrix} \tag{81}$$

If applied to a vector, then:

$$\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_m \end{pmatrix}^{\ddagger} = \begin{pmatrix} \mathbf{v}_1^{\ddagger} & \dots \mathbf{v}_m^{\ddagger} \end{pmatrix} \tag{82}$$

5.1.2 Axiomatic Definition of the Algebra, in 2D

Let $\mathbb{V}$ be an m-dimensional vector space over $\mathbb{G}(2,\mathbb{R})$. A subset of vectors in $\mathbb{V}$ forms an algebra of natural states $\mathcal{A}(\mathbb{V})$ iff the following holds:

1. $\forall \boldsymbol{\psi} \in \mathcal{A}(\mathbb{V})$, the bilinear map:

$$\begin{array}{rccc} \langle \cdot, \cdot \rangle & : & \mathbb{V} \times \mathbb{V} & \longrightarrow & \mathbb{G}(2,\mathbb{R}) \\ & & \langle \mathbf{u}, \mathbf{v} \rangle & \longmapsto & \mathbf{u}^{\ddagger}\mathbf{v} \end{array} \tag{83}$$

   is positive-definite:

$$\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle \in \mathbb{R}_{>0} \tag{84}$$

2. $\forall \boldsymbol{\psi} \in \mathcal{A}(\mathbb{V})$, then for each element $\psi(q) \in \boldsymbol{\psi}$, the function:

$$\rho(\psi(q), \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \psi(q)^{\ddagger} \psi(q) \tag{85}$$

   is positive-definite:

$$\rho(\psi(q), \boldsymbol{\psi}) \in \mathbb{R}_{>0} \tag{86}$$

We note the following comments and definitions:

- From (1) and (2) it follows that $\forall \boldsymbol{\psi} \in \mathcal{A}(\mathbb{V})$, the probabilities sum to unity:

$$\sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \boldsymbol{\psi}) = 1 \tag{87}$$

- $\boldsymbol{\psi}$ is called a *natural* (or physical) state.

- $\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle$ is called the *partition function* of $\boldsymbol{\psi}$.
- $\rho(q, \boldsymbol{\psi})$ is called the *probability measure* (or generalized Born rule) of $\psi(q)$.
- The set of all matrices $\mathbf{T}$ acting on $\boldsymbol{\psi}$, as $\mathbf{T}\boldsymbol{\psi} \to \boldsymbol{\psi}'$, which leaves the sum of probabilities normalized (invariant):

$$\sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \mathbf{T}\boldsymbol{\psi}) = \sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \boldsymbol{\psi}) = 1 \tag{88}$$

  are the *natural* transformations of $\boldsymbol{\psi}$.
- A matrix $\mathbf{O}$ such that $\forall \mathbf{u} \forall \mathbf{v} \in \mathcal{A}(\mathbb{V})$:

$$\langle \mathbf{O}\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{O}\mathbf{v} \rangle \tag{89}$$

  is called an observable.
- The expectation value of an observable $\mathbf{O}$ is:

$$\langle \mathbf{O} \rangle = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \langle \mathbf{O}\boldsymbol{\psi}, \boldsymbol{\psi} \rangle \tag{90}$$

5.1.3 Reduction to Complex Hilbert Spaces

It is fairly easy to see that if we reduce the expression of our multi-vectors ($A + \mathbf{X} + \mathbf{B}|_{\mathbf{X} \to 0} = A + \mathbf{B}$ and further restrict $\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle \in \mathbb{R}_{>0}$ to $\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle = 1$, then we recover the unit vectors of the complex Hilbert spaces:

- Reduction to the conjugate transpose:

$$\Big(\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^{\ddagger} \mathbf{v} \Big|_{\mathbf{x} \to 0} \implies \langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^{\dagger} \mathbf{v} \tag{91}$$

- Reduction to the unitary transformations:

$$\big(\langle \mathbf{T}\mathbf{u}, \mathbf{T}\mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle \big|_{\mathbf{x} \to 0} \implies \mathbf{T}^{\dagger}\mathbf{T} = I \tag{92}$$

- Reduction to the Born rule:

$$\left(\rho(q, \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \psi(q)^{\ddagger} \psi(q) \right|_{\mathbf{X} \to 0} \implies \rho(q, \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \psi(q)^{\dagger} \psi(q) \tag{93}$$

- Reduction of observables to Hermitian operators:

$$\big(\langle \mathbf{O}\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{O}\mathbf{v} \rangle \big|_{\mathbf{X} \to 0} \implies \mathbf{O}^{\dagger} = \mathbf{O} \tag{94}$$

Under this reduction, the formalism becomes equivalent to the Dirac-Von-Neumann formalism of quantum mechanics.

5.1.4 Observable, in 2D — Self-Adjoint Operator

Let us now investigate the general case of an observable is 2D. A matrix $\mathbf{O}$ is an observable iff it is a self-adjoint operator; defined as:

$$\langle \mathbf{O}\boldsymbol{\phi}, \boldsymbol{\psi} \rangle = \langle \boldsymbol{\phi}, \mathbf{O}\boldsymbol{\psi} \rangle \tag{95}$$

$\forall \mathbf{u} \forall \mathbf{v} \in \mathbb{V}$.

Setup: Let $\mathbf{O} = \begin{pmatrix} O_{00} & O_{01} \\ O_{10} & O_{11} \end{pmatrix}$ be an observable. Let $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ be 2 two-state vectors $\boldsymbol{\phi} = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$ and $\boldsymbol{\psi} = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$. Here, the components ϕ_1, ϕ_2, ψ_1, ψ_2, O_{00}, O_{01}, O_{10}, O_{11} are multi-vectors of $\mathbb{G}(2, \mathbb{R})$.

Derivation: 1. Let us now calculate $\langle \mathbf{O}\boldsymbol{\phi}, \boldsymbol{\psi} \rangle$:

$$\begin{aligned} 2\langle \mathbf{O}\boldsymbol{\phi}, \boldsymbol{\psi} \rangle &= (O_{00}\phi_1 + O_{01}\phi_2)^\ddagger \psi_1 + \psi_1^\ddagger (O_{00}\phi_1 + O_{01}\phi_2) \\ &\quad + (O_{10}\phi_1 + O_{11}\phi_2)^\ddagger \psi_2 + \psi_2^\ddagger (O_{10}\phi_1 + O_{11}\phi_2) \qquad (96) \\ &= \phi_1^\ddagger O_{00}^\ddagger \psi_1 + \phi_2^\ddagger O_{01}^\ddagger \psi_1 + \psi_1^\ddagger O_{00}\phi_1 + \psi_1^\ddagger O_{01}\phi_2 \\ &\quad + \phi_1^\ddagger O_{10}^\ddagger \psi_2 + \phi_2^\ddagger O_{11}^\ddagger \psi_2 + \psi_2^\ddagger O_{10}\phi_1 + \psi_2^\ddagger O_{11}\phi_2 \qquad (97) \end{aligned}$$

2. Now, $\langle \boldsymbol{\phi}, \mathbf{O}\boldsymbol{\psi} \rangle$:

$$\begin{aligned} 2\langle \boldsymbol{\phi}, \mathbf{O}\boldsymbol{\psi} \rangle &= \phi_1^\ddagger (O_{00}\psi_1 + O_{01}v_2) + (O_{00}\psi_1 + O_{01}\psi_2)^\ddagger \phi_1 \\ &\quad + \phi_2^\ddagger (O_{10}\psi_1 + O_{11}\psi_2) + (O_{10}\psi_1 + O_{11}\psi_2)^\ddagger \phi_1 \qquad (98) \\ &= \phi_1^\ddagger O_{00}\psi_1 + \phi_1^\ddagger O_{01}\psi_2 + \psi_1^\ddagger O_{00}^\ddagger \phi_1 + \psi_2^\ddagger O_{01}^\ddagger \phi_1 \\ &\quad + \phi_2^\ddagger O_{10}\psi_1 + \phi_2^\ddagger O_{11}\psi_2 + \psi_1^\ddagger O_{10}^\ddagger \phi_1 + \psi_2^\ddagger O_{11}^\ddagger \phi_1 \qquad (99) \end{aligned}$$

For $\langle \mathbf{O}\boldsymbol{\phi}, \boldsymbol{\psi} \rangle = \langle \boldsymbol{\phi}, \mathbf{O}\boldsymbol{\psi} \rangle$ to be realized, it follows that these relations must hold:

$$O_{00}^\ddagger = O_{00} \tag{100}$$

$$O_{01}^\ddagger = O_{10} \tag{101}$$

$$O_{10}^\ddagger = O_{01} \tag{102}$$

$$O_{11}^\ddagger = O_{11} \tag{103}$$

Therefore, it follows that it must be the case that $\mathbf{O}$ must be equal to its own Clifford transpose. Thus, $\mathbf{O}$ is an observable iff:

$$\mathbf{O}^{\ddagger} = \mathbf{O} \tag{104}$$

which is the equivalent of the self-adjoint operator $\mathbf{O}^{\dagger} = \mathbf{O}$ of complex Hilbert spaces.

5.1.5 Observable, in 2D — Eigenvalues / Spectral Theorem

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

$$\mathbf{O} = \begin{pmatrix} a_{00} & a - xe_1 - ye_2 - be_{12} \\ a + xe_1 + ye_2 + be_{12} & a_{11} \end{pmatrix} \tag{105}$$

In this case, it follows that $\mathbf{O}^{\ddagger} = \mathbf{O}$:

$$\mathbf{O}^{\ddagger} = \begin{pmatrix} a_{00} & a - xe_1 - ye_2 - be_{12} \\ a + xe_1 + ye_2 + be_{12} & a_{11} \end{pmatrix} \tag{106}$$

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

$$0 = \det(\mathbf{O} - \lambda I) = \det \begin{pmatrix} a_{00} - \lambda & a - xe_1 - ye_2 - be_{12} \\ a + xe_1 + ye_2 + be_{12} & a_{11} - \lambda \end{pmatrix} \tag{107}$$

implies:

$$0 = (a_{00} - \lambda)(a_{11} - \lambda) - (a - xe_1 - ye_2 - be_{12})(a + xe_1 + ye_2 + be_{12} + a_{11}) \tag{108}$$

$$0 = (a_{00} - \lambda)(a_{11} - \lambda) - (a^2 - x^2 - y^2 + b^2) \tag{109}$$

finally:

$$\lambda = \{\frac{1}{2}\left(a_{00} + a_{11} - \sqrt{(a_{00} - a_{11})^2 + 4(a^2 - x^2 - y^2 + b^2)}\right), \tag{110}$$

$$\frac{1}{2}\left(a_{00} + a_{11} + \sqrt{(a_{00} - a_{11})^2 + 4(a^2 - x^2 - y^2 + b^2)}\right)\} \tag{111}$$

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 determinant of real matrices must be greater than zero because the exponential maps to the orientation-preserving general linear group— 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, $\mathbf{O}^{\ddagger} = \mathbf{O}$ — implies, for orientation-preserving[1] transformations, that its roots are real-valued, and thus constitute a 'geometric' observable in the traditional sense of an observable whose eigenvalues are real-valued.

[1] 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.

5.2 Algebra of Natural States, in 3D (brief)

The 3D case will be a stepping stone for the 4D case. A general multi-vector of $\mathbb{G}(3,\mathbb{R})$ can be written as follows:

$$\mathbf{u} = A + \mathbf{X} + \mathbf{V} + \mathbf{B} \tag{112}$$

where A is a scalar, $\mathbf{X}$ is a vector, $\mathbf{V}$ is a pseudo-vector and $\mathbf{B}$ is a pseudo-scalar. Such multi-vectors form a complete representation of 2×2 complex matrices:

$$A + X\sigma_1 + Y\sigma_2 + Z\sigma_3 + V_1 i\sigma_1 + V_2 i\sigma_2 + V_3 i\sigma_3 + B\sigma_1 \wedge \sigma_2 \wedge \sigma_3 \tag{113}$$

$$\cong \begin{pmatrix} A + iB + iV_2 + Z & V_1 + iV_3 + X - iY \\ -V_1 + iV_3 + X + iY & A + iB - iV_2 - Z \end{pmatrix} \tag{114}$$

and the determinant of this matrix connects to the determinant of the multi-vector as follows:

$$\begin{array}{rccc} \det \cdot & : & \mathbb{G}(3,\mathbb{R}) & \longrightarrow & \mathbb{C} \\ & & \mathbf{u} & \longmapsto & \mathbf{u}^\ddagger \mathbf{u} \end{array} \tag{115}$$

where $\mathbf{u}^\ddagger$ is the Clifford conjugate in 3D:

$$\mathbf{u}^\ddagger := \langle \mathbf{u} \rangle_0 - \langle \mathbf{u} \rangle_1 - \langle \mathbf{u} \rangle_2 + \langle \mathbf{u} \rangle_3 \tag{116}$$

To produce a real number a further multiplication by its complex conjugate is required:

$$\begin{array}{rccc} |\cdot| & : & \mathbb{G}(3,\mathbb{R}) & \longrightarrow & \mathbb{R} \\ & & \mathbf{u} & \longmapsto & (\mathbf{u}^\ddagger \mathbf{u})^\dagger \mathbf{u}^\ddagger \mathbf{u} \end{array} \tag{117}$$

where $\mathbf{u}^\dagger$ is defined as:

$$\mathbf{u}^\ddagger := \langle \mathbf{u} \rangle_0 + \langle \mathbf{u} \rangle_1 + \langle \mathbf{u} \rangle_2 - \langle \mathbf{u} \rangle_3 \tag{118}$$

5.2.1 Axiomatic Definition of the Algebra, in 3D

Let $\mathbb{V}$ be an m-dimensional vector space over $\mathbb{G}(3,\mathbb{R})$. A subset of vectors in $\mathbb{V}$ forms an algebra of natural states $\mathcal{A}(\mathbb{V})$ iff the following holds:

1. $\forall \boldsymbol{\psi} \in \mathcal{A}(\mathbb{V})$, the quadri-linear map:

$$\begin{array}{rccc} \langle \cdot,\cdot,\cdot,\cdot \rangle & : & \mathbb{V} \times \mathbb{V} \times \mathbb{V} \times \mathbb{V} & \longrightarrow & \mathbb{G}(3,\mathbb{R}) \\ & & \langle \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{x} \rangle & \longmapsto & (\mathbf{u}^\ddagger \mathbf{v})^\dagger \mathbf{w}^\ddagger \mathbf{x} \end{array} \tag{119}$$

   is positive-definite:

$$\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle \in \mathbb{R}_{>0} \tag{120}$$

2. $\forall \boldsymbol{\psi} \in \mathcal{A}(\mathbb{V})$, then for each element $\psi(q) \in \boldsymbol{\psi}$, the function:

$$\rho(\psi(q), \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} (\psi(q)^{\ddagger} \psi(q))^{\dagger} \psi(q)^{\ddagger} \psi(q) \tag{121}$$

is positive-definite:

$$\rho(\psi(q), \boldsymbol{\psi}) \in \mathbb{R}_{>0} \tag{122}$$

5.2.2 Reduction to Complex Hilbert Spaces

We now consider an algebra of natural states that comprises only those multi-vectors of the form $\mathbf{u}' \propto \mathbf{u}^{\ddagger}\mathbf{u}$ (called a sub-algebra, sub-ring, or 'ideal' of the algebra). We also consider, as we obtain an exponential map from our entropy maximization procedure, only multi-vectors which are exponentiated. Then, the algebra reduces to the foundation of quantum mechanics on complex Hilbert spaces (with an extra internal geometric structure). For example, a wave-function would be of this form:

$$\boldsymbol{\psi} = \begin{pmatrix} (\exp \frac{1}{2}\mathbf{u}_1)^{\ddagger} \exp \frac{1}{2}\mathbf{u}_1 \\ \vdots \\ (\exp \frac{1}{2}\mathbf{u}_m)^{\ddagger} \exp \frac{1}{2}\mathbf{u}_m \end{pmatrix} \tag{123}$$

Each element of $\boldsymbol{\psi}$ are of this form:

$$(\exp \frac{1}{2}\mathbf{u})^{\ddagger} \exp \frac{1}{2}\mathbf{u} = \exp \frac{1}{2}(A - \mathbf{X} - \mathbf{V} + \mathbf{B}) \exp \frac{1}{2}(A + \mathbf{X} + \mathbf{V} + \mathbf{B}) \tag{124}$$

$$= \exp(A + \mathbf{B}) \exp \frac{1}{2}(-\mathbf{X} - \mathbf{V}) \exp \frac{1}{2}(\mathbf{X} + \mathbf{V}) \tag{125}$$

$$= \exp(A + \mathbf{B}) \tag{126}$$

Restricting the algebra to such states reduces the quadri-linear map to a bilinear form:

$$\begin{array}{rccc} \langle \cdot, \cdot \rangle & : & \mathcal{A}(\mathbb{V}) \times \mathcal{A}(\mathbb{V}) & \longrightarrow & \mathbb{C} \\ & & \langle \psi, \phi \rangle & \longmapsto & \psi^{\dagger}\phi \end{array} \tag{127}$$

yielding, when applied to said reduced subset of vectors, the same theory as that of quantum mechanics on complex Hilbert space, but with an extra geometric structure for its observables. The 3D case is a stepping stone for the 4D case, where this extra geometric structure will be revealed to be (in the 4D case) the relativistic wave-function given in the form of a spinor field.

5.3 Algebra of Natural States, in 4D

We will now consider the general case for a vector space over 4×4 matrices.

5.3.1 Geometric Representation (in 4D)

The notation will be significantly improved if we use a geometric representation of matrices. Let $\mathbb{G}(4,\mathbb{R})$ be the two-dimensional geometric algebra over the reals. We can write a general multi-vector of $\mathbb{G}(4,\mathbb{R})$ as follows:

$$\mathbf{u} = A + \mathbf{X} + \mathbf{F} + \mathbf{V} + \mathbf{B} \tag{128}$$

where A is a scalar, $\mathbf{X}$ is a vector, $\mathbf{F}$ is a bivector, $\mathbf{V}$ is a pseudo-vector, and $\mathbf{B}$ is a pseudo-scalar. Each multi-vector has a structure-preserving (addition/multiplication) matrix representation. Explicitly, the multi-vectors of $\mathbb{G}(4,\mathbb{R})$ are represented as follows:

Definition 25 (Geometric representation of a matrix (4×4))**.**

$$\begin{aligned}
&A + T\gamma_0 + X\gamma_1 + Y\gamma_2 + Z\gamma_3 \\
&\quad + F_{01}\gamma_0\wedge\gamma_1 + F_{02}\gamma_0\wedge\gamma_2 + F_{03}\gamma_0\wedge\gamma_3 + F_{23}\gamma_2\wedge\gamma_3 + F_{13}\gamma_1\wedge\gamma_3 + F_{12}\gamma_1\wedge\gamma_2 \\
&\quad + V_t\gamma_1\wedge\gamma_2\wedge\gamma_3 + V_x\gamma_0\wedge\gamma_2\wedge\gamma_3 + V_y\gamma_0\wedge\gamma_1\wedge\gamma_3 + V_z\gamma_0\wedge\gamma_1\wedge\gamma_2 \\
&\quad + B\gamma_0\wedge\gamma_1\wedge\gamma_2\wedge\gamma_3 \\
&\qquad \cong \begin{pmatrix}
A + X_0 - iF_{12} - iV_3 & F_{13} - iF_{23} + V_2 - iV_1 & -iB + X_3 + F_{03} - iV_0 & X_1 - iX_2 + F_{01} - iF_{02} \\
-F_{13} - iF_{23} - V_2 - iV_1 & A + X_0 + iF_{12} + iV_3 & X_1 + iX_2 + F_{01} + iF_{02} & -iB - X_3 - F_{03} - iV_0 \\
-iB - X_3 + F_{03} + iV_0 & -X_1 + iX_2 + F_{01} - iF_{02} & A - X_0 - iF_{12} + iV_3 & F_{13} - iF_{23} - V_2 + iV_1 \\
-X_1 - iX_2 + F_{01} + iF_{02} & -iB + X_3 - F_{03} + iV_0 & -F_{13} - iF_{23} + V_2 + iV_1 & A - X_0 + iF_{12} - iV_3
\end{pmatrix}
\end{aligned} \tag{129}$$

And the converse is also true, each 4×4 real matrix is represented as a multi-vector of $\mathbb{G}(4,\mathbb{R})$.

We can define the determinant solely using constructs of geometric algebra[8].

Definition 26 (Clifford conjugate (of a $\mathbb{G}(4,\mathbb{R})$ multi-vector))**.**

$$\mathbf{u}^{\ddagger} := \langle\mathbf{u}\rangle_0 - \langle\mathbf{u}\rangle_1 - \langle\mathbf{u}\rangle_2 + \langle\mathbf{u}\rangle_3 + \langle\mathbf{u}\rangle_4 \tag{130}$$

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):

$$\lfloor\mathbf{u}\rfloor_{\{3,4\}} := \langle\mathbf{u}\rangle_0 + \langle\mathbf{u}\rangle_1 + \langle\mathbf{u}\rangle_2 - \langle\mathbf{u}\rangle_3 - \langle\mathbf{u}\rangle_4 \tag{131}$$

The, the determinant of $\mathbf{u}$ is:

Definition 27 (Geometric representation of the determinant (of a 4×4 matrix))**.**

$$\begin{array}{rccc}
\det & : & \mathbb{G}(4,\mathbb{R}) & \longrightarrow & \mathbb{R} \\
 & & \mathbf{u} & \longmapsto & \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u}
\end{array} \tag{132}$$

5.3.2 Axiomatic Definition of the Algebra, in 4D

Let $\mathbb{V}$ be a m-dimensional vector space over the 4×4 real matrices. A subset of vectors in $\mathbb{V}$ forms an algebra of natural states $\mathcal{A}(\mathbb{V})$ iff the following holds:

1. $\forall \boldsymbol{\psi} \in \mathcal{A}(\mathbb{V})$, the quadri-linear form:

$$\begin{array}{rccc} \langle \cdot, \cdot, \cdot, \cdot \rangle & : & \mathbb{V} \times \mathbb{V} \times \mathbb{V} \times \mathbb{V} & \longrightarrow & \mathbb{G}(4, \mathbb{R}) \\ & & \langle \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{x} \rangle & \longmapsto & \lfloor \mathbf{u}^\ddagger \mathbf{v} \rfloor_{3,4} \mathbf{w}^\ddagger \mathbf{x} \end{array} \tag{133}$$

   is positive-definite:

$$\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle \in \mathbb{R}_{>0} \tag{134}$$

2. $\forall \boldsymbol{\psi} \in \mathcal{A}(\mathbb{V})$, then for each element $\psi(q) \in \boldsymbol{\psi}$, the function:

$$\rho(\psi(q), \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \lfloor \psi(q)^\ddagger \psi(q) \rfloor_{3,4} \psi(q)^\ddagger \psi(q) \tag{135}$$

   is positive-definite:

$$\rho(\psi(q), \boldsymbol{\psi}) \in \mathbb{R}_{>0} \tag{136}$$

We note the following properties, features and comments:

- $\boldsymbol{\psi}$ is called a *natural* (or physical) state.
- $\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle$ is called the *partition function* of $\boldsymbol{\psi}$.
- $\rho(\psi(q), \boldsymbol{\psi})$ is called the *probability measure* (or generalized Born rule) of $\psi(q)$.
- The set of all matrices $\mathbf{T}$ acting on $\boldsymbol{\psi}$ such as $\mathbf{T}\boldsymbol{\psi} \to \boldsymbol{\psi}'$ which leaves the sum of probabilities normalized (invariant):

$$\sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \mathbf{T}\boldsymbol{\psi}) = \sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \boldsymbol{\psi}) = 1 \tag{137}$$

  are the *natural* transformations of $\boldsymbol{\psi}$.

- A matrix $\mathbf{O}$ such that $\forall \mathbf{u} \forall \mathbf{v} \forall \mathbf{w} \forall \mathbf{x} \in \mathbb{V}$:

$$\langle \mathbf{O}\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{x} \rangle = \langle \mathbf{u}, \mathbf{O}\mathbf{v}, \mathbf{w}, \mathbf{x} \rangle = \langle \mathbf{u}, \mathbf{v}, \mathbf{O}\mathbf{w}, \mathbf{x} \rangle = \langle \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{O}\mathbf{x} \rangle \tag{138}$$

  is called an observable.

- The expectation value of an observable $\mathbf{O}$ is:

$$\langle \mathbf{O} \rangle = \frac{\langle \mathbf{O}\boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \tag{139}$$

5.3.3 Reduction to Complex Hilbert Space

Let us select a subset of multi-vectors. The subset will contain all multi-vectors resulting from the multiplication of an even-multi-vector by its own Clifford conjugate. Consistent with our entropy maximization procedure, the elements will also be exponentiated.

$$\boldsymbol{\psi} = \begin{pmatrix} (\exp \frac{1}{2}\mathbf{u}_1)^\ddagger \exp \frac{1}{2}\mathbf{u}_1 \\ \vdots \\ (\exp \frac{1}{2}\mathbf{u}_m)^\ddagger \exp \frac{1}{2}\mathbf{u}_m \end{pmatrix} \tag{140}$$

The form of the elements of $\boldsymbol{\psi}$ is:

$$\psi^\ddagger\psi = \exp \frac{1}{2}(A - \mathbf{F} + \mathbf{B}) \exp \frac{1}{2}(A + \mathbf{F} + \mathbf{B}) \tag{141}$$

$$= \exp \frac{1}{2}A \exp -\frac{1}{2}\mathbf{F} \exp \frac{1}{2}\mathbf{B} \exp \frac{1}{2}A \exp \frac{1}{2}\mathbf{F} \exp \frac{1}{2}\mathbf{B} \tag{142}$$

$$= \exp A \exp \mathbf{B} \tag{143}$$

On such states, the quadri-linear map is reduced to the Born rule (a bilinear map):

$$\begin{array}{rcccc} \langle \cdot, \cdot \rangle & : & \mathcal{A}(\mathbb{V}) \times \mathcal{A}(\mathbb{V}) & \longrightarrow & \mathbb{C} \\ & & \langle \psi, \phi \rangle & \longmapsto & \psi^\dagger \phi \end{array} \tag{144}$$

In our example, and with this bilinear map, $\langle \psi^\ddagger\psi, \psi^\ddagger\psi \rangle = \exp 2A$.

We note the similarity of this sub-algebra to David Hestenes[9]'s geometric algebra formulation of the relativistic wave-function, given as $\psi = \sqrt{\rho}e^{iB/2}e^{\mathbf{F}/2}$. David Hestenes connects his wave-function to a complex number via the reverse $\tilde{\psi} := \sqrt{\rho}e^{iB/2}e^{-\mathbf{F}/2}$, such that $\psi\tilde{\psi} = \rho e^{iB}$.

5.4 Law of Motion

5.4.1 Probability-Preserving Evolution (Left Action in 2D)

A left action on a wave-function : $\mathbf{G}\,|\psi\rangle$, connects to the bilinear form as follows: $\langle\psi|\,\mathbf{G}^\ddagger\mathbf{G}\,|\psi\rangle$. The invariance requirement on $\mathbf{G}$ is as follows:

$$\langle\psi|\,\mathbf{G}^\ddagger\mathbf{G}\,|\psi\rangle = \langle\psi|\psi\rangle \tag{145}$$

We are thus interested in the group of matrices such that:

$$\mathbf{G}^\ddagger\mathbf{G} = I \tag{146}$$

Let us consider a two-state system. A general transformation is:

$$\mathbf{G} = \begin{pmatrix} u & v \\ w & x \end{pmatrix} \tag{147}$$

where u, v, w, x are multi-vectors of 2 dimensions. The expression $\mathbf{G}^\ddagger\mathbf{G}$ is:

$$\mathbf{G}^\ddagger\mathbf{G} = \begin{pmatrix} v^\ddagger & u^\ddagger \\ w^\ddagger & x^\ddagger \end{pmatrix} \begin{pmatrix} v & w \\ u & x \end{pmatrix} = \begin{pmatrix} v^\ddagger v + u^\ddagger u & v^\ddagger w + u^\ddagger x \\ w^\ddagger v + x^\ddagger u & w^\ddagger w + x^\ddagger x \end{pmatrix} \tag{148}$$

For the results to be the identity, it must be the case that:

$$v^\ddagger v + u^\ddagger u = 1 \tag{149}$$
$$v^\ddagger w + u^\ddagger x = 0 \tag{150}$$
$$w^\ddagger v + x^\ddagger u = 0 \tag{151}$$
$$w^\ddagger w + x^\ddagger x = 1 \tag{152}$$

This is the case if

$$\mathbf{G} = \frac{1}{\sqrt{v^\ddagger v + u^\ddagger u}} \begin{pmatrix} v & u \\ -e^\varphi u^\ddagger & e^\varphi v^\ddagger \end{pmatrix} \tag{153}$$

where u, v are multi-vectors of 2 dimensions, and where e^φ is a unit multi-vector. Comparatively, the unitary case is obtained with $\mathbf{X} \to 0$, and is:

$$\mathbf{U} = \frac{1}{\sqrt{|a|^2 + |b|^2}} \begin{pmatrix} a & b \\ -e^{i\theta} b^\dagger & e^{i\theta} a^\dagger \end{pmatrix} \tag{154}$$

We can show that $\mathbf{G}^\ddagger\mathbf{G} = I$ as follows:

$$\implies \mathbf{G}^\ddagger\mathbf{G} = \frac{1}{v^\ddagger v + u^\ddagger u} \begin{pmatrix} v^\ddagger & -e^{-\varphi} u \\ u^\ddagger & e^{-\varphi} v \end{pmatrix} \begin{pmatrix} v & u \\ -e^\varphi u^\ddagger & e^\varphi v^\ddagger \end{pmatrix} \tag{155}$$
$$= \frac{1}{v^\ddagger v + u^\ddagger u} \begin{pmatrix} v^\ddagger v + u^\ddagger u & v^\ddagger u - v^\ddagger u \\ u^\ddagger v - u^\ddagger v & u^\ddagger u + v^\ddagger v \end{pmatrix} \tag{156}$$
$$= I \tag{157}$$

In the case where $\mathbf{G}$ and $|\psi\rangle$ are n-dimensional, we can find an expression for it starting from a diagonal matrix:

$$\mathbf{D} = \begin{pmatrix} e^{x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1} & 0 \\ 0 & e^{x_2\hat{\mathbf{x}}+y_2\hat{\mathbf{y}}+ib_2} \end{pmatrix} \tag{158}$$

where $\mathbf{G} = P\mathbf{D}P^{-1}$. It follows quite easily that $D^\ddagger D = I$, because each diagonal entry produces unity: $e^{-x_1\hat{\mathbf{x}}-y_1\hat{\mathbf{y}}-ib_1}e^{x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1} = 1$. Then,

5.4.2 Probability-Preserving Evolution (Adjoint Action in 2D)

Since the elements of $|\psi\rangle$ are matrices, in the general case, the transformation is given by adjoint action:

$$\mathbf{G}\,|\psi\rangle\,\mathbf{G}^{-1} \tag{159}$$

The bilinear form is:

$$(\mathbf{G}\,|\psi\rangle\,\mathbf{G}^{-1})^\ddagger(\mathbf{G}\,|\psi\rangle\,\mathbf{G}^{-1}) = (\mathbf{G}^{-1})^\ddagger\,\langle\psi|\,\mathbf{G}^\ddagger\mathbf{G}\,|\psi\rangle\,\mathbf{G}^{-1} \tag{160}$$

and the invariance requirement on $\mathbf{G}$ is as follows:

$$(\mathbf{G}^{-1})^\ddagger\,\langle\psi|\,\mathbf{G}^\ddagger\mathbf{G}\,|\psi\rangle\,\mathbf{G}^{-1} = \langle\psi|\psi\rangle \tag{161}$$

With a diagonal matrix, this occurs for general linear transformations:

$$\mathbf{D} = \begin{pmatrix} e^{a_1+x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1} & 0 & 0 \\ 0 & e^{a_2+x_2\hat{\mathbf{x}}+y_2\hat{\mathbf{y}}+ib_2} & 0 \\ 0 & 0 & \ddots \end{pmatrix} \tag{162}$$

where $\mathbf{G} = P\mathbf{D}P^{-1}$.
Taking a single diagonal entry as an example, the reduction is:

$$e^{-a_1+x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1}\psi_1^\ddagger e^{a_1-x_1\hat{\mathbf{x}}-y_1\hat{\mathbf{y}}-ib_1}e^{a_1+x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1}\psi_1 e^{-a_1-x_1\hat{\mathbf{x}}-y_1\hat{\mathbf{y}}-ib_1} \tag{163}$$

$$= e^{-a_1+x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1}\psi_1^\ddagger e^{2a_1}\psi_1 e^{-a_1-x_1\hat{\mathbf{x}}-y_1\hat{\mathbf{y}}-ib_1} \tag{164}$$

We note that $\psi^\ddagger\psi$ is a scalar, therefore

$$= \psi_1^\ddagger\psi_1 e^{2a_1}e^{-a_1+x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1}e^{-a_1-x_1\hat{\mathbf{x}}-y_1\hat{\mathbf{y}}-ib_1} \tag{165}$$

$$= \psi_1^\ddagger\psi_1 e^{2a_1}e^{-a_1}e^{-a_1} = \psi_1^\ddagger\psi_1 \tag{166}$$

5.4.3 General Linear Schrödinger Equation (Left Action)

The standard Schrödinger equation can be derived as follows. First, assume $U(t) = e^{-itH}$, and its Taylor expansion to the first linear term: $U(\delta t) \approx 1 - i\delta t H$. Then:

$$\left|\psi(t+\delta t)\right\rangle = U(\delta t)\left|\psi(t)\right\rangle \approx (1 - i\delta t H)\left|\psi\right\rangle \tag{167}$$

$$\implies \left|\psi(t+\delta t)\right\rangle - \left|\psi\right\rangle \approx -i\delta t H\left|\psi\right\rangle \tag{168}$$

$$\implies i\frac{\left|\psi(t+\delta t)\right\rangle - \left|\psi\right\rangle}{\delta t} \approx H\left|\psi\right\rangle \tag{169}$$

$$\implies i\frac{\partial\left|\psi(t)\right\rangle}{\partial t} = H\left|\psi\right\rangle \tag{170}$$

Now, we wish to use the same derivation, but apply it to the 2D general linear version of the unitary group:

$$U^\dagger U = I \to (\mathbf{G}^{-1})^\ddagger \left\langle\psi\right| \mathbf{G}^\ddagger \mathbf{G}\left|\psi\right\rangle \mathbf{G}^{-1} = \left\langle\psi|\psi\right\rangle \tag{171}$$

In the general linear case, the imaginary number i is replaced with an arbitrary matrix $\mathbf{M}$, via the relation:

$$\mathbf{G} = e^{-\mathbf{M}\tau H} \tag{172}$$

where H is self-adjoint : $H^\ddagger = H$.

Then, the general linear Schrödinger equation for the one-parameter group of the general linear group $\mathbf{G}(\tau) = e^{-\mathbf{M}\tau H}$, for the left action is:

$$\left|\psi(\tau+\delta\tau)\right\rangle = G(\delta\tau)\left|\psi(\tau)\right\rangle \tag{173}$$

$$\approx (1 - \mathbf{M}\delta\tau H)\left|\psi(\tau)\right\rangle \tag{174}$$

$$= \left|\psi(\tau)\right\rangle - \mathbf{M}\delta\tau H\left|\psi(\tau)\right\rangle \tag{175}$$

$$\implies -\frac{\partial\left|\psi(\tau)\right\rangle}{\partial\tau} = \mathbf{M}H\left|\psi(\tau)\right\rangle \tag{176}$$

and iff $\exists\mathbf{M}^{-1}$, then

$$-\mathbf{M}^{-1}\frac{\partial\left|\psi(\tau)\right\rangle}{\partial\tau} = H\left|\psi(\tau)\right\rangle \tag{177}$$

5.4.4 General Linear Schrödinger Equation (Adjoint Action)

And for the adjoint action, it is:

$$\left|\psi(\tau+\delta\tau)\right\rangle = G(\delta\tau)\left|\psi(\tau)\right\rangle G(\delta\tau)^{-1} \tag{178}$$
$$\approx (1-\mathbf{M}\delta\tau H)\left|\psi(\tau)\right\rangle (1+\mathbf{M}\delta\tau H) \tag{179}$$
$$= \left|\psi(\tau)\right\rangle (1+\mathbf{M}\delta\tau H) - \mathbf{M}\delta\tau H\left|\psi(\tau)\right\rangle (1+\mathbf{M}\delta\tau H) \tag{180}$$
$$= \left|\psi(\tau)\right\rangle + \left|\psi(\tau)\right\rangle \mathbf{M}\delta\tau H - \mathbf{M}\delta\tau H\left|\psi(\tau)\right\rangle - \mathbf{M}\delta\tau H\left|\psi(\tau)\right\rangle \mathbf{M}\delta\tau H \tag{181}$$
$$\approx \left|\psi(\tau)\right\rangle + \left|\psi(\tau)\right\rangle \mathbf{M}\delta\tau H - \mathbf{M}\delta\tau H\left|\psi(\tau)\right\rangle \tag{182}$$
$$= \left|\psi(\tau)\right\rangle + \delta\tau\left[\mathbf{M}H, \left|\psi(\tau)\right\rangle\right] \tag{183}$$
$$\implies \frac{\partial\left|\psi(\tau)\right\rangle}{\partial\tau} = \left[\mathbf{M}H, \left|\psi(\tau)\right\rangle\right] \tag{184}$$

5.4.5 Conservation of Probability (Left Action in 2D)

For a parametrization of ψ, the probability must normalize. For instance, a x parametrization would yield:

$$\int \psi(\tau,x)^{\ddagger}\psi(\tau,x)\,\mathrm{d}x = N(\tau) \tag{185}$$

To lighten the notation we will not explicitly write the dependance of ψ in (τ, x).

$$\frac{dN(\tau)}{d\tau} = 0 = \int \frac{\partial\psi^{\ddagger}\psi}{\partial\tau}\,\mathrm{d}x \tag{186}$$
$$= \int \frac{\partial\psi^{\ddagger}}{\partial\tau}\psi\,\mathrm{d}x + \int \psi^{\ddagger}\frac{\partial\psi}{\partial\tau}\,\mathrm{d}x \tag{187}$$

We now inject the following relation (derived from the general linear Schrödinger equation):

$$\frac{\partial\psi}{\partial\tau} = \mathbf{M}H\psi \tag{188}$$
$$\frac{\partial\psi^{\ddagger}}{\partial\tau} = (\mathbf{M}H\psi)^{\ddagger} = \psi^{\ddagger}H^{\ddagger}\mathbf{M}^{\ddagger} \tag{189}$$

Injecting them, we get:

$$\frac{dN(\tau)}{d\tau} = \int \frac{\partial\psi^{\ddagger}}{\partial\tau}\psi\,\mathrm{d}x + \int \psi^{\ddagger}\frac{\partial\psi}{\partial\tau}\,\mathrm{d}x \tag{190}$$
$$= \int \psi^{\ddagger}\mathbf{M}^{\ddagger}H^{\ddagger}\psi\,\mathrm{d}x + \int \psi^{\ddagger}\mathbf{M}H\psi\,\mathrm{d}x \tag{191}$$

We note that $\mathbf{M}^\ddagger = -\mathbf{M}$ (this requires that $\langle \mathbf{M} \rangle_0 = 0$, which is the case for the left action) and that $H^\ddagger = H$ (which it is iff it is self-adjoint), therefore:

$$= -\int \psi^\ddagger \mathbf{M} H \psi \, \mathrm{d}x + \int \psi^\ddagger \mathbf{M} H \psi \, \mathrm{d}x \tag{192}$$

Finally, adding the requirement that $[\mathbf{M}, H] = 0$, we get the conservation of probability:

$$= 0 \tag{193}$$

The general linear form of the Schrödinger equation is a conservation of probability law of the general linear case.

5.4.6 Conservation of Probability (Adjoint Action in 2D)

For a parametrization of ψ, the probability must normalize. For instance, a x parametrization would yield:

$$\int \psi(\tau, x)^\ddagger \psi(\tau, x) \, \mathrm{d}x = N(\tau) \tag{194}$$

To lighten the notation we will not explicitly write the dependance of ψ in (τ, x).

$$\frac{dN(\tau)}{d\tau} = 0 = \int \frac{\partial \psi^\ddagger \psi}{\partial \tau} \, \mathrm{d}x \tag{195}$$

$$= \int \frac{\partial \psi^\ddagger}{\partial \tau} \psi \, \mathrm{d}x + \int \psi^\ddagger \frac{\partial \psi}{\partial \tau} \, \mathrm{d}x \tag{196}$$

We now inject the following relation (derived from the general linear Schrödinger equation):

$$\frac{\partial \psi}{\partial \tau} = \mathbf{M} H \psi - \psi \mathbf{M} H \tag{197}$$

$$\frac{\partial \psi^\ddagger}{\partial \tau} = (\mathbf{M} H \psi - \psi \mathbf{M} H)^\ddagger = \psi^\ddagger H^\ddagger \mathbf{M}^\ddagger - H^\ddagger \mathbf{M}^\ddagger \psi^\ddagger \tag{198}$$

Injecting them, we get:

$$\frac{dN(\tau)}{d\tau} = \int \frac{\partial \psi^\ddagger}{\partial \tau} \psi \, \mathrm{d}x + \int \psi^\ddagger \frac{\partial \psi}{\partial \tau} \, \mathrm{d}x \tag{199}$$

$$= \int \psi^\ddagger H^\ddagger \mathbf{M}^\ddagger \psi - H^\ddagger \mathbf{M}^\ddagger \psi^\ddagger \psi \, \mathrm{d}x + \int \psi^\ddagger \mathbf{M} H \psi - \psi^\ddagger \psi \mathbf{M} H \, \mathrm{d}x \tag{200}$$

$$= \int \psi^\ddagger H^\ddagger \mathbf{M}^\ddagger \psi - H^\ddagger \mathbf{M}^\ddagger \psi^\ddagger \psi + \psi^\ddagger \mathbf{M} H \psi - \psi^\ddagger \psi \mathbf{M} H \, \mathrm{d}x \tag{201}$$

We note that $\mathbf{M}^\ddagger = A - \mathbf{m}$ and that $H^\ddagger = H$, therefore (we also pose $\mathbf{M} = A + \mathbf{m}$):

$$= \int \psi^\ddagger H(A - \mathbf{m})\psi - H(A - \mathbf{m})\psi^\ddagger\psi + \psi^\ddagger(A + \mathbf{m})H\psi - \psi^\ddagger\psi(A + \mathbf{m})H \,\mathrm{d}x \tag{202}$$

$$= \int -\psi^\ddagger H\mathbf{m}\psi + H\mathbf{m}\psi^\ddagger\psi + \psi^\ddagger\mathbf{m}H\psi - \psi^\ddagger\psi\mathbf{m}H \,\mathrm{d}x \tag{203}$$

Finally, adding the requirement that $[\mathbf{m}, H] = 0$, we get the conservation of probability:

$$= 0 \tag{204}$$

The general linear form of the Schrödinger equation is a conservation of probability law of the general linear case.

5.5 Gravity as a Gauge Theory

5.5.1 Unitary Gauge (Recap)

The typical gauge theory in quantum electrodynamics is obtained by the production of a gauge covariant derivative over a $U(1)$ invariance associated with the use of the complex norm in any probability measure of quantum mechanics. Localizing the invariance group $\theta \to \theta(x)$ yields the corresponding covariant derivative:

$$D_\mu = \partial_\mu + iqA_\mu(x) \tag{205}$$

Where $A_\mu(x)$ is the gauge field. The $U(1)$ invariance results from the usage of the complex norm to construct a probability measure in a quantum theory, and the presence of the derivative is the result of constructing said probability measure as the Lagrangian of a Dirac field. If one then applies a gauge transformation to ψ and A_μ:

$$\psi \to e^{-iq\theta(x)}\psi \quad \text{and} \quad A_\mu \to A_\mu + \partial_\mu\theta(x) \tag{206}$$

Then, applies the covariant derivation, one gets:

$$D_\mu\psi = \partial_\mu\psi + iqA_\mu\psi \tag{207}$$

$$\to \partial_\mu(e^{-iq\theta(x)}\psi) + iq(A_\mu + \partial_\mu\theta(x))(e^{-iq\theta(x)}\psi) \tag{208}$$

$$= e^{-iq\theta(x)}D_\mu\psi \tag{209}$$

Finally, the field is given as follows:

$$F_{\mu\nu} = [\mathcal{D}_\mu, \mathcal{D}_\nu] \tag{210}$$

where $\mathcal{D}_\mu$ is the covariant derivative with respect to the potential one-form $A_\mu = A_\mu^\alpha T_\alpha$, and where T_α are the generators of the lie algebra of $U(1)$.

5.5.2 General Linear Gauge

The fundamental invariance group of our measure is the orientation-preserving general linear group $\mathrm{GL}^+(n,\mathbb{R})$, if the algebra is even, or the complex general linear group $\mathrm{GL}(n,\mathbb{C})$ if the algebra is odd, rather than $U(1)$. Gauging the $\mathrm{GL}^+(n,\mathbb{R})$ group is known to substantially connect to general relativity, as the resulting $\mathrm{GL}(4,\mathbb{R})$-valued field can be viewed as the Christoffel symbols Γ^μ.

A general linear transformation of ψ:

$$\psi'(x) \to g\psi(x)g^{-1} \tag{211}$$

leaves the probability measure invariant.

The gauge-covariant derivative is:

$$D_\mu\psi = \partial_\mu\psi - [iqA_\mu, \psi] \tag{212}$$

Finally, the field is given as follows:

$$R_{\mu\nu} = [D_\mu, D_\nu] \tag{213}$$

where $R_{\mu\nu}$ is the Riemann tensor.

Since this is the result of the GL gauge invariance, then gravity is fundamentally integrated throughout the present quantum mechanical framework because of GL invariance, for the same reason that electromagnetism is fundamentally integrated within quantum theory over the complex norm because of U(1) invariance.

6 Testable Prediction

Certain linear transformations of the wave-function, under the general linear group and its subgroups, would produce richer interference patterns that what is possible merely with complex interference. The possibility of richer interference patterns has been proposed before; specifically, I note the work of B. I. Lev.[10] which suggests (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.

We note that interference experiments have paid off substantial dividends in the history of physics and are somewhat easy to construct and more affordable that many alternative experiments.

6.1 Geometric Interference

Let us start by introducing a notation for a dot product, then we will list the various possible interference patterns.

6.1.1 Geometric Algebra Dot Product

Let us introduce a notation. We will define a bilinear form using the dot product notation, as follows:

$$\begin{array}{rccc} \cdot \;:\; & \mathbb{G}(2n,\mathbb{R}) \times \mathbb{G}(2n,\mathbb{R}) & \longrightarrow & \mathbb{R} \\ & \mathbf{u} \cdot \mathbf{v} & \longmapsto & \frac{1}{2}(\det(\mathbf{u}+\mathbf{v}) - \det \mathbf{u} - \det \mathbf{v}) \end{array} \tag{214}$$

For example,

$$\mathbf{u} = A_1 + X_1 e_1 + Y_1 e_2 + B_1 e_{12} \tag{215}$$

$$\mathbf{v} = A_2 + X_2 e_1 + Y_2 e_2 + B_2 e_{12} \tag{216}$$

$$\implies \mathbf{u} \cdot \mathbf{v} = A_1 A_2 + B_1 B_2 - X_1 X_2 - Y_1 Y_2 \tag{217}$$

Iff $\det \mathbf{u} > 0$ and $\det \mathbf{v} > 0$ then $\mathbf{u} \cdot \mathbf{v}$ is always positive, and therefore qualifies as a positive inner product (over the positive det group), but no greater than either $\det \mathbf{u}$ or $\det \mathbf{v}$, whichever is larger. This definition of the dot product extends to multi-vectors of 4 dimensions.

2D: In 2D the dot product is equivalent to this form:

$$\frac{1}{2}(\det(\mathbf{u}+\mathbf{v}) - \det \mathbf{u} - \det \mathbf{v}) = \frac{1}{2}\left((\mathbf{u}+\mathbf{v})^\ddagger(\mathbf{u}+\mathbf{v}) - \mathbf{u}^\ddagger\mathbf{u} - \mathbf{v}^\ddagger\mathbf{v}\right) \tag{218}$$

$$= \mathbf{u}^\ddagger\mathbf{u} + \mathbf{u}^\ddagger\mathbf{v} + \mathbf{v}^\ddagger\mathbf{u} + \mathbf{v}^\ddagger\mathbf{v} - \mathbf{u}^\ddagger\mathbf{u} - \mathbf{v}^\ddagger\mathbf{v} \tag{219}$$

$$= \mathbf{u}^\ddagger\mathbf{v} + \mathbf{v}^\ddagger\mathbf{u} \tag{220}$$

4D: In 4D it is substantially more verbose:

$$\frac{1}{2}(\det(\mathbf{u}+\mathbf{v}) - \det\mathbf{u} - \det\mathbf{v}) \tag{221}$$

$$= \frac{1}{2}\left(\lfloor(\mathbf{u}+\mathbf{v})^\ddagger(\mathbf{u}+\mathbf{v})\rfloor_{3,4}(\mathbf{u}+\mathbf{v})^\ddagger(\mathbf{u}+\mathbf{v}) - \lfloor\mathbf{u}^\ddagger\mathbf{u}\rfloor_{3,4}\mathbf{u}^\ddagger\mathbf{u} - \lfloor\mathbf{v}^\ddagger\mathbf{v}\rfloor_{3,4}\mathbf{v}^\ddagger\mathbf{v}\right) \tag{222}$$

$$= \frac{1}{2}\left(\lfloor\mathbf{u}^\ddagger\mathbf{u} + \mathbf{u}^\ddagger\mathbf{v} + \mathbf{v}^\ddagger\mathbf{u} + \mathbf{v}^\ddagger\mathbf{v}\rfloor_{3,4}(\mathbf{u}^\ddagger\mathbf{u} + \mathbf{u}^\ddagger\mathbf{v} + \mathbf{v}^\ddagger\mathbf{u} + \mathbf{v}^\ddagger\mathbf{v}) - \dots\right) \tag{223}$$

$$\begin{aligned}= {} & \lfloor\mathbf{u}^\ddagger\mathbf{u}\rfloor_{3,4}\mathbf{u}^\ddagger\mathbf{u} + \lfloor\mathbf{u}^\ddagger\mathbf{u}\rfloor_{3,4}\mathbf{u}^\ddagger\mathbf{v} + \lfloor\mathbf{u}^\ddagger\mathbf{u}\rfloor_{3,4}\mathbf{v}^\ddagger\mathbf{u} + \lfloor\mathbf{u}^\ddagger\mathbf{u}\rfloor_{3,4}\mathbf{v}^\ddagger\mathbf{v} \\ & + \lfloor\mathbf{u}^\ddagger\mathbf{v}\rfloor_{3,4}\mathbf{u}^\ddagger\mathbf{u} + \lfloor\mathbf{u}^\ddagger\mathbf{v}\rfloor_{3,4}\mathbf{u}^\ddagger\mathbf{v} + \lfloor\mathbf{u}^\ddagger\mathbf{v}\rfloor_{3,4}\mathbf{v}^\ddagger\mathbf{u} + \lfloor\mathbf{u}^\ddagger\mathbf{v}\rfloor_{3,4}\mathbf{v}^\ddagger\mathbf{v} \\ & + \lfloor\mathbf{v}^\ddagger\mathbf{u}\rfloor_{3,4}\mathbf{u}^\ddagger\mathbf{u} + \lfloor\mathbf{v}^\ddagger\mathbf{u}\rfloor_{3,4}\mathbf{u}^\ddagger\mathbf{v} + \lfloor\mathbf{v}^\ddagger\mathbf{u}\rfloor_{3,4}\mathbf{v}^\ddagger\mathbf{u} + \lfloor\mathbf{v}^\ddagger\mathbf{u}\rfloor_{3,4}\mathbf{v}^\ddagger\mathbf{v} \\ & + \lfloor\mathbf{v}^\ddagger\mathbf{v}\rfloor_{3,4}\mathbf{u}^\ddagger\mathbf{u} + \lfloor\mathbf{v}^\ddagger\mathbf{v}\rfloor_{3,4}\mathbf{u}^\ddagger\mathbf{v} + \lfloor\mathbf{v}^\ddagger\mathbf{v}\rfloor_{3,4}\mathbf{v}^\ddagger\mathbf{u} + \lfloor\mathbf{v}^\ddagger\mathbf{v}\rfloor_{3,4}\mathbf{v}^\ddagger\mathbf{v} - \dots\end{aligned} \tag{224}$$

$$\begin{aligned}= {} & \lfloor\mathbf{u}^\ddagger\mathbf{u}\rfloor_{3,4}\mathbf{u}^\ddagger\mathbf{v} + \lfloor\mathbf{u}^\ddagger\mathbf{u}\rfloor_{3,4}\mathbf{v}^\ddagger\mathbf{u} + \lfloor\mathbf{u}^\ddagger\mathbf{u}\rfloor_{3,4}\mathbf{v}^\ddagger\mathbf{v} \\ & + \lfloor\mathbf{u}^\ddagger\mathbf{v}\rfloor_{3,4}\mathbf{u}^\ddagger\mathbf{u} + \lfloor\mathbf{u}^\ddagger\mathbf{v}\rfloor_{3,4}\mathbf{u}^\ddagger\mathbf{v} + \lfloor\mathbf{u}^\ddagger\mathbf{v}\rfloor_{3,4}\mathbf{v}^\ddagger\mathbf{u} + \lfloor\mathbf{u}^\ddagger\mathbf{v}\rfloor_{3,4}\mathbf{v}^\ddagger\mathbf{v} \\ & + \lfloor\mathbf{v}^\ddagger\mathbf{u}\rfloor_{3,4}\mathbf{u}^\ddagger\mathbf{u} + \lfloor\mathbf{v}^\ddagger\mathbf{u}\rfloor_{3,4}\mathbf{u}^\ddagger\mathbf{v} + \lfloor\mathbf{v}^\ddagger\mathbf{u}\rfloor_{3,4}\mathbf{v}^\ddagger\mathbf{u} + \lfloor\mathbf{v}^\ddagger\mathbf{u}\rfloor_{3,4}\mathbf{v}^\ddagger\mathbf{v} \\ & + \lfloor\mathbf{v}^\ddagger\mathbf{v}\rfloor_{3,4}\mathbf{u}^\ddagger\mathbf{u} + \lfloor\mathbf{v}^\ddagger\mathbf{v}\rfloor_{3,4}\mathbf{u}^\ddagger\mathbf{v} + \lfloor\mathbf{v}^\ddagger\mathbf{v}\rfloor_{3,4}\mathbf{v}^\ddagger\mathbf{u}\end{aligned} \tag{225}$$

6.1.2 Geometric Interference (General Form)

A multi-vector can be written as $\mathbf{u} = a + \mathbf{s}$, where a is a scalar and $\mathbf{s}$ is the multi-vectorial part. In general, the exponential $\exp\mathbf{u}$ equals $\exp a \exp\mathbf{s}$ because a commutes with $\mathbf{s}$.

One can thus write a general two-state system as follows:

$$\psi = \psi_1 + \psi_2 = e^{\frac{2}{n}A_1}e^{\frac{2}{n}\mathbf{S}_1} + e^{\frac{2}{n}A_2}e^{\frac{2}{n}\mathbf{S}_2} \tag{226}$$

$$\tag{227}$$

The general interference pattern will be of the following form:

$$\det\psi_1 + \psi_2 = \det\psi_1 + \det\psi_2 + \psi_1\cdot\psi_2 \tag{228}$$

$$= e^{2A_1} + e^{2A_2} + \psi_1\cdot\psi_2 \tag{229}$$

where $\det\psi_1 + \det\psi_2$ is a sum of probabilities and where $\psi_1\cdot\psi_2$ is the interference pattern.

6.1.3 Complex Interference (Recall)

Consider a two-state wave-function:

$$\psi = \psi_1 + \psi_2 = e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{\mathbf{B}_2} \tag{230}$$

The interference pattern familiar to quantum mechanics is the result of the complex norm:

$$\psi^\dagger\psi = \psi_1^\dagger\psi_1 + \psi_2^\dagger\psi_2 + \psi_1^\dagger\psi_2 + \psi_2^\dagger\psi_1 \tag{231}$$

$$= e^{A_1}e^{-\mathbf{B}_1}e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{-\mathbf{B}_2}e^{A_2}e^{\mathbf{B}_2} + e^{A_1}e^{-\mathbf{B}_1}e^{A_2}e^{\mathbf{B}_2} + e^{A_2}e^{-\mathbf{B}_2}e^{A_1}e^{\mathbf{B}_1} \tag{232}$$

$$= e^{2A_1} + e^{2A_2} + e^{A_1+A_2}(e^{-\mathbf{B}_1+\mathbf{B}_2} + e^{-(-\mathbf{B}_1+\mathbf{B}_2)}) \tag{233}$$

$$= \underbrace{e^{2A_1} + e^{2A_2}}_{\text{sum}} + \underbrace{2e^{A_1+A_2}\cos(B_1 - B_2)}_{\text{interference}} \tag{234}$$

6.1.4 Geometric Interference in 2D

Consider a two-state wave-function:

$$\psi = \psi_1 + \psi_2 = e^{A_1}e^{\mathbf{X}_1+\mathbf{B}_1} + e^{A_2}e^{\mathbf{X}_2+\mathbf{B}_2} \tag{235}$$

To lighten the notation we will write it as follows:

$$\psi = \psi_1 + \psi_2 = e^{A_1}e^{\mathbf{S}_1} + e^{A_2}e^{\mathbf{S}_2} \tag{236}$$

where

$$\mathbf{S} = \mathbf{X} + \mathbf{B} \tag{237}$$

The interference pattern for a full general linear transformation on a two-state wave-function in 2D is:

$$\psi^\dagger\psi = \psi_1^\dagger\psi_1 + \psi_2^\dagger\psi_2 + \psi_1^\dagger\psi_2 + \psi_2^\dagger\psi_1 \tag{238}$$

$$= e^{A_1}(e^{\mathbf{S}_1})^\ddagger e^{A_1}e^{\mathbf{S}_1} + e^{A_2}(e^{\mathbf{S}_2})^\ddagger e^{A_2}e^{\mathbf{S}_2} + e^{A_1}(e^{\mathbf{S}_1})^\ddagger e^{A_2}e^{\mathbf{S}_2} + e^{A_2}(e^{\mathbf{S}_2})^\ddagger e^{A_1}e^{\mathbf{S}_1} \tag{239}$$

$$= e^{2A_1} + e^{2A_2} + e^{A_1+A_2}((e^{\mathbf{S}_1})^\ddagger e^{\mathbf{S}_2} + (e^{\mathbf{S}_2})^\ddagger e^{\mathbf{S}_1}) \tag{240}$$

$$= \underbrace{e^{2A_1} + e^{2A_2}}_{\text{sum}} + \underbrace{e^{A_1+A_2}(e^{-\mathbf{X}_1-\mathbf{B}_1}e^{\mathbf{X}_2+\mathbf{B}_2} + e^{-\mathbf{X}_2-\mathbf{B}_2}e^{\mathbf{X}_1+\mathbf{B}_1})}_{\text{interference}} \tag{241}$$

6.1.5 Geometric Interference in 4D

Consider a two-state wave-function:

$$\psi = \psi_1 + \psi_2 = e^{\frac{1}{2}A_1}e^{\frac{1}{2}(\mathbf{X}_1+\mathbf{F}_1+\mathbf{V}_1+\mathbf{B}_1)} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}(\mathbf{X}_2+\mathbf{F}_2+\mathbf{V}_2+\mathbf{B}_2)} \tag{242}$$

To lighten the notation we will write it as follows:

$$\psi = \psi_1 + \psi_2 = e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{S}_1} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{S}_2} \tag{243}$$

where

$$\mathbf{S} = \mathbf{X} + \mathbf{F} + \mathbf{V} + \mathbf{B} \tag{244}$$

The geometric interference patterns for a full general linear transformation in 4D is given by the product:

$$\lfloor \psi^{\ddagger} \psi \rfloor_{3,4} \psi^{\ddagger} \psi = \lfloor \psi_1^{\ddagger} \psi_1 \rfloor_{3,4} \psi_1^{\ddagger} \psi_1 + \lfloor \psi_2^{\ddagger} \psi_2 \rfloor_{3,4} \psi_2^{\ddagger} \psi_2 + \psi_1 \cdot \psi_2 \tag{245}$$

$$= e^{2A_1} + e^{2A_2} + \left(e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{S}_1} \right) \cdot \left(e^{\frac{1}{2}A_2} e^{\frac{1}{2}\mathbf{S}_2} \right) \tag{246}$$

In many cases of interest, the pattern simplifies.

6.1.6 Geometric Interference in 4D (Shallow Phase Rotation)

If we consider a sub-algebra in 4D comprised of even-multi-vector products $\psi^{\ddagger}\psi$, then a two-state system is given as:

$$\psi = \psi_1 + \psi_2 \tag{247}$$

where

$$\psi_1 = (e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1})^{\ddagger} (e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1}) = e^{A_1} e^{\mathbf{B}_1} \tag{248}$$

$$\psi_2 = (e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1})^{\ddagger} (e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1}) = e^{A_2} e^{\mathbf{B}_2} \tag{249}$$

Thus

$$\psi = e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{B}_2} \tag{250}$$

The quadri-linear map becomes a bilinear map:

$$\psi^{\dagger} \psi = (e^{A_1} e^{-\mathbf{B}_1} + e^{A_2} e^{-\mathbf{B}_2})(e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{B}_2}) \tag{251}$$

$$= e^{A_1} e^{-\mathbf{B}_1} e^{A_1} e^{\mathbf{B}_1} + e^{A_1} e^{-\mathbf{B}_1} e^{A_2} e^{\mathbf{B}_2} + e^{A_2} e^{-\mathbf{B}_2} e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{-\mathbf{B}_2} e^{A_2} e^{\mathbf{B}_2} \tag{252}$$

$$= \underbrace{e^{2A_1} + e^{2A_2}}_{\text{sum}} + \underbrace{2 e^{A_1 + A_2} \cos(B_1 - B_2)}_{\text{complex interference}} \tag{253}$$

6.1.7 Geometric Interference in 4D (Deep Phase Rotation)

A phase rotation on the base algebra (rather than the sub-algebra) produces a difference interference pattern. Consider a two-state wave-function:

$$\psi = \psi_1 + \psi_2 = e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2} e^{\frac{1}{2}\mathbf{B}_2} \tag{254}$$

The sub-product part is:

$$\psi^{\ddagger}\psi = (e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2})(e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2}) \tag{255}$$

$$= e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1}e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1}e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2}e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2}e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2} \tag{256}$$

$$= e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{\mathbf{B}_2} + 2e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} \tag{257}$$

The final product is:

$$\begin{aligned}\lfloor\psi^{\ddagger}\psi\rfloor_{3,4}\psi^{\ddagger}\psi &= (e^{A_1}e^{-\mathbf{B}_1} + e^{A_2}e^{-\mathbf{B}_2} + 2e^{\frac{1}{2}(A_1+A_2)}e^{-\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}) \\ &\quad\times (e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{\mathbf{B}_2} + 2e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}) \qquad (258)\\ &= e^{A_1}e^{-\mathbf{B}_1}e^{A_1}e^{\mathbf{B}_1} + e^{A_1}e^{-\mathbf{B}_1}e^{A_2}e^{\mathbf{B}_2} + e^{A_1}e^{-\mathbf{B}_1}2e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} \\ &\quad + e^{A_2}e^{-\mathbf{B}_2}e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{-\mathbf{B}_2}e^{A_2}e^{\mathbf{B}_2} + e^{A_2}e^{-\mathbf{B}_2}2e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} \\ &\quad + 2e^{\frac{1}{2}(A_1+A_2)}e^{-\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}e^{A_1}e^{\mathbf{B}_1} \\ &\quad\quad + 2e^{\frac{1}{2}(A_1+A_2)}e^{-\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}e^{A_2}e^{\mathbf{B}_2} \\ &\quad\quad + 2e^{\frac{1}{2}(A_1+A_2)}e^{-\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}2e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} \qquad (259)\\ &= e^{2A_1} + e^{2A_2} + 2e^{A_1+A_2}\cos(B_1 - B_2) \\ &\quad + e^{A_1}e^{-\mathbf{B}_1}2e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} \\ &\quad + e^{A_2}e^{-\mathbf{B}_2}2e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} \\ &\quad + 2e^{\frac{1}{2}(A_1+A_2)}e^{-\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}e^{A_1}e^{\mathbf{B}_1} \\ &\quad + 2e^{\frac{1}{2}(A_1+A_2)}e^{-\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}e^{A_2}e^{\mathbf{B}_2} \\ &\quad + 4e^{A_1+A_2} \qquad (260)\\ &= \underbrace{e^{2A_1} + e^{2A_2}}_{\text{sum}} + \underbrace{2e^{A_1+A_2}\cos(B_1 - B_2)}_{\text{complex interference}} \\ &\quad \underbrace{+2e^{\frac{1}{2}(A_1+A_2)}(e^{A_1} + e^{A_2})\cos\frac{1}{2}(B_1 - B_2) + 4e^{A_1+A_2}}_{\text{deep phase interference}} \qquad (261)\end{aligned}$$

6.1.8 Geometric Interference in 4D (Deep Spinor Rotation)

Consider a two-state wave-function (we note that $[\mathbf{F}, \mathbf{B}] = 0$):

$$\psi = \psi_1 + \psi_2 = e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{F}_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{F}_2}e^{\frac{1}{2}\mathbf{B}_2} \tag{262}$$

The geometric interference pattern for a full general linear transformation in 4D is given by the product:

$$\lfloor \psi^\ddagger \psi \rfloor_{3,4} \psi^\ddagger \psi \tag{263}$$

Let us start with the sub-product:

$$\psi^\ddagger \psi = (e^{\frac{1}{2}A_1} e^{-\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2} e^{-\frac{1}{2}\mathbf{F}_2} e^{\frac{1}{2}\mathbf{B}_2})(e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2} e^{\frac{1}{2}\mathbf{F}_2} e^{\frac{1}{2}\mathbf{B}_2}) \tag{264}$$

$$\begin{aligned} &= e^{\frac{1}{2}A_1} e^{-\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1} e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_1} e^{-\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1} e^{\frac{1}{2}A_2} e^{\frac{1}{2}\mathbf{F}_2} e^{\frac{1}{2}\mathbf{B}_2} \\ &\quad + e^{\frac{1}{2}A_2} e^{-\frac{1}{2}\mathbf{F}_2} e^{\frac{1}{2}\mathbf{B}_2} e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2} e^{-\frac{1}{2}\mathbf{F}_2} e^{\frac{1}{2}\mathbf{B}_2} e^{\frac{1}{2}A_2} e^{\frac{1}{2}\mathbf{F}_2} e^{\frac{1}{2}\mathbf{B}_2} \end{aligned} \tag{265}$$

$$= e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{B}_2} + e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} (e^{-\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{F}_2}) + e^{-\frac{1}{2}\mathbf{F}_2} e^{\frac{1}{2}\mathbf{F}_1}) \tag{266}$$

$$= e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{B}_2} + e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) \tag{267}$$

where $R = e^{\frac{1}{2}\mathbf{F}}$, and where $\tilde{R} = e^{-\frac{1}{2}\mathbf{F}}$.
The full product is:

$$\lfloor \psi^\ddagger \psi \rfloor_{3,4} \psi^\ddagger \psi = \Big(e^{A_1} e^{-\mathbf{B}_1} + e^{A_2} e^{-\mathbf{B}_2} + e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(-\mathbf{B}_1-\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) \Big) \times \Big(e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{B}_2} + e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) \Big) \tag{268}$$

$$\begin{aligned} &= e^{A_1} e^{-\mathbf{B}_1} e^{A_1} e^{\mathbf{B}_1} + e^{A_1} e^{-\mathbf{B}_1} e^{A_2} e^{\mathbf{B}_2} + e^{A_1} e^{-\mathbf{B}_1} e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) \\ &\quad + e^{A_2} e^{-\mathbf{B}_2} e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{-\mathbf{B}_2} e^{A_2} e^{\mathbf{B}_2} + e^{A_2} e^{-\mathbf{B}_2} e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) \\ &\quad + e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(-\mathbf{B}_1-\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) e^{A_1} e^{\mathbf{B}_1} \\ &\quad\quad + e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(-\mathbf{B}_1-\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) e^{A_2} e^{\mathbf{B}_2} \\ &\quad\quad + e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(-\mathbf{B}_1-\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) \end{aligned} \tag{269}$$

$$= e^{2A_1} + e^{2A_2} + 2e^{A_1+A_2} \cos(B_1 - B_2) \tag{270}$$

$$+ e^{\frac{1}{2}(A_1+A_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) (\tag{271}$$

$$e^{A_1} (e^{\frac{1}{2}(-\mathbf{B}_1+\mathbf{B}_2)} + e^{\frac{1}{2}(\mathbf{B}_1-\mathbf{B}_2)}) \tag{272}$$

$$+ e^{A_2} (e^{\frac{1}{2}(\mathbf{B}_1-\mathbf{B}_2)} + e^{\frac{1}{2}(-\mathbf{B}_1+\mathbf{B}_2)})) \tag{273}$$

$$+ e^{A_1+A_2} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1)^2 \tag{274}$$

$$= \underbrace{e^{2A_1} + e^{2A_2}}_{\text{sum}} + \underbrace{2e^{A_1+A_2} \cos(B_1 - B_2)}_{\text{complex interference}}$$

$$+ \underbrace{2e^{\frac{1}{2}(A_1+A_2)} (e^{A_1} + e^{A_2}) (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) (\cos \frac{1}{2}(B_1 - B_2)) + e^{A_1+A_2} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1)^2}_{\text{deep spinor interference}} \tag{275}$$

6.1.9 Geometric Interference Experiment (Sketch)

In the case of the general linear group, 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.

General linear interference can be understood as a generalization of complex interference, which is recovered under a "shallow" phase rotation in 4D and under just a plain normal phase rotation in 2D. Furthermore, 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.

Such 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 as the identification tool. Identification of the full general linear interference pattern (with all the elements $A, \mathbf{X}, \mathbf{F}, \mathbf{V}, \mathbf{B}$) in a lab experiment would suggest a gauge-theoretical theory of gravity, whereas identification of a reduced interference pattern (produced by $A, \mathbf{F}, \mathbf{B}$) and subsequently showing a failure to observe the full general linear interference ($\mathbf{X} \to 0, \mathbf{V} \to 0$) would suggest at most spinor-level interference.

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.

One can further utilize 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 construction 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.

7 Discussion — Existence of the Universe

The overall implications of the existence of this type of proof from first principles are of course numerous. Here, let us discuss what we believe is likely it's most significant implication. Using our formal system of science, we have established the following point:

- The model which maximizes the axiomatic information produced by constructing the knowledge-base via the scientific method, is that of the universe (formulated as a theory of quantum gravity).

This explains why scientists (in the wild) eventually converged towards a scientific model of reality in which they were inhabitants of a universe as the kind described by the laws of physics here-in derived. The production of such model as always been a guaranteed result obtained at the end of scientific process and holds for all possible knowledge-bases. The universe is not special, it is merely an artefact of the structure of knowledge and how informative models of knowledge are produced via entropy maximization — although without the tools developed in this paper the connection is too obscure to notice.

The derivation of the model holds whether the 'universe exists physically or not', decoupling itself from the need to settle this question with an answer. This is why the system resolves the problem of the existence of the universe elegantly; by deriving it as a model regardless of its true physical status, it can show that all possible observers practicing science will eventually conclude themselves to be residents of such a universe as the will recover said model at the end of the inquiry. Since it holds for all possible knowledge-bases, and knowledge-base are the correct starting point to describe reality (they are first principles), the proof holds from true first principles.

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] Gregory J Chaitin. Meta math! the quest for omega. *arXiv preprint math/0404335*, 2004.

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

[5] Andrew M Gleason. Measures on the closed subspaces of a hilbert space. In *The Logico-Algebraic Approach to Quantum Mechanics*, pages 123–133. Springer, 1975.

[6] Ruth E Kastner. *The transactional interpretation of quantum mechanics: the reality of possibility*. Cambridge University Press, 2013.

[7] Lluís Masanes, Thomas D Galley, and Markus P Müller. The measurement postulates of quantum mechanics are operationally redundant. *Nature communications*, 10(1):1–6, 2019.

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

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

[10] 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 m dimensions over a field $\mathbb{F}$ is noted as $\mathbb{G}(m, \mathbb{F})$. 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 $(\langle \mathbf{v} \rangle_0 + \langle \mathbf{v} \rangle_2 + \langle \mathbf{v} \rangle_4 + \ldots)$ or odd grades $(\langle \mathbf{v} \rangle_1 + \langle \mathbf{v} \rangle_3 + \ldots)$ 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. 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 $a_0a_1 + x_0x_1\hat{\mathbf{x}} + y_0y_1\hat{\mathbf{y}} + b_0b_1\hat{\mathbf{x}} \wedge \hat{\mathbf{y}}$.

B Lagrange equation

The Lagrangian equation to maximize is:

$$\mathcal{L}(\rho, \alpha, \tau) = -k_B \sum_{q \in \mathbb{Q}} \rho(q) \ln \rho(q) + \alpha \left(1 - \sum_{q \in \mathbb{Q}} \rho(q) \right) + \tau \operatorname{tr} \left(\overline{\mathbf{M}} - \sum_{q \in \mathbb{Q}} \rho(q) \mathbf{M}(q) \right) \tag{276}$$

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:

$$\frac{\partial \mathcal{L}}{\partial \rho(p)} = -k_B \ln \rho(p) - k_B - \alpha - \tau \operatorname{tr} \mathbf{M}(p) \tag{277}$$

$$0 = k_B \ln \rho(p) + k_B + \alpha + \tau \operatorname{tr} \mathbf{M}(p) \tag{278}$$

$$\implies \ln \rho(p) = \frac{1}{k_B} \left(-k_B - \alpha - \tau \operatorname{tr} \mathbf{M}(p) \right) \tag{279}$$

$$\implies \rho(p) = \exp\left(\frac{-k_B - \alpha}{k_B} \right) \exp\left(-\frac{\tau}{k_B} \operatorname{tr} \mathbf{M}(p) \right) \tag{280}$$

$$= \frac{1}{Z} \det \exp\left(-\frac{\tau}{k_B} \mathbf{M}(p) \right) \tag{281}$$

where Z is obtained as follows:

$$1 = \sum_{q \in \mathbb{Q}} \exp\left(\frac{-k_B - \alpha}{k_B} \right) \exp\left(-\frac{\tau}{k_B} \operatorname{tr} \mathbf{M}(q) \right) \tag{282}$$

$$\implies \left(\exp\left(\frac{-k_B - \alpha}{k_B} \right) \right)^{-1} = \sum_{q \in \mathbb{Q}} \exp\left(-\frac{\tau}{k_B} \operatorname{tr} \mathbf{M}(q) \right) \tag{283}$$

$$Z := \sum_{q \in \mathbb{Q}} \det \exp\left(-\frac{\tau}{k_B} \mathbf{M}(q) \right) \tag{284}$$

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:

$$\overline{\mathbf{M}}_1 = \sum_{q\in\mathbb{Q}} \rho(q)\mathbf{M}_1(q) \tag{285}$$

$$\vdots \tag{286}$$

$$\overline{\mathbf{M}}_n = \sum_{q\in\mathbb{Q}} \rho(q)\mathbf{M}_n(q) \tag{287}$$

Then the Lagrange equation becomes:

$$\mathcal{L} = -k_B \sum_{q\in\mathbb{Q}} \rho(q)\ln\rho(q) + \alpha\left(1 - \sum_{q\in\mathbb{Q}}\rho(q)\right) + \tau_1 \operatorname{tr}\left(\overline{\mathbf{M}}_1 - \sum_{q\in\mathbb{Q}}\rho(q)\mathbf{M}_1(q)\right) + \ldots + \tau_n \operatorname{tr}\left(\overline{\mathbf{M}}_n - \sum_{q\in\mathbb{Q}}\rho(q)\mathbf{M}_n(q)\right) \tag{288}$$

and the measure references all n constraints:

$$\rho(q) = \frac{1}{Z}\det\exp\left(-\frac{\tau_1}{k_B}\mathbf{M}_1(q) - \cdots - \frac{\tau_n}{k_B}\mathbf{M}_n(q)\right) \tag{289}$$

B.2 Multiple constraints - General Case

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

$$\overline{M}_{00}\begin{pmatrix} 1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 0 \end{pmatrix} = \sum_{q\in\mathbb{Q}} \rho(q)M_{00}(q)\begin{pmatrix} 1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 0 \end{pmatrix} \tag{290}$$

$$\vdots \tag{291}$$

$$\overline{M}_{01}\begin{pmatrix} 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix} = \sum_{q\in\mathbb{Q}} \rho(q)M_{01}(q)\begin{pmatrix} 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix} \tag{292}$$

$$\vdots \tag{293}$$

$$\overline{M}_{nn}\begin{pmatrix} 0 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 1 \end{pmatrix} = \sum_{q\in\mathbb{Q}} \rho(q)M_{nn}(q)\begin{pmatrix} 0 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 1 \end{pmatrix} \tag{294}$$

For a $n \times n$ matrix, there are n^2 constraints.
The probability measure which maximizes the entropy is as follows:

$$\rho(q) = \frac{1}{Z} \det \exp\left(-\frac{1}{k_B} \boldsymbol{\tau} \odot \mathbf{M}(q) \right) \tag{295}$$

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