

The universe from a single particle

Michael Freedman^a and Modjtaba Shokrian Zini^{b,1}

^aMicrosoft Research, Station Q, and Department of Mathematics,
University of California, Santa Barbara,
Santa Barbara, CA 93106, U.S.A.

^bPerimeter Institute for Theoretical Physics,
Waterloo, ON N2L 2Y5, Canada

E-mail: michaelf@microsoft.com, mshokrianzini@pitp.ca

ABSTRACT: We explore the emergence of many-body physics from quantum mechanics via spontaneous symmetry breaking. To this end, we study potentials which are functionals on the space of Hamiltonians enjoying an unstable critical point corresponding to a random quantum mechanical system (the Gaussian unitary ensemble), but also less symmetrical local minima corresponding to interacting systems at the level of operators.

KEYWORDS: Effective Field Theories, Spontaneous Symmetry Breaking

ARXIV EPRINT: [2011.05917](https://arxiv.org/abs/2011.05917)

¹Research was done as part of an internship of the second named author at Microsoft.

Contents

1	Introduction	1
1.1	The summary of results	8
2	Numerical methods	9
2.1	Loss function	9
2.1.1	Gaussian perturbed integral	9
2.1.2	Ricci scalar curvature	13
2.2	Initialization of the gradient descent	14
2.3	Adam optimization algorithm	15
2.3.1	Scheduling and hyperparameters	16
3	Optimization results and conclusions	17
3.1	DiagPerturbId and GenPerturbId results	19
3.2	PenMetricOmega and PenMetricPattern results	22
3.3	Ricci scalar curvature results	22
4	Summary and outlook	22
A	The structure constants can only distinguish weights	23
B	Gradient descent and degeneracy patterns	26
C	Individual diagrams are convex	28

1 Introduction

Since Wilson [1] much attention both in condensed matter and high energy has been devoted to the question: can theory B emerge as a low energy limit of theory A? And even possible concatenations $A \rightarrow B \rightarrow C \rightarrow \dots$? Effective field theories are undoubtedly important links in such chains. This paper explores, in a toy model, the possibility that field theory itself, even at high energy, may arise through a kind of symmetry breaking applied to a much simpler system, single particle quantum mechanics.

In our toy model, single particle quantum mechanics is represented by theory A with a finite dimensional Hilbert space of states $\mathcal{H}_A \cong \mathbb{C}^N$. Initially, and for simplicity, we take N to be a power of 2, $N = 2^n$, but, as we will see, this choice simplifies the discussion but may not be essential. We study symmetry breaking to a theory B which is a toy model for field theory. Theory B has a Hilbert space \mathcal{H}_B isometric to $\underbrace{\mathbb{C}^2 \otimes \dots \otimes \mathbb{C}^2}_{n \text{ times}}$ but \mathcal{H}_B is regarded as describing a system of n interacting bosons, each with two states (a very similar choice, not explored here, is for \mathcal{H}_B to describe n interacting spin- $\frac{1}{2}$ Fermions). The difference between models A and B is the probability distribution from which the system

Hamiltonian, which we think of as the Hamiltonian of the universe, is drawn. In case A, we consider that the Hamiltonian H_A is drawn *randomly* from the Gaussian Unitary Ensemble (GUE) \mathcal{Z} on $\mathfrak{su}(N)$, the Lie algebra of the group of symmetries of \mathcal{H}_A . We consider the simple Lie group $SU(N)$ to be the symmetry group of \mathcal{H}_A as overall phase is unobservable. Since the elements of $\mathfrak{su}(N)$ are skew-Hermitian, we multiply by $-i$ to get a Hermitian Hamiltonian. To complete the definition of theory A we specify the metric g_{ij} on $\mathfrak{su}(N)$. We choose g_{ij} to be the (ad-invariant) killing form: $\langle a, b \rangle \propto \text{tr}(\text{ad}(a) \circ \text{ad}(b))$. Specifying the metric makes the distribution \mathcal{Z} well-defined and induces a measure on $\mathfrak{su}(N)$, giving meaning to the word “randomly” above. In the case g_{ij} is the Killing form, the resulting distribution is the GUE.

Concretely, imagine a particle with access to N states $|1\rangle, \dots, |N\rangle$, with the system Hamiltonian H_A built by assembling N^2 i.i.d real Gaussian random variables into

1. the diagonal entries of H_A ,
2. and the real and imaginary parts of $(H_A)_{ij}, i > j$ with $(H_A)_{ij} = \overline{(H_A)_{ji}}$.

Constructing H_A in this way reflects a complete agnosticism on the possible transitions, making H_A the Hamiltonian for a random single particle system.

By contrast, in case of B, some qubit structure has been chosen in the form of an isometry: $\mathbb{C}^{2^n} \xrightarrow{J} \underbrace{\mathbb{C}^2 \otimes \dots \otimes \mathbb{C}^2}_n$. Thus the moduli space of qubits structures is the homogeneous space

$$\mathfrak{m} = \text{U}(2^n) / \underbrace{\text{U}(2) \times \dots \times \text{U}(2)}_n,$$

where J determines the inclusion. There is now a central definition:

Definition 1.1. A non-singular inner product (i.e. a metric) g_{ij} on $\mathfrak{su}(2^n)$ which is *not* bi-invariant, **knows about qubits** (is **kaq**) if there is a basis $\{iH_1, \dots, iH_{4^n-1}\}$ for $\mathfrak{su}(2^n)$ consisting of principal axes for g_{ij} with the property that there is an isomorphism of Lie algebras

$$j : \underbrace{\mathfrak{u}(2) \otimes \dots \otimes \mathfrak{u}(2)}_{n \text{ copies}} \rightarrow \mathfrak{u}(2^n)$$

such that $iH_k = j(O_{1,k} \otimes \dots \otimes O_{n,k})$ for $1 \leq k \leq 4^n - 1$, where $O_{i,k}$ lives in the i -th copy of $\mathfrak{u}(2)$; that is, there is a basis of principal axes which are *pure tensors* relative to j .

The previous choice J determines j . The notions of isometry and principal axis above are with respect to a unique (up to scale) bi-invariant metric $\langle a, b \rangle = \frac{1}{\dim \mathfrak{su}(N)} \text{tr}(\text{ad}(a) \circ \text{ad}(b))$. Implicitly, this definition exploits the natural inclusion $SU(N) \subset U(N)$, since our g_{ij} is defined only on the former.

For clarity, we remark that if the (inverse) bi-invariant metric is used to raise an index of g_{ij} to create a symmetric operator g_i^j , then the principal axes of g_{ij} are precisely the eigenvectors of g_i^j . The reason the definition refers to a basis of principal axes (eigenvectors) instead of *the* basis is that in the metrics we encounter there is usually degeneracy (to numerical precision) so we need the freedom to look within eigenspaces for a basis of

simple (pure tensor form) eigenvectors. In practice, one might wish to soften the definition slightly and allow eigenspaces with *nearly* identical eigenvalues also to be combined, thus allowing more freedom to find simple linear combinations.

We can generalize the definition above to **knows-about-qunits (kaqn)**. We have used **qubits** to define **kaq** but the reader can imagine other primes or combination of primes not restricted to 2. In this initial paper, we have not attempted computations in this direction but expect to study this in future works.

One challenge we face is the large dimension $4^n - 3n - 1$ of \mathfrak{m} . To decide if a given metric g_{ij} on $\mathfrak{su}(2^n)$ is or is not **kaq**, it may be necessary to numerically search \mathfrak{m} . For $n = 2$, we did this on several occasions. But a particularly simple family of **kaq** metrics are those where there is a basis of principal axes of g_{ij} consisting of Pauli words. The structure J determines a *Pauli-word* basis, called \mathbf{PB}_n on the Lie algebra $\mathfrak{su}(2^n) =$

$$\{-iX \otimes 1 \otimes \cdots \otimes 1, iY \otimes 1 \otimes \cdots \otimes 1, -iZ \otimes 1 \otimes \cdots \otimes 1, -iX \otimes Y \otimes 1 \cdots \otimes 1, \dots, -iZ \otimes Z \otimes \cdots \otimes Z\},$$

consisting of all words of length n in the letters $\{1, X, Y, Z\}$ except the all 1 word which is in $\mathfrak{u}(2^n)$ but not $\mathfrak{su}(2^n)$. Given J , this basis is canonical up to the action of $SU(2)$ on each factor.

Hence forth, we normalize all metrics (also the killing form) so that $\det(g_{ij}) = 1$. A very interesting special case of **kaq** metrics on $\mathfrak{su}(2^n)$ are the previously studied *penalty* metrics of [2–4], [5, 6], and [7] where

$$g_{ij} \propto e^{rw(i)} \delta_{ij} \tag{1.1}$$

where r is some positive constant and $w(i)$ is the *weight* of the i th Pauli-word meaning the number of letters it contains, X, Y, or Z, which are different from 1. Of course:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.2}$$

We should emphasize that **kaq** metrics are exceedingly rare having roughly square root the number of parameters as general metrics. A **kaq** metric needs $4^n - 1$ parameters to specify j (on which $\text{Aut}(\mathfrak{su}(2^n))$ acts freely and transitively), and also $4n$ parameters for each product of $O_{i,k}$'s of which there are $4^n - 1$ (as $1 \leq k \leq 4^n - 1$). In total this makes $(4^n - 1) + (4^n - 1)(4n) = (4n + 1)(4^n - 1)$ parameters for a **kaq** metric, whereas the metrics space g_{ij} on $\mathfrak{su}(2^n)$ has dimension $\frac{4^n(4^n-1)}{2}$.

This paper studies symmetry breaking from theory A to theory B. This is a symmetry breaking at the level of (probability distributions on) operators, not of states as is familiar, for example, in the Ginzburg-Landau theory of superconductivity. To do this, we consider (one of several possible) functionals:

$$f : \{\text{metrics}\} \rightarrow \mathbb{R} \tag{1.3}$$

on the space of $\det = 1$ inner products (i.e. normalized metrics) on $\mathfrak{su}(2^n)$ and ask: does f break the symmetry of the bi-invariant Killing form to a **kaq** metric?

Let us repeat what we are looking for and why we are looking for it.

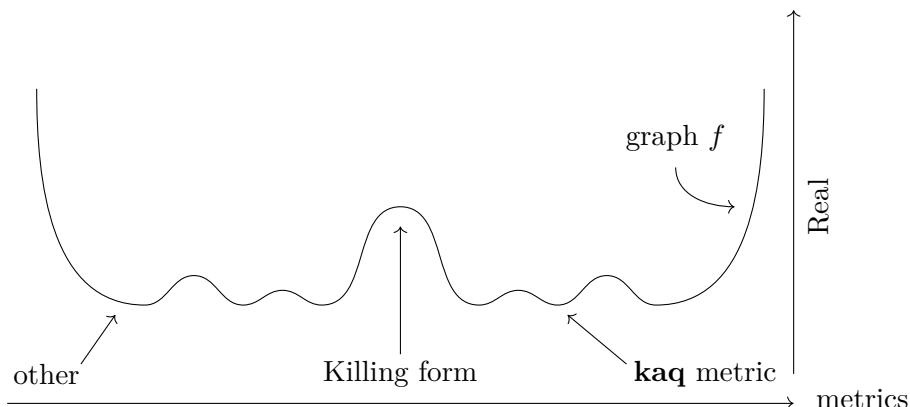


Figure 1. Conjectural symmetry breaking from Killing to **kaq**.

First the “what”. We are looking for a functional f on metrics which, if minimized, would confer the usual features of interacting physics on what, at a fundamental level, is actually a single particle system. The systems (A and B) we study are each fully quantum mechanical but are each drawn from their separate statistical ensembles: Gaussians defined from quite different inner products. Because the overwhelming preponderance of **kaq** metrics produce measures with infinitesimally small overlap with the ad-invariant measure (induced by the Killing form), a **kaq** metric will display observable features of many-body physics, almost certainly not seen in a Killing-random (GUE) Hamiltonian from theory A. The distinction is most clear in the special case of penalty metric in which the expected strength of k -body terms in the Hamiltonian H_B , such as

$$11 \underbrace{XYZ \dots ZY}_{k \text{ letters}} 1 \dots 1 \tag{1.4}$$

(\otimes is suppressed and the non-identity letters are not required to be consecutive), will decay exponentially with k .

If one were handed a seemingly random $2^n \times 2^n$ Hermitian matrix and discovered (not an easy computational task) that it was mostly captured, in some tensor coordinates, by low weight terms, one would believe many-body physics was at work. The SYK model $H = \sum J_{ijkl} \gamma_i \gamma_j \gamma_k \gamma_l$ [8] is a (fermionic) illustration. It assembles the system Hamiltonian for a black hole from random 4-Majorana interactions (analogous to 2-body interactions in the bosonic case) governed by a complete graph (no “spatial locality”). The SYK model can easily be enhanced, in the spirit of penalty metrics, to allow exponentially weakening higher body interactions:

$$H = \sum_{p=2}^n e^{\text{const. } p} \left(\sum_{i_1, \dots, i_{2p}} J_{i_1 \dots i_{2p}} \gamma_{i_1} \dots \gamma_{i_{2p}} \right) \tag{1.5}$$

Such Hamiltonians are part and parcel of many-body physics. It would indeed be surprising to be told that the system was “really” a single particle whose governing Hamiltonian had rolled off an unstable equilibrium into a **kaq** well. This is the surprise we are investigating.

Before addressing the “why,” a word about the parameter space {metrics} and the functionals f we investigate.

The geometry of compact simple Lie groups with bi-invariant metrics is very well understood [9]. They are symmetric spaces whose geodesics are the cosets of 1-parameter subgroups. In this case much global information is known: for example, all points on the cut locus (say from id) are also conjugate points. On the other hand, if the Riemannian metric is merely left-invariant the geometry is a rich subject of contemporary research. Local computations are straightforward [10, 11] but global properties are harder to derive (in [7] it was recently shown that for a fixed constant $c > 0$ the diameter of the penalty metrics on $SU(2^n)$ is larger by a factor, exponential in n , than the Killing metric).

Choosing a left-invariant metric on a Lie group G is of course the same thing as choosing an inner product on its Lie algebra \mathfrak{g} (and moving that inner product about by the differential of left multiplication). So when we study {metrics} on $\mathfrak{su}(2^n)$ we are really studying left invariant Riemannian geometries on $SU(2^n)$. The normalization condition $\det(g_{ij}) = 1$ ensures a fixed volume for $SU(2^n)$ as the metric varies. The penalty metrics make high weight directions exponentially expensive so a typical H_B for such a metric will consist mostly of low-weight (i.e. few-body) interactions, i.e. look like typical many-body physics. So we are looking for functionals which break symmetry to penalty metrics or at least **kaq** metrics. It is clearly *cheating* to mention a qubit structure in the *definition* of f . f should be defined only from what is intrinsic to noninteracting physics, the structure constants of $\mathfrak{su}(2^n)$, c_{ij}^k , and the metric g_{ij} .

As geometers, a very natural functional might be $f = -\text{diameter}$. For the Killing metric, the diameter is π but can be substantially increased by varying g_{ij} from id. While $f = -\text{diameter}$ might eventually reward investigation, we must restrict ourselves to studying functionals f where computations are efficient. So we have tried to assemble for study the most natural functionals built from the tensors g_{ij} and c_{ij}^k . For f , we settled on Ricci scalar curvature which is easily expressed in terms of g_{ij} and c_{ij}^k , as well as a variety of (cubically) perturbed finite dimensional Gaussian integrals. The practical reason for studying these is that it is easy to work out the first few terms of their perturbative expansions in terms of tensor contraction diagrams. Once (a few terms of) the perturbative expansion is in hand, there are efficient methods, e.g. Padé approximates [12] to achieve *rapid* convergence (to be explored in future works) but merely summing the first few terms is useful when the coupling k to the cubic term is small. These integrals are described in the next section.

Regrading the accuracy of our methods, a mathematical admission is in order. We do not have a proof that the integrals we use to define our functional f have an analytic meaning. This is currently under investigation with preliminary remarks in section 4. We proceed instead, as one often does in field theory, trusting that the integral expression, through some regularization, actually defines a function of a (complex) expansion parameter and trust that detailed information on this function can be extracted from its asymptotic expansion (possibly post-processed via Padé approximates). Of course, this issue is not present when $f = \text{Ricci scalar curvature}$.

Finally, why? The Wilsonian dream is that at highest energy the universe has a simple mathematical description and that any complexity in the apparent physical laws

$$\text{boundary CFT} \xrightarrow{\text{dual}} \left(\begin{array}{c} \text{stringy} \\ \text{bulk theory} \end{array} \right) \xrightarrow{\text{SSB}} ? \rightarrow \dots \xrightarrow{\text{SSB}} \text{Standard Model} \\ \cup \\ \text{Quantum mechanics (QM)} \\ \text{(single particle)}$$

Figure 2. A typical Wilsonian dream.

$$\left(\begin{array}{c} \text{QM on} \\ \text{Hilbert space} \\ \cong \mathbb{C}^N \end{array} \right) \xrightarrow{\text{SSB}} \text{boundary CFT} \xrightarrow{\text{dual}} \left(\begin{array}{c} \text{stringy} \\ \text{bulk theory} \end{array} \right) \xrightarrow{\text{SSB}} ? \rightarrow \dots \xrightarrow{\text{SSB}} \text{Standard Model} \\ \cup \\ \text{QM}$$

Figure 3. Dream with prequel.

arose through spontaneous symmetry breaking (SSB). For example:

All the theories in any such story have necessarily been interacting because they are intended to model an obviously interacting universe. The goal of this paper is to propose, in the context of a highly restricted toy model, that a prequel is possible. That single particle quantum mechanics could also be the beginning of the story. Adding quantum mechanics also to the far left in figure 2, we propose that SSB, at the operator level, could allow QM to simulate the interacting systems from which the familiar story is told, as shown in figure 3.

Although the vision is broad, the goal of this paper is modest: explore symmetry breaking from QM to a **kaq**-geometry. The reader will naturally wonder when we will make contact with experiment or observation. Our aim here, is quite low but fundamental: to address the fact that the universe appears to contain more than one thing. It would be nice to deduce: the dimension(s) of spacetime, the structure of physical laws, the existence of fermions, and the value of the fine-structure constant. But the truth is that we are nowhere near ready to address these questions. However, to the large ensemble of untouched questions, we would like to add a speculation.

It may have, and should have, seemed unnatural that for theory A, our random finite dimensional QM system was placed on a Hilbert space of dimension $N = 2^n$. What would one expect if N was just a randomly selected large integer? Could there still be a symmetry breaking story to a collection of various qunits? If N is chosen at random in a certain size range (say 100 digit numbers), we can ask what is the expected size of its largest prime factor. The answer is that it likely will have about 62 digits. This phenomenon is governed by the Golomb-Dickman constant $\approx 0.624329989\dots$. Much beyond this is known regarding the statistics of prime factorizations of large random integers. If N instead of being a power of 2, is random, perhaps a suitable functional f will break the large system into qunits of dimensions the prime factors of N .

As a test of our symmetry breaking thesis, one should look near the r.h.s. of figure 3 for some phenomenological residue of a factorization of the initial random integer N . N should indeed be large. Since the entropy of measured black holes can be 10^{90} [13, 14]

we would need $n \gg 10^{90}$ and $N \gg 2^{10^{90}}$, roughly what was once called Googolplex. To understand how the prime factorization of a large random integer might be imprinted into the low energy physics is a major challenge, and one approach to gathering physical evidence for the symmetry breaking hypothesis that we study here on a mathematical plane. The distribution of primes is tightly reflected in the Riemann zeta function, and similar zeta functions built from the length distribution of closed geodesics on a hyperbolic surface are linked through the Selberg trace formula [15] to the spectrum of the Laplacian on these surfaces. So statistical properties of factorization could have an audible echo in string dynamics if we know what to listen for.

Another mathematical possibility to be explored is that the prime 2 is energetically favored because of the relation to Majorana/Clifford algebras, and that for a general integer N , SSB will simply round down to a power of 2, splitting the Hilbert space into a direct sum of glassy physics, and a left-over degrees of freedom. The size of the power of 2 may depend on energy/entropy balance, as there should be many more shallow, smaller Clifford algebra minima and fewer larger and deeper ones. Leftover degrees of freedom may separately organize themselves into additional interacting systems. A Hamiltonian which is a superposition of several separate interacting sectors would govern separate world branches which, like the live and dead cat, do not know about each other.

In future works, we hope to investigate SSB to “partial”- $\mathbf{kaq}(\mathbf{n})$ minima in which a proper summand of the Hilbert space is $\mathbf{kaq}(\mathbf{n})$, or the possibility of “multi”- $\mathbf{kaq}(\mathbf{n})$ minima where the Hilbert space decomposes into several subspaces each of which has a $\mathbf{kaq}(\mathbf{n})$ decomposition. For example, one might find a 17D Hilbert space decomposing as $(\mathbb{C}^3)^{\otimes 2} \oplus (\mathbb{C}^2)^{\otimes 3}$.

Let us finish this introductory section with integrals used to define functionals $f : \{\text{metrics}\} \rightarrow \mathbb{R}$, and a discussion of their symmetries. A discussion of the numerics and ML used to enhance the numerics is in the next sections.

We define the complex-valued functions: $F_k : \{\text{metrics}\} \rightarrow \mathbb{C}$, and consider functions of the form

$$f_{k,1} = \text{Re}(F_k), \quad f_{k,2} = \text{Im}(F_k), \quad f_{k,3} = |F_k|^2, \quad k \in \mathbb{C} \quad (1.6)$$

$$F_k = \int_{\vec{x} \in \mathbb{R}^{3(4^n-1)}} d\vec{x} e^{ik(G_{IJ}x^I x^J + c_{ijk}y_1^i y_2^j y_3^k)} \quad (1.7)$$

for $x = (y_1, y_2, y_3)$ with $y_o \in \mathbb{R}^{4^n-1}$, $o \in \{1, 2, 3\}$, and for $I = (i, o)$, $x^I = y_o^i \in \mathbb{R}$, and $G_{IJ}x^I x^J = g_{ij}y_1^i y_1^j + g_{ij}y_2^i y_2^j + g_{ij}y_3^i y_3^j$, i.e. $G = \begin{pmatrix} g & 0 & 0 \\ 0 & g & 0 \\ 0 & 0 & g \end{pmatrix}$.

The structure constants c_{ij}^k are defined by

$$[y_i, y_j] = c_{ij}^k y_k \quad \text{and} \quad c_{ijk} = c_{ij}^{k'} g_{k'k} \quad (1.8)$$

This seems to be the simplest perturbed Gaussian that can be formed from the two tensors g and c which will not vanish due to symmetry consideration. For example, if we merely integrated over one copy, instead of three, of the Lie algebra $\mathfrak{su}(2^n) \cong \mathbb{R}^{4^n-1}$, using

g_{ij} instead of G_{IJ} the integral would become

$$\tilde{F}_k = \int_{\vec{y} \in \mathbb{R}^{4n-1}} d\vec{y} e^{ik(g_{ij}y^i y^j + c_{ijk}y^i y^j y^k)}. \tag{1.9}$$

But observe that the cubic perturbation $c_{ijk}y^i y^j y^k = 0$ since $c_{ijk} = -c_{jik}$, by the skew symmetry of the Lie bracket $[\cdot, \cdot]$. The formula in (1.7) indeed seems the most natural way to build a nontrivial perturbed Gaussian integral from the tensors g and c .

Another functional that will be considered is the Euclidean version of the above, where $-i$ in the exponent in (1.7) is replaced by -1 .

We should say a little about the symmetries of F_k restricted to the **kaq** subset $\mathcal{M}_{\mathbf{kaq}}^n \subset \mathcal{M}^n := \{\text{metrics on } \mathfrak{su}(N)\}$. There is the adjoint action of $SU(N)$ on $\mathfrak{su}(N)$ inducing an action on \mathcal{M}^n , and since $\mathcal{M}_{\mathbf{kaq}}^n$ is defined via the *existence* of some qubit decomposition, this action preserves $\mathcal{M}_{\mathbf{kaq}}^n$. The isomorphism to $\underbrace{\mathbb{C}^2 \otimes \dots \otimes \mathbb{C}^2}_n$ is merely precomposed by the inverse of the element of $SU(N)$. This action, for example has isotopy properly containing $\underbrace{(SU(2) \times \dots \times SU(2))}_{n \text{ copies}}$ on $\mathcal{M}_{\text{penalty}}^n$.

For numerical reasons, we are best able to study $f_{k,2}$ for k positive real in the range of $100 \leq k \leq 1000$ and $f_{k,1}$ for k positive imaginary in a similar range.

Once we locate a locally minimizing metric g we diagonalize it and call the $4^n - 1$ eigenvalues the *large spectrum*. Each of the corresponding $4^n - 1$ eigenvalues is a $2^n \times 2^n$ skew-Hermitian matrix and we refer to their spectra (which are imaginary) as the *little spectra*. When there is eigenvalue degeneracy, there is additional choice choosing principle axes (eigenvectors of g_i^j) and the concomitant little spectra.

There is a very quick and often reliable check that a metric is in $\mathcal{M}_{\mathbf{kaq}}^n$: the little spectra should all be consistent with the proposed tensor product structure. This is useful primarily when a non-degenerate eigenvalue in the little spectrum (a 1D eigenspace). A further more definitive test is discussed in section 4. In practice we can usually understand the **kaq** status of the local minima we locate.

The rest of the paper is organized as follows:

- Section 2. A detailed look at the numerical methods and the machine learning techniques that enabled them.
- Section 3. A summary of what computations were carried out and the most significant conclusions for symmetry breaking to **kaqs**.
- Section 4. Summary and outlook.
- Appendices A–C. Appendices on the theoretical and linear algebra results.

1.1 The summary of results

Numerical study of a variety of functionals on inner product on $\mathfrak{su}(4)$ and $\mathfrak{su}(8)$ finds, in addition to an unstable critical point at the bi-invariant Killing form, metric structure often involving multiple local minima. In many cases we have identified these as **kaq** metrics

(see figures 9 and 10). Although our study was necessarily limited to very low dimensions, it provides evidence for our proposed SSB scenario. It is furthermore intriguing that the choice of functional does not need to be fine tuned, indicating that perhaps many natural functionals formed from g_{ij} and c_{ij}^k commonly have **kaq** minima. One could add another layer of probability and consider not just our system Hamiltonian H selected from a metric dependent Gaussian, but also that metric g_{ij} drawn from local minima of functionals f themselves drawn from a distribution \mathcal{D} of functionals.

Next, a few disclaimers: Ricci scalar curvature was numerically challenging and we were only able to locate critical points whose locations had previously been determined analytically [16] and we already knew would be **kaq**. These local minima are associated in the two cases with the Lie subalgebras $\mathfrak{so}(4) \subset \mathfrak{su}(4)$ and $\mathfrak{sp}(2) \subset \mathfrak{su}(4)$ respectively. Surprisingly, for the functionals studied on $\mathfrak{su}(4)$, some of the solutions appear numerically to be $\mathfrak{sp}(2)$ Jensen local minima as well. The metric is small on the Lie subalgebra and large on its orthogonal complement. For other local minima for $\mathfrak{su}(4)$ and $\mathfrak{su}(8)$, similar Lie **sub**algebras have been found that correspond to one (or a combination of some) of the eigenspaces. A more detailed description is given in section 3.

This points to the fact that in addition to the class **kaq**, there is a class on metrics which we shall call **sub_n** having the structure described above. Particularly for small n , **sub_n** and **kaq_n** have considerable intersection among metrics on $\mathfrak{su}(2^n)$.

We must consider to what extent the evidence is actually for **kaq** and not **sub**. So far, we have not seen local minima in **sub_n** \ **kaq_n**.

In summary, we find considerable numerical evidence to support the plausibility of spontaneous symmetry breaking to **kaq** metrics, as the prequel to interacting physics.

2 Numerical methods

2.1 Loss function

2.1.1 Gaussian perturbed integral

Expanding the perturbative series of (1.7) up to third order, yields a summation of two, four and six vertex trivalent tensor networks, where vertices are labelled by c and edges by g or g^{-1} . This expansion follows a similar procedure outlined in [17, equation 1.7 onwards]. We start with the bosonic Gaussian perturbed integral in (1.7), and make a simple change of variable $\vec{x} \rightarrow \vec{x} = \sqrt{k}\vec{x}$,

$$F_k = k^{-3(4^n-1)/2} \int_{\mathbb{R}^{3(4^n-1)}} d\vec{x} e^{\frac{i}{2}G_{IJ}x^I x^J} e^{\frac{i}{\sqrt{k}}c_{ijk}y_1^i y_2^j y_3^k} \tag{2.1}$$

$$= k^{-3(4^n-1)/2} \int_{\mathbb{R}^{3(4^n-1)}} d\vec{x} e^{\frac{i}{2}G_{IJ}x^I x^J} \sum_{m=0}^{\infty} \frac{i^m}{m!k^{m/2}} (c_{ijk}y_1^i y_2^j y_3^k)^m \tag{2.2}$$

Focusing on the m -th term of the expansion gives a Gaussian integral

$$\int_{\mathbb{R}^{3(4^n-1)}} d\vec{x} e^{\frac{i}{2}G_{IJ}x^I x^J} (c_{ijk}y_1^i y_2^j y_3^k)^m \tag{2.3}$$

which can be computed using standard methods

$$\left[\left(c_{ijk} \frac{-i\partial}{\partial V_1^i} \frac{-i\partial}{\partial V_2^j} \frac{-i\partial}{\partial V_3^k} \right)^m \int_{\mathbb{R}^{3(4^n-1)}} d\vec{x} e^{\frac{i}{2}(g_{ij}y_1^i y_1^j + g_{ij}y_2^i y_2^j + g_{ij}y_3^i y_3^j) + V_1^i y_1^i + V_2^j y_2^j + V_3^k y_3^k} \right]_{\vec{V}=0} \quad (2.4)$$

$$\propto \left[\left(c_{ijk} \frac{-i\partial}{\partial V_1^i} \frac{-i\partial}{\partial V_2^j} \frac{-i\partial}{\partial V_3^k} \right)^m e^{-\frac{i}{2}G^{IJ}V_I V_J} \right]_{\vec{V}=0} \quad (2.5)$$

Writing $e^{-\frac{i}{2}G^{IJ}V_I V_J}$ as a Taylor series $\sum_p \frac{(-\frac{i}{2}G^{IJ}V_I V_J)^p}{p!}$, since the exponent $G^{IJ}V_I V_J$ is a quadratic expression, the only nonzero terms after differentiation and setting $\vec{V} = 0$ correspond to the power $p = \frac{3m}{2}$, i.e.

$$\left(c_{ijk} \frac{-i\partial}{\partial V_1^i} \frac{-i\partial}{\partial V_2^j} \frac{-i\partial}{\partial V_3^k} \right)^m \left(-\frac{i}{2}G^{IJ}V_I V_J \right)^{\frac{3m}{2}}. \quad (2.6)$$

Indeed all other terms are zero, when $p < \frac{3m}{2}$, as there are $3m$ partials and less ($< 2 \times \frac{3m}{2}$) variables, and when $p > \frac{3m}{2}$, as $\vec{V} = 0$.

By following the power of imaginary number i in the previous equations, we observe that $\text{Im}(F_k) = f_{k,2}$ corresponds to $m \equiv 2 \pmod{4}$, while the rest is $\text{Re}(F_k) = f_{k,1}$. Similarly, the Euclidean version of this integral has alternating signs ± 1 depending on $m \pmod{4}$.

Our numerical experiments are based on $m = 2, 6$ and on $m = 2, 4$ for the Euclidean version.

We can show that the expression above can be viewed as a contraction of a trivalent network without any loops, and m vertices c_{ijk} and edges g^{ii}, g^{jj}, g^{kk} . Note each partial needs to be paired with its corresponding term in $(-\frac{i}{2}G^{IJ}V_I V_J)^{\frac{3m}{2}}$. Let us fix a j and consider the example of $\frac{-i\partial}{\partial V_2^j}$ which is associated with some term c_{-j-} . Inside $(-\frac{i}{2}G^{IJ}V_I V_J)^{\frac{3m}{2}}$, the term V_2^j comes with another $g^{jj'} V_2^{j'}$. Therefore, there can not be any loops, or any pairing between two different indices, like i and j . As $V_2^{j'}$ is similarly paired with $\frac{-i\partial}{\partial V_2^{j'}}$, which is associated with some term $c_{-j'-}$, we get an *edge* connecting the two *vertices* c_{-j-} and $c_{-j'-}$ with $g^{jj'}$ labeling that edge.

We note a further simplification, by using (1.8), allowing us to conclude that vertices can be labelled by c_{ij}^k instead of c_{ijk} , while edges are labelled by $g^{ii'}, g^{jj'}$ and $g_{kk'}$ instead of $g^{kk'}$. The reason for the change in the latter edge type is that when two vertices c_{ijk} and $c_{i'j'k'}$ are connected along the kk' indices, two factors of $g_{kk'}$ come from c_{ijk} and $c_{i'j'k'}$ (thus transforming them to $c_{ij}^k, c_{i'j'}^{k'}$), and one gets cancelled by $g^{kk'}$ coming from $g^{kk'} V_3^k V_3^{k'}$.

Each diagram also comes with a coefficient. The coefficient can be computed by following the scalars in equations above for each m , and considering all permutations to each diagram labelling.

We list all diagrams for $m = 2, 4, 6$ in figures 4, 5, 6 along with their coefficients.

For $m = 2$, there is only the diagram

- Theta

with coefficient $\frac{1}{2!}$. For $m = 4$, there is a common coefficient $\frac{1}{4!}$.

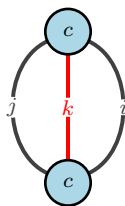


Figure 4. Theta diagram. All diagrams are trivalent networks without any loop, and vertices are the structure constants c_{ij}^k . Each vertex has indices i, j, k which are paired with their counterpart in another vertex. This pairing is done using g along edge of type k (colored red) and g^{-1} for type i and j .

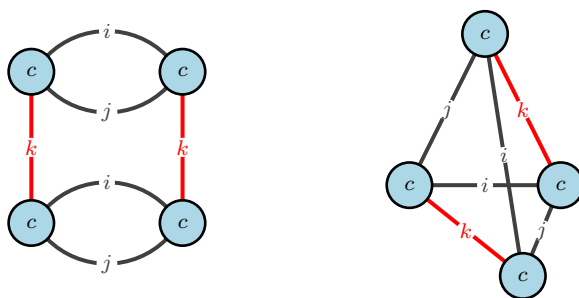


Figure 5. Tincan and Tetrahedron with some sample labeling. Red lines are labelled by g and black lines by g^{-1} .

- Two thetas
- Tincan
- Tetrahedron

For $m = 6$, there is a common coefficient $\frac{1}{6!}$.

- Three thetas
- Tincan and theta
- Tetrahedron and theta
- Prism
- $K_{3,3}$
- Extended tincan
- Necklace
- TetraTheta

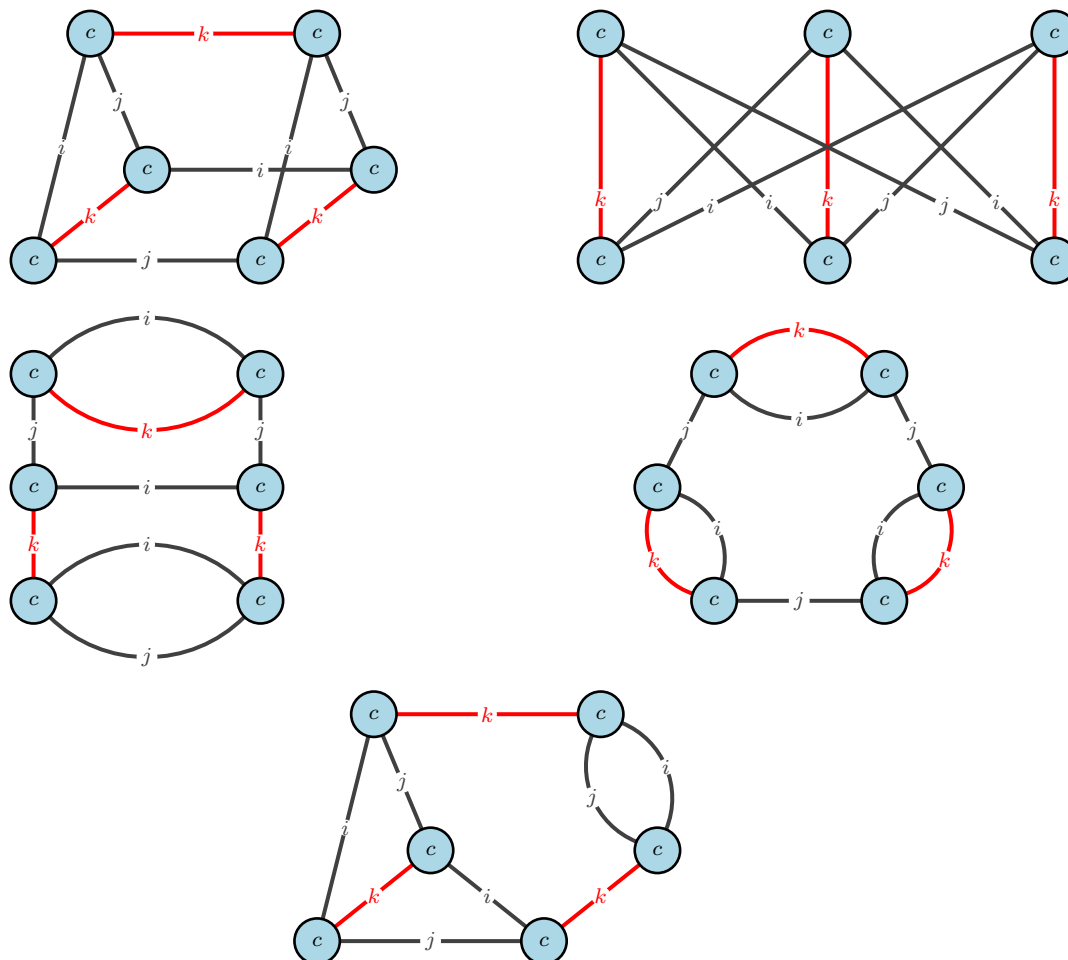


Figure 6. From left to right, top to bottom: Prism, $K_{3,3}$, Extended tincan, Necklace, and Tetra-Theta diagrams, each with some sample labeling. Red lines are labelled by g and black lines by g^{-1} .

Hence, the functions of study are

$$F_{26}(c, g, k) = \frac{1}{6!}(\text{sum of 6-vertex diagrams}) - \frac{k^2}{2!}(\text{the 2-vertex diagram Theta}) \quad (2.7)$$

$$F_{24}(c, g, k) = \frac{1}{4!}(\text{sum of 4-vertex diagrams}) - \frac{k}{2!}(\text{the 2-vertex diagram Theta}) \quad (2.8)$$

where the power of k for each m is $-\frac{3(4^n-1)+m}{2}$, thus the ratio yields the term k^2 in F_{26} and k in F_{24} for $m = 2$.

Remark 2.1. As mentioned in the introduction, and illustrated in experiments and in appendix C, the tensor contractions above are each convex with minimum at Id. This is why it is expected that a signed sum of these diagrams gives local minima around a local maximum at Id (figure 1).

In order to fix the volume by restricting to the space of metrics with $\det(g) = 1$, we add the term $(\det(g) - 1)^2$ with a high enough coefficient to the loss function. The more obvious normalization is a suitable fractional power of $\frac{1}{\det(g)}$. However, not only this power

$$R = -\frac{1}{2} \begin{array}{c} \text{c} \\ \text{Id}_{jk} \quad \text{Id}_{kj} \quad g_{ii'} \\ \text{c} \end{array} - \frac{1}{4} \begin{array}{c} \text{c} \\ g_{jj'} \quad g^{kk'} \quad g_{ii'} \\ \text{c} \end{array}$$

Figure 7. Ricci scalar curvature.

is different for networks with different number of vertices, but also, such an expression leads to more severe numerical round-off errors because of the fractional power, as well as a more involved formula for the gradient than $(\det(g) - 1)^2$ (see remark 3.4).

Furthermore, in many instances, the value of F_{26}, F_{24} is too high, which makes the use of scaling factors for the functional and $\det(g)$ necessary in order to avoid numerical instability, sometimes leading to $\det(g)$ being sent to zero.

The general form of the loss function is:

$$L_{26}(c, g, k) = r_1^{-1} F_{26}(c, g, k) + r_2 (\det(g) - 1)^2, \quad (2.9)$$

$$L_{24}(c, g, k) = r_1^{-1} F_{24}(c, g, k) + r_2 (\det(g) - 1)^2, \quad (2.10)$$

where $r_1 \geq 1, r_2 \gg 1$. The term *solution* will be exclusively used to refer to the critical points of the loss function found through gradient descent, which experimentally (through many random small perturbations) have been checked to be local minima. This last check is necessary, as we are looking for local minima of F_{26}, F_{24} inside the $\det(g) = 1$ subspace, and it is not immediately clear that the set of such local minima is equal to that of L_{26}, L_{24} in the space of all metrics g .

2.1.2 Ricci scalar curvature

The Ricci scalar curvature for all unimodular Lie algebras can be computed as follows [16]:

$$R = -\frac{1}{2} \sum_{i,i',j,k} c_{ij}^k c_{i'k}^j g_{ii'} - \frac{1}{4} \sum_{i,j,k} \sum_{i',j',k'} c_{ij}^k c_{i'j'}^{k'} g_{ii'} g_{jj'} g^{kk'} \quad (2.11)$$

which in diagrammatic notation figure 7.

For $\mathfrak{su}(2^n)$, this can be simplified to

$$R = \frac{1}{2} \sum_{i,j,k} (c_{ij}^k)^2 g_{ii} - \frac{1}{4} \sum_{i,j,k} \sum_{i',j',k'} c_{ij}^k c_{i'j'}^{k'} g_{ii'} g_{jj'} g^{kk'}. \quad (2.12)$$

To find the critical points, we need to minimize $\|\nabla R\|$ which can be computed explicitly using the formula above. Minimizing $\|\nabla R\|$ through gradient descent is too slow, likely because the functional has a shape very close to flat. In this kind of situation, in machine learning applications, one method that generally provides better results is the *evolutionary* method. The idea is to do an approximate gradient descent, but much faster and with more potential in finding lower minima in almost flat-shaped graphs.

Instead of computing a gradient, we perturb by small amount the given point (*parent*) and compute the value of M many of its perturbations (*children*). The child with the smallest evaluation is chosen as the new parent. In our application, we choose $M = 20$ and perturb each entry of the parameter by using a uniform distribution on $(1, 1/30)$.

2.2 Initialization of the gradient descent

Different initializations of the metric g lead to the discovery of different local minima of the loss function. In addition, some forms of initializations are preserved under gradient descent. We need to discuss some terminology and lemmas to discuss the initializations.

Definition 2.1. Given a metric g , the *degeneracy pattern* is the set $\{d_1, \dots, d_t\}$ of the dimensions of its eigenspaces. An *ordered* degeneracy pattern is a tuple (d_1, \dots, d_t) where the corresponding eigenvalues $0 < \lambda_{d_1} < \dots < \lambda_{d_t}$ are in ascending order.

As an example, the ordered *penalty* degeneracy pattern (or just ordered penalty pattern) is $(6, 9), (9, 27, 27)$ for $\mathfrak{su}(4), \mathfrak{su}(8)$, respectively.

The gradient flow preserves particular degeneracy patterns related to the **Clifford group**: given n qubits, the Clifford group $C_n = \{V \in U_{2^n} | V\mathbf{P}_n V^\dagger = \mathbf{P}_n\}$ is the group of unitary gates that normalize the Pauli group $\mathbf{P}_n = \{e^{i\theta\pi/2} \sigma_{i_1} \otimes \dots \otimes \sigma_{i_n} | \theta \in \{0, 1, 2, 3\}, i_j \in \{0, x, y, z\}\}$, where $\sigma_0 = \text{Id}$.

Let a subgroup $H < C_n$ act on $\mathbf{P}_n / \{\pm 1, \pm i\}$, where we identify words that are proportional to each other. Consider the orbits of H , which also provide a decomposition of the Pauli word basis \mathbf{PB}_n of $\mathfrak{su}(2^n)$, thus determining a degeneracy pattern. We have the following theorem.

Theorem 2.1 (Appendix B.3). *Let D be diagonal with a degeneracy pattern determined according to the orbits of a subgroup $H < C_n$ as explained above. The gradient flow preserves not only the diagonal form of the metric, but also the degeneracy pattern modulo the merging of the eigenspaces, e.g $\{d_1, d_2, d_3\}$ could become $\{d_1, d_2 + d_3\}$.*

Remark 2.2. As a direct corollary, gradient descent reaches a critical point (experimentally checked to be local minima) with that degeneracy pattern modulo merging.

We can now discuss different types of initialization to start the gradient descent with. All such initialization are normalized to have $\det = 1$, and determine the set of parameters on which gradient descent is performed.

- **DiagPerturbId**: A uniform random diagonal perturbation of the local maxima at identity, the set of parameters being the diagonal entries.
- **PenMetricOmega**: Penalty metric with gradient on the weight parameter $w = e^r$ in (1.1), which means the local minima are found in the space of penalty metrics. Thus, a solution may not necessarily be a local minimum in the space of all metrics (but see remark 2.3). This is to search for the minimum that the penalty metric can achieve and compare it to the lowest local minima found in other ways.

- **PenMetricPattern**: Penalty metric with gradient descent on w_1, \dots, w_n where w_i is the entry for Pauli word of weight i . All w_i are initialized using a random w , i.e. $w_i = w^i$ at the start. Once the local minimum is reached, we can observe how far the local minima found is from a penalty metric by comparing the ratios w_{i+1}/w_i .
- **GenPerturbId**: A general (nondiagonal) random Gaussian perturbation of identity, the set of parameters being entries of the matrix.
- **GenMetric**: A random metric, the parameters being entries initialized by standard normal distribution.

Remark 2.3. For $\mathfrak{su}(4)$, **PenMetricOmega** is effectively the same as the more general **PenMetricPattern**: there is only one ratio (w_2/w_1) to calculate and it gives the weight parameter w . On the other hand, a penalty pattern for $\mathfrak{su}(8)$ could experience a merging of eigenspaces or lead to a solution where the eigenvalues do not represent a penalty pattern (i.e. $w_2/w_1 \neq w_3/w_2$), and all local minima found with this initialization for $\mathfrak{su}(8)$ are as such.

2.3 Adam optimization algorithm

The Adam optimization algorithm [18] has become one of the most used algorithms for deep learning and more generally machine learning applications. Similar to the vanilla stochastic gradient descent, it is used to update parameters of a functional iteratively. However, it has many advantages compared to the stochastic gradient descent, and has been experimentally shown to lead to lower minima faster compared to other stochastic optimization algorithms [18]. Most important to our application, it satisfies some desirable characteristics:

- Computationally efficient with little memory requirements.
- Invariant to diagonal rescale of the gradients.
- Appropriate for problems with sparse gradients.
- Hyper-parameters have intuitive interpretation and typically require little tuning.

In contrast to stochastic gradient descent, Adam has a learning rate for each parameter which is separately adapted during learning. The name Adam comes from *adaptive moment estimation*, as it computes individual adaptive learning rates for each parameter from estimates of first and second moments of the gradients. This helps with sparse gradients situations and cases where the gradient is highly noisy (like in nonstationary settings). For our purpose, we would like an algorithm that avoids saddle points and has a better chance of finding the more interesting local minima by not getting stuck in small basins.

Figure 8 shows the implementation of the algorithm. The hyperparameters of Adam with their default values are described below.

- α (default 0.001): The learning rate (LR) or stepsize. The proportion that weights are updated. We prefer small values to have a more thorough search and find more local minima.

```

Require:  $\alpha$ : Stepsize
Require:  $\beta_1, \beta_2 \in [0, 1)$ : Exponential decay rates for the moment estimates
Require:  $f(\theta)$ : Stochastic objective function with parameters  $\theta$ 
Require:  $\theta_0$ : Initial parameter vector
 $m_0 \leftarrow 0$  (Initialize 1st moment vector)
 $v_0 \leftarrow 0$  (Initialize 2nd moment vector)
 $t \leftarrow 0$  (Initialize timestep)
while  $\theta_t$  not converged do
   $t \leftarrow t + 1$ 
   $g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1})$  (Get gradients w.r.t. stochastic objective at timestep  $t$ )
   $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$  (Update biased first moment estimate)
   $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$  (Update biased second raw moment estimate)
   $\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$  (Compute bias-corrected first moment estimate)
   $\hat{v}_t \leftarrow v_t / (1 - \beta_2^t)$  (Compute bias-corrected second raw moment estimate)
   $\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$  (Update parameters)
end while
return  $\theta_t$  (Resulting parameters)

```

Figure 8. Adam algorithm as in [18].

$k = 100, 200$	DiagPerturbId	GenPerturbId	PenMetricPattern	
r_1	1	1	1	
r_2	10^6	10^6	10^6	

$k = 500, 1000$	DiagPerturbId	GenPerturbId	PenMetricOmega	PenMetricPattern
r_1	10^4	10^4	10^4	10^4
r_2	10^3	10^4	10^3	10^3 or 10^4

Table 1. The scaling factors for $L_{2,6}$ for $\mathfrak{su}(4)(k = 100, 200), \mathfrak{su}(8)(k = 500, 1000)$.

- β_1 (default 0.9): The exponential decay rate for the first moment estimates. This hyperparameter like the next ones required no tuning.
- β_2 (default 0.999): The exponential decay rate for the second-moment estimates. In regions with a small or sparse gradients, β_2 should be close to 1.
- ϵ (default 10^{-8}): This is to avoid any division by zero in the implementation.

2.3.1 Scheduling and hyperparameters

LR decay can be used with Adam. In all cases, LR is 10^{-3} but is halved during the gradient descent every 20000 steps, as the algorithm struggles to converge in its final stages, as it jumps over and back on lower minima basins due to larger than appropriate LR. All other hyperparameters are taken with their default values. Scaling factors mostly depend only on the type of the integral and value of k . The scaling factors chosen for each initialization are given in tables 1 and 2. Generally, we prefer to choose higher r_1 to scale F_{26} or F_{24} if they attain very high values, which happens as k grows larger. If they attain high values but we can still be confident about the accuracy of the computations, r_1 is set to 1 while r_2 is chosen high enough (10^6) to avoid $\det(g)$ vanishing.

$k = 100, 200$	DiagPerturbId	GenPerturbId	PenMetricPattern
r_1	10	10	10
r_2	10^3	10^3	10^3

$k = 1000$	DiagPerturbId	GenPerturbId	PenMetricPattern
r_1	10	10	10
r_2	10^3	10^3	10^3

Table 2. The scaling factors for $L_{2,4}$ for $\mathfrak{su}(4)(k = 100, 200)$, $\mathfrak{su}(8)(k = 1000)$.

3 Optimization results and conclusions

We shall make a few remarks before discussing the results.

Remark 3.1. In tables above, **GenMetric** is absent. It was observed through experiments that they

- did not give any new local minima (in terms of degeneracy pattern), or
- they did not converge, as either the values of the functionals were generally too high for us to have confidence in the numerics, or the complicated scheduling necessary in the hyperparameter tuning made it too hard to have a stable gradient descent.

Remark 3.2. As mentioned in remark 2.3, for $\mathfrak{su}(4)$, “**PenMetricPattern = PenMetricOmega**”, but for $\mathfrak{su}(8)$, they are gradient descents in different spaces. In the case of $L_{2,4}$ (table 2), we could not collect any nontrivial result for $\mathfrak{su}(8)$ by using **PenMetricOmega**, hence why its column is absent.

Remark 3.3. Our experiments involved two values of k although the values in-between were tried to some extent, and did not deliver any new results. If a degeneracy pattern reappears for higher values of k , it is always the case that eigenvalues are further apart.

We assess our solutions using the following type of evidences, listed in the order used in the paper.

- Finding Lie subalgebra structure by computing the brackets of (a collection of) eigenspaces.
- We propose two different ways for providing evidence/proof regarding the existence of a map j as defined in 1.1. In this paper, we explore the first one to some extent:
 1. An evidence for the existence of j can be found by looking for what is called *pattern breaking*. First one finds a partition of the given degeneracy pattern that has the same dimensions as the penalty pattern; e.g. for $\mathfrak{su}(4)$, the pattern $(3, 1, 1, 8, 2)$ (see further below) has one partition such as $(\{3, 1, 2\}, \{1, 8\})$ where the dimensions in each set add up to the penalty pattern $(6, 9)$. Next, we check if the Lie algebra generated by the set of eigenspaces $(\{3, 1, 2\}$ in the example) corresponding to the weight one Pauli words is isomorphic to the Lie subalgebra

$\mathfrak{su}(2) \oplus \dots \oplus \mathfrak{su}(2)$ ($n = 2$ times in the example). This ensures that there exists a map j , at least for the restriction to weight one Pauli words. The final step, not carried out in this work, would be to see if the rest of the eigenvectors can also be mapped under such a j to a local tensor product form.

2. A more direct approach is a gradient descent in the space of all Lie algebra isomorphisms j^{-1} . We can check whether $j^{-1}(H_i)$ is in tensor product form by computing the entanglement entropy and aim to minimize the overall entanglement entropy of $j^{-1}(H_i)$ s.

Consider the case of $\mathfrak{su}(4)$ where we would like to measure the distance of $j^{-1}(H_i)$ to a decomposition into a tensor product $O_1 \otimes O_2$. Fixing j (or J), determines an isomorphism of the underlying \mathbb{C}^4 (on which $\mathfrak{su}(4)$ are the skew-Hermitian transformations) into $\mathbb{C}^4 \cong \mathbb{C}_1^2 \otimes \mathbb{C}_2^2$. Think of $j^{-1}(H_i)$ as a vector in $\mathbb{C}_1^2 \otimes \mathbb{C}_2^2 \otimes (\mathbb{C}_1^2)^* \otimes (\mathbb{C}_2^2)^*$ so that the partial trace $\rho_i = \text{Tr}_{\mathbb{C}_2^2}(j^{-1}(H_i)) \in \mathbb{C}_1^2 \otimes (\mathbb{C}_2^2)^*$ under the natural map.

By computing the entanglement entropy of $j^{-1}(H_i)$ which is $S(j^{-1}(H_i)) := S_i = \text{Tr}(\rho_i \log \rho_i)$, we can define the loss function as $\sum_{i=1}^{\dim \mathfrak{su}(4)=15} S_i$ and aim to minimize it over $\{j\}$. If $\sum S_i$ can be made “close” to zero, then we conclude that $\{H_i\}_{i=1}^{\dim \mathfrak{su}(4)=15}$, is **kaq** up to some error (depending on how “close” to zero); i.e. the eigenvectors for g have attained something close to a tensor product form. Note that for higher number of qubits, like in $\mathfrak{su}(8)$, this test needs to be performed successively to prove decomposition into $O_1 \otimes O_2 \otimes O_3$.

- If we find a nondegenerate eigenvalue, we make a little spectra test on the corresponding skew-Hermitian matrix, to find if the spectrum can rule out that the matrix is not in the form of a tensor product.
- In the case of **PenMetricPattern**, we make a direct inspection and look at the ratios of eigenvalues of different weight spaces (remark 2.3).
- In the case of **PenMetricOmega**, we check if the solution is a local minima in the space of all metrics, and if the weight parameter is smaller than one.

We used the **TensorNetwork** package to compute the diagrams and **PyTorch** to do gradient descent using Adam. For the hardware, we used one NVIDIA Tesla V100 GPU. Most simulations took at most 2 hours to converge, with the diagonal initializations converging faster.

Remark 3.4. PyTorch uses Automatic differentiation (Autograd) to compute gradient of a function f . It does so by first building a computational tree of f using simpler functions, such as linear, trigonometric, etc., for which the gradient formula is built into the program and can be calculated exactly. Then, it computes the gradient using the chain rule.

As such, to make gradient descent computations faster on the computer, *we replace g with g^{-1}* , making sure that one type of edge (k) gets labelled by g^{-1} , and two labelled by g . *Thus, when we examine the results (eigenvalues of g), we will need to take the inverse again.* Note that as the parameters of our program are the entries of g , the computation

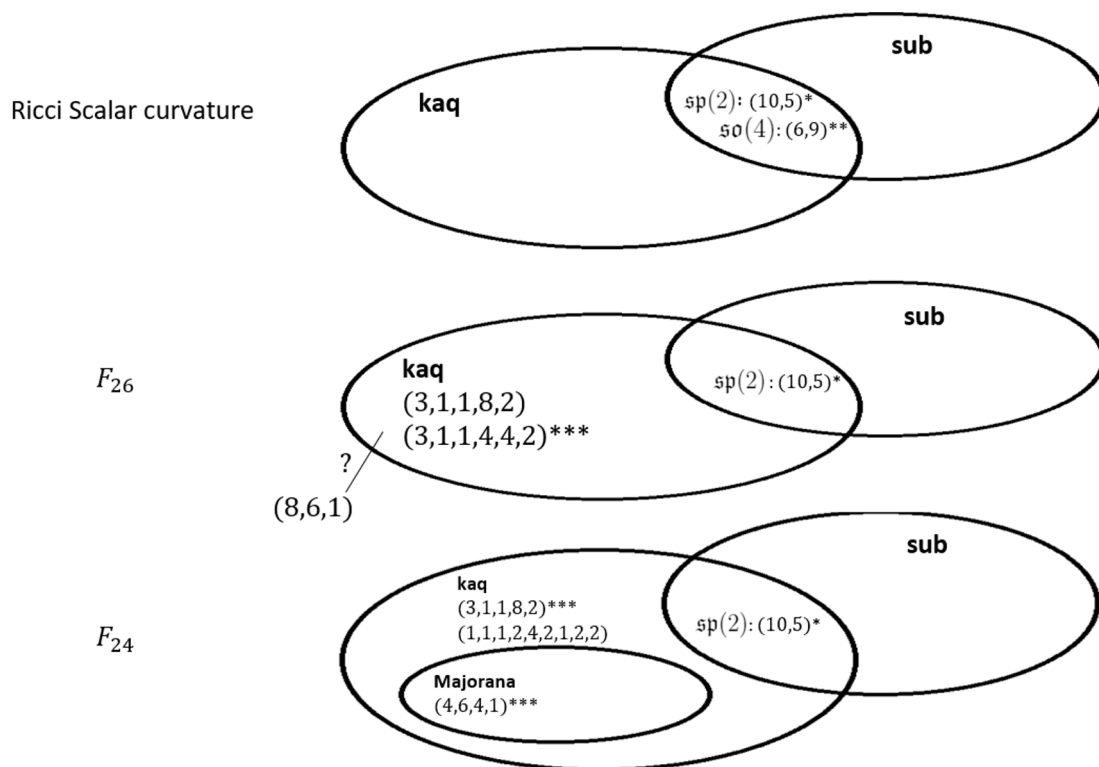


Figure 9. Patterns found for the $\mathfrak{su}(4)$ for each functional and their possible classification. *: all $\mathfrak{sp}(2)$ s are found in diagonal solutions. **: the penalty metric. ***: the ordered patterns are slightly modified given that the eigenvalues were numerically close; this is done to represent the ordered degeneracy pattern in a way that relates it to a previously found $((3, 1, 1, 8, 2))$ or well-known (Majorana) pattern. ?: the classification of the pattern $(8, 6, 1)$ is unclear.

for gradients of terms involving g is far simpler than those involving g^{-1} , as the formula for the latter involves many long expressions in terms of g entries (determinants of minors and division by $\det(g)$). Therefore, it makes sense to make such a replacement as it allows us to cut the number of edges labelled by g^{-1} in half (and double the number of those labelled by g).

3.1 DiagPerturbId and GenPerturbId results

There are different degeneracy patterns among the solutions of **DiagPerturbId** and **GenPerturbId**. We will list the (lowest) local minima ordered degeneracy patterns along with eigenvalues for the highest k (if it appears for both k), and provide explanation regarding the **kaq/sub**-ness of the metrics. In these lists, LM = local minima and LLM = lowest local minima. A summary of the results is presented in figures 9 and 10.

1. $L_{2,6}$ and $\mathfrak{su}(4)$ with $k = 100, 200$:

- LM: The second pattern occurs exclusively for $k = 100$ while the last two patterns do so for $k = 200$.

1. $(10, 5), (0.315, 10.079)$

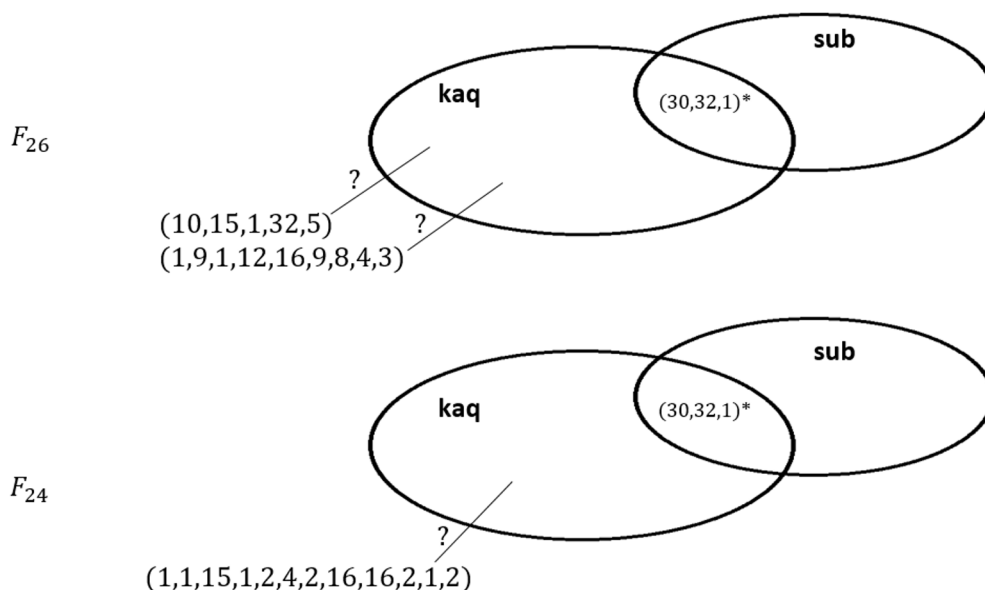


Figure 10. Patterns found for the $\mathfrak{su}(8)$ for each functional. *: the diagonal patterns lead to very similar $(30, 32, 1)$ pattern in each functional (see the text for details). ?: the classification of all the other patterns such as $(10, 15, 1, 32, 5)$ is unclear.

2. $(3, 1, 1, 8, 2), (0.650, 0.678, 0.686, 0.951, 3.417)$
 3. $(1, 3, 1, 4, 4, 2), (0.309, 0.311, 0.333, 0.381, 3.424, 10.526)$
 4. $(8, 6, 1), (0.396, 2.523, 44.247)$
- LLM for both k : $(10, 5)$ where the 10D eigenspace is $\mathfrak{sp}(2)$.

We found nondiagonal metrics for all patterns above, and diagonal metrics with the $(10, 5)$ pattern only (for both k). The second pattern is interesting as we find it is a *pattern breaking* of $(6, 9)$, where the 1D eigenspace with corresponding eigenvalue 0.686 along with the 2D and 3D eigenspaces, form a Lie algebra isomorphic to $\mathfrak{su}(2) \oplus \mathfrak{su}(2)$ (similar to Pauli words with weight one), while the 4×4 skew-Hermitian matrix in the other 1D eigenspace has the eigenvalue structure of a tensor product. For $k = 200$, some of solutions with pattern $(1, 3, 1, 4, 4, 2)$ have that property as well. Finally, the 6D degeneracy for the last pattern does not form a Lie algebra, and the matrix in the 1D eigenspace does not have the eigenvalue structure of a tensor product.

2. $L_{2,6}$ and $\mathfrak{su}(8)$ with $k = 500, 1000$:

- LM: Only the last pattern occurs exclusively for $k = 1000$.
 1. $(30, 32, 1), (0.341, 2.375, 9.901)$
 2. $(10, 15, 1, 32, 5), (0.3771, 0.3835, 0.4432, 1.3020, 2.8011)$
 3. $(1, 9, 1, 12, 16, 9, 8, 4, 3), (0.352, 0.359, 0.381, 0.671, 0.689, 1.197, 1.25, 12.5, 20.0)$
- LLM: $(10, 15, 1, 32, 5)$ for both k .

In diagonal solutions, only $(30, 32, 1)$ appears, while the other two patterns occur in non-diagonal solutions. We know by taking all brackets that the eigenvectors in $1 + 30$ (and 30 by itself) give a Lie algebra, but we have not classified it. The same property holds for $(10, 15, 1, 32, 5)$ where $1 + 5 + 10 + 15$ (and $5 + 10 + 15$) is a 31D (30D) Lie algebra.

Given that $(9, 27, 27)$ is the penalty pattern, it can only be related to $(1, 9, 1, 12, 16, 9, 8, 4, 3)$ if we were to observe some *pattern breaking* similar to the instance in $\mathfrak{su}(4)$. It is the case (up to $1e-3$ errors) that one of the two nines has structure constants similar to that of Pauli words with weight one, i.e. $\mathfrak{su}(2) \oplus \mathfrak{su}(2) \oplus \mathfrak{su}(2)$. However, the matrix in each of the 1D eigenspaces can not be a tensor product of three 2×2 matrices, although they can be a tensor product of a 4×4 and 2×2 matrix. Given that we are only approximating the functional, this suggests that maybe locality does not fully form and one should consider the possibility of partial **kaq**-ness where there are tensor products of **qunits** instead of (\mathbb{C}^2) qubits.

3. $L_{2,4}$ and $\mathfrak{su}(4)$ with $k = 100, 200$:

- LM:
 1. $(10, 5), (0.790, 1.602)$ only for $k = 100$.
 2. $(1, 3, 1, 8, 2), (0.347, 0.407, 0.448, 0.971, 10.921)$ for both k .
 3. $(6, 4, 1, 4), (0.383, 0.448, 1.645, 4.794)$ only for $k = 200$.
 4. $(1, 1, 2, 4, 2, 1, 2, 2), (0.365, 0.404, 0.548, 0.586, 0.653, 1.222, 2.118, 9.039)$ only for $k = 200$.
- LLM for both k : $(1, 3, 1, 8, 2)$.

We found diagonal metrics with the $(10, 5)$ pattern ($k = 100$) and $(6, 4, 1, 4)$ ($k = 200$). The latter corresponds to the Majorana fermion degeneracy pattern, where ψ_1, \dots, ψ_4 form a penalty pattern $(4, 6, 4, 1)$, but note the order is not the same. This could be due to the competition the functional feels between forming two local minima, the $\mathfrak{so}(4)$ local minima (same as the penalty pattern $(6, 9)$) and the Majorana pattern $(4, 6, 4, 1)$, explaining why 6 falls before 4 in the ordered pattern.

The pattern $(1, 1, 3, 8, 2)$ is related to the penalty pattern similar to $(3, 1, 1, 8, 2)$ in $L_{2,6}$ for $\mathfrak{su}(4)$. The case of $(1, 1, 2, 4, 2, 1, 2, 2)$ is yet unknown, however we can show numerically that it is not a pattern breaking of a penalty pattern.

4. $L_{2,4}$ and $\mathfrak{su}(8)$ with $k = 1000$:

- LM:
 1. $(30, 32, 1), (0.520, 1.620, 61.728)$
 2. $(1, 1, 15, 1, 2, 4, 2, 16, 16, 2, 1, 2), (0.543, 0.549, 0.552, 0.606, 0.742, 0.746, 0.776, 0.86, 1.268, 2.994, 3.416, 5.714)$
- LLM: $(1, 1, 15, 1, 2, 4, 2, 16, 16, 2, 1, 2)$.

Only $(30, 32, 1)$ is found among diagonal metrics and it is related to a Lie algebra similar to the pattern $(30, 32, 1)$ in $L_{2,6}$. The second pattern is also related to a similar Lie algebra as $1 + 1 + 1 + 1 + 2 + 2 + 2 + 2 + 4 + 15$ forms a 31D Lie algebra. The 1D eigenspaces are matrices with eigenvalue structure that can be realized by a tensor product (4×4 and 2×2). However, none of the patterns can be a breaking of the penalty pattern $(9, 27, 27)$.

3.2 PenMetricOmega and PenMetricPattern results

Using penalty metric initializations, local minima with such patterns are obtained. As noted in remark 2.3, we need to check if the eigenvalues are generated by a weight parameter which is smaller than one.

For $\mathfrak{su}(4)$ and both of $L_{2,4}, L_{2,6}$, the answer is always affirmative for the lowest local minima in **PenMetricPattern**. The loss value attained is $\geq 95\%$ of the lowest loss value found (in LLMs in the previous part), and the weight parameter is pushed more towards zero as k increases.

For $\mathfrak{su}(8)$, there are local minima in **PenMetricPattern**, with $\geq 95\%$ of the loss value of LLM previously listed, which have the same *ordered* degeneracy pattern as the penalty pattern. However the two ratios $w_2/w_1, w_3/w_2$ do not match; the relative difference is about 30%, more than what could be attributed to numerical errors (although it could be due to missing terms in our expansion of the integral). The lowest local minima are given by degeneracy patterns with order $(27 + 9, 27)$ for $L_{2,6}$, where weight one and three have merged, and $(27, 27, 9)$ with weight two being the smallest for $L_{2,4}$. Finally, The solutions with **PenMetricOmega** initialization achieved $\sim 90\%$ of the LLM loss value, but none were local minima in the space of all metrics.

3.3 Ricci scalar curvature results

As mentioned in the introduction, our search for minimizing $\|\nabla R\|$ yielded the results in [16]. Thus, only diagonal solutions in $\mathbf{sub}_2 \cap \mathbf{kaq}_2$ were found corresponding to $\mathfrak{so}(4)$ (6D) and $\mathfrak{sp}(2)$ (10D) Lie subalgebras.

4 Summary and outlook

There are many directions in which this work can be refined. One could start by investigating the nature of those 31D Lie subalgebras in $\mathfrak{su}(8)$, while collecting more data for **PenMetricOmega** and more importantly a larger set of (especially higher) values of k . Results in this work and preliminary unreported results for $k > 200$ for $\mathfrak{su}(4)$, suggest that new patterns emerge for higher values of k , which are consolidation of previous ones known for lower k s. Perhaps there is a tendency for energy-close eigenspaces to cluster as k increases while the energies of truly distinct eigenspaces grow further apart. We can then ask about the low-energy limit, and what eigenspaces remain.

With regards to **kaq**-ness, we still do not have a mechanism/theory to prove it given a certain degeneracy pattern, other than checking the Lie subalgebra structure. By doing so, we are ignoring the possibility that a solution is **kaq** but does not have any Lie subalgebra

among (any proper collection of) its eigenspaces. One way to address this issue is a gradient descent in the space of all Lie algebra isomorphisms j^{-1} , as mentioned in section 3.

Also, we did not use any post-processing technique like Padé approximates, which could provide a more accurate estimation of the functional.

Another direction is the study of complex fermionic perturbed Gaussian integrals which could reveal similar patterns.

Finally, we would like to have similar numerical simulations for other $\mathfrak{su}(N)$ s like $\mathfrak{su}(6)$, to see if **kaq**-ness would still manifest itself in the form of $\mathbb{C}^3 \otimes \mathbb{C}^2$, showing that there is nothing special about $N = 2^n$ regarding decomposition to local tensor factors.

Acknowledgments

We would like to thank Carl Bender, Dror Bar-Natan, Scott Morrison, and Xiaoliang Qi for stimulating conversations. This research was done as part of an internship of the second named author at Microsoft, and the writing of this paper was finished while he was supported by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Research, Innovation and Science.

A The structure constants can only distinguish weights

We are interested in exploring the symmetries of the *anti-commutativity* graph of Pauli words $G_{\mathbf{PB}_n}$ defined below.

Definition A.1. The graph of Pauli words $G_{\mathbf{PB}_n}$ has vertices \mathbf{PB}_n , and two vertices are connected when they anti-commute.

As a simple initial observation, note that all Pauli words have identical spectra, half the eigenvalues are 1 and half -1, thus they are all in the same adjoint orbit, so it turns out to be a rather subtle matter to find asymmetric structures in $G_{\mathbf{PB}_n}$ which have any hope of distinguishing, for example, high weight from low weight Pauli words. In this appendix, we initiate the study of such subtle asymmetries in the hope that it will allow us in the future to reverse-engineer functionals which break to penalty metrics.

If there are asymmetries in this graph, especially among the vertices with different weights, one can hope to derive a metric sensitive to such asymmetries. Recall a graph is k -ultrahomogeneous if any given isomorphism between two of its induced subgraph with at most k vertices can be extended to an automorphism of the graph. As the following fact demonstrates, $G_{\mathbf{PB}_n}$ is symmetric.

Fact A.1. $G_{\mathbf{PB}_n}$ is a 2^{2n-1} -regular 2-ultrahomogeneous graph. The 2-ultrahomogeneity is facilitated by the Clifford group transformations.

Remark A.1. Following this fact, one has graph automorphisms of $G_{\mathbf{PB}_n}$ where vertices of different weights are sent to each other. Moreover, it is a standard Lie algebra fact

that the Lie algebra automorphisms of $\mathfrak{u}(N)$ all come from conjugations by $U(N)$ and a complex conjugation (\mathbb{Z}_2 of the Dynkin diagram), hence by definition of C_n , all Lie algebra automorphisms that preserve \mathbf{P}_n come from the Clifford group and complex conjugation, and the latter acts trivially on \mathbf{PB}_n .

Remark A.2. The fact above also means that a simple preference for more commutativity (and less anti-commutativity or less $\| [a, b] \|$) can not be the sole reason behind the possible emergence of a penalty metric. Hence minimizing simple functionals like $\sum \| [a, b] \|$ will not deliver the desired result.

However, we will show that the additional structure imposed by c_{ij}^k on this graph is exactly what is needed to distinguish the vertices with different weight. We recall a fact that will be useful for our next theorem.

Fact A.2. For every $a, b \in \mathbf{P}_n$, either $[a, b] = 0$ or $ab = -ba$. Hence, $c_{ij}^k = 0, \forall k$ or $c_{ij}^k \neq 0$ for a unique k , implying $ij = \frac{c_{ij}^k}{2}k$, and c_{ij}^k satisfies skew-symmetry, i.e. $c_{ik}^j = -c_{ij}^k$, etc.

We ask if it is possible to have a graph automorphism which preserves the structure constants c_{ij}^k while not preserving weights (see (1.1) for the definition of weight)? The negative answer below gives meaning to the title of this section and provides the theoretical motivation for the possibility of penalty metric to emerge from optimization of Gaussian perturbed integrals:

Theorem A.1. *Given any $\phi \in \text{Aut}(G_{\mathbf{PB}_n})$, $c_{\phi(i)\phi(j)}^{\phi(k)} = c_{ij}^k, \forall i, j, k$ if and only if ϕ is a permutation of tensor factors composed with some local automorphism, i.e. $\phi \in \text{Aut}(G_{\mathbf{PB}_1})^{\otimes n} \times S_n$, consequently preserving the weight of each word.*

Proof. In the following, any Pauli word basis element like $-iX \otimes 1 \otimes \dots \otimes 1$ is represented a $X1\dots 1$, where tensors and $-i$ are dropped for convenience. We wish to prove that $\phi(X1\dots 1)$ has weight one. Assuming otherwise, without loss of generality, possibly after composing with a suitable automorphism in $\text{Aut}(G_{\mathbf{PB}_1})^{\otimes n} \times S_n$, we have $\phi(X1\dots 1) = X\dots X1\dots 1$ where there are $l > 1$ many X s.

Consider the triangle ($i = X1\dots 1, j = Y1\dots 1, k = Z1\dots 1$) which has $c_{ij}^k = +2$. Sitting at the opposite side of each of i, j, k with respect to the triangle edges, and not connected to them, we have the subgraphs $S_i = X\mathfrak{su}(2^{n-1}), S_j = Y\mathfrak{su}(2^{n-1}), S_k = Z\mathfrak{su}(2^{n-1})$, respectively. Furthermore, every vertex $s_j \in S_j$ is connected to i (and k), and has a counterpart $i.s_j = s_k \in S_k$ such that $c_{is_j}^{s_k} = +2$. There exists a similar structure for j, S_k, S_i and k, S_i, S_j . Finally, there exists a fourth subgraph $S_{ijk} = 1\mathfrak{su}(2^{n-1})$ not connected to the triangle ijk . This gives the structure of the whole $G_{\mathbf{PB}_n}$ when viewed from this triangle (see figure 11).

Using fact A.1, Given any triangle ijk with $c_{ij}^k \neq 0$, the previous description of $G_{\mathbf{PB}_n}$ holds for any such i, j, k , with the exception of the fact about $c_{is_j}^{s_k} = +2$ for all s_j, s_k corresponding pairs. This means all triangles with $c_{ij}^k \neq 0$ have three subgraphs corresponding to each vertex and not connected to them, and each vertex is connected to the other two subgraphs completely. Also, there is a fourth subgraph not connected to the triangle.

As a result, if $\phi(X1\dots 1) = X\dots X1\dots 1$, since ϕ is a graph automorphism, it must send the triangle and structure described above to that of ($\phi_x = X\dots X1\dots 1, \phi_y =$

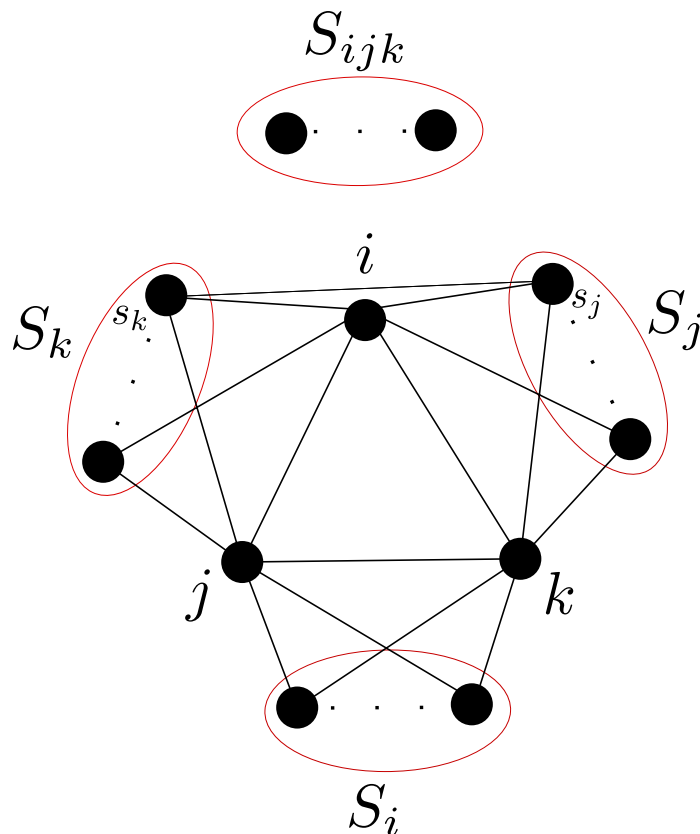


Figure 11. For each triangle with $ij = -ji = k$, one can partition the vertices into four sets S_i, S_j, S_k, S_{ijk} , where i is connected to all vertices in S_j, S_k (and similarly for j, S_i, S_k and k, S_i, S_j) and S_{ijk} is not connected to any of i, j, k . Further, for each $s_j \in S_j$, there is a corresponding vertex $s_k \in S_k$ such that $c_{is_j}^{s_k} \neq 0$ and $is_j = \frac{c_{is_j}^{s_k}}{2} s_k$.

$\phi(Y1\dots 1), \phi_z = \phi(Z1\dots 1)$). Therefore, as ϕ preserves c , we must have $c_{\phi_x s_{\phi_y}}^{s_{\phi_z}} = +2$ for all corresponding pairs $s_{\phi_y}, \phi_x \cdot s_{\phi_y} = s_{\phi_z}$. This implies that vertices of S_{ϕ_y} are Pauli words where the first l letters have odd many Ys and even many Zs. However, since $c_{\phi_x \phi_y}^{\phi_z} = +2$, the same is true about ϕ_y . As ϕ_y is not connected to S_{ϕ_y} , this means the set of Pauli words where the first l letters have odd many Ys and even many Zs must have an *isolated* vertex. In its first $l > 1$ letters,

- if ϕ_y has at least three Ys, then one can exchange two of those with an X and a 1 , to find a neighbor of ϕ_y in S_{ϕ_y} .
- If instead ϕ_y has one Y in its first l letters, then the whole word ϕ_y has either
 - two Zs in its first l letters, in which case we replace them by X and 1 , or
 - an X in its first l letters, in which case we replace Y by 1 and X by Y , or
 - an X or Y or Z outside of its first l letters, in which case we replace it by Y, X or Z ,

and in all these cases we find a neighbor in S_{ϕ_y} .

The only remaining case is when $\phi_y = Y1\dots 1$. Applying the same argument to ϕ_z (this time with odd many Z s and even many Y s) implies $\phi_z = Z1\dots 1$, and since $\phi_x \cdot \phi_y = \phi_z \implies \phi_x = X1\dots 1$.

This proves that ϕ has to preserve the weight one Pauli words and modulo some local automorphism and permutation, that $\phi(X1\dots 1) = X1\dots 1$. As S_{ϕ_y} must have odd many Y s in its first $l = 1$ letters and $Y1\dots 1$ is the only isolated vertex among such words, we have $\phi_y = Y1\dots 1$, and similarly $\phi_z = Z1\dots 1$. Restricting ϕ to each of the four subgraphs $S_i = X\mathfrak{su}(2^{n-1}), S_j = Y\mathfrak{su}(2^{n-1}), S_k = Z\mathfrak{su}(2^{n-1}), 1\mathfrak{su}(2^{n-1})$, the proof follows by induction. \square

Remark A.3. Let $\phi \in C_n$. For any $a \in \mathbf{PB}_n$, one can write $\phi(a) = \eta_a \phi_a$, for some $\phi_a \in \mathbf{PB}_n, \eta_a \in \{\pm 1\}$ (notice $\eta_a \neq \pm i$ as a, ϕ_a are skew-Hermitian). Then, if $[a, b] \neq 0$, for some unique $d \in \mathbf{PB}_n$:

$$ab = \frac{c_{ab}^d}{2}d \implies \phi(a)\phi(b) = \frac{c_{ab}^d}{2}\phi(d) \implies c_{\phi_a\phi_b}^{\phi_d} = \frac{\eta_c}{\eta_a\eta_b}c_{ab}^d.$$

The theorem above states that there is no $\phi \in C_n$, for which $\frac{\eta_d}{\eta_a\eta_b} = 1$ whenever $c_{ab}^d \neq 0$.

B Gradient descent and degeneracy patterns

In this section, we prove theorem 2.1. First, we show that gradient descent preserves diagonal metrics.

Theorem B.1. *Given any trivalent tensor network $T(c, g)$ without loops, with vertices labelled by c (structure constants of Pauli word basis), and edges labelled by g, g, g^{-1} for type i, j, k edges, we have*

$$\frac{\partial T}{\partial g_{ab}}|_D = 0,$$

for all diagonal metrics D and $a \neq b$.

Proof. To compute a tensor contraction, one selects indices at the two ends of each edge, e.g. for an edge of type i the selection i, i' yields a term like $g_{ii'}c_{i-}^-c_{i'-}^-$.

Thus, differentiating such an expression is done by selecting any edge and choosing labels a, b on the two ends, while the rest of the edges are labelled using the diagonal metric D (see figure 12 for more details). We wish to show that if $a \neq b$, any such network contracts to zero. To do so, we prove the structure constants multiplier $\prod_{i,j,k \text{ labels on edges}} c_{ij}^k$ obtained from any label selection over the edges, is zero if *exactly* one edge is picked nondiagonally, like in a, b .

Without loss of generality, assume that a, b is chosen on an edge of type i . Edges of type j and i form a cycle decomposition of the graph. Consider the product of the indices over those edges; as every edge has two equal indices on both ends (except for a, b), the product is some scalar times $a.b$.

On the other hand, if the structure constants multiplier is nonzero, it means at every vertex, i, j, k were chosen such that $c_{ij}^k \neq 0$. Furthermore, in that product, every i is

$$\frac{\partial T}{\partial g_{ab}} \Big|_{g=D} = E_{ab} \overset{c}{\circ} \overset{c}{\circ} D^{-1} D + D \overset{c}{\circ} \overset{c}{\circ} D^{-1} E_{ab} - \frac{\det(D_{a,b})}{\det(D)} D \overset{c}{\circ} \overset{c}{\circ} E_{ba} D$$

Figure 12. Differentiating the Theta diagram at a diagonal metric D . Differentiating the i or j edges is easy, as it just requires replacing the matrix by the elementary matrix E_{ab} . Note the last term requires computing $\frac{\partial g^{-1}}{\partial g_{ab}}$ at D . Considering the entries, we need to compute $(-1)^{m+n} \frac{\partial \det(g_{mn})}{\partial g_{ab}}$, where g_{mn} are the minors. It is straightforward to show that evaluating this expression at a diagonal metric D leaves a nonzero term only for $m = b, n = a$. The matrix $D_{a,b}$ is the diagonal matrix obtained by removing the two a, b -th rows and the two a, b -th columns.

followed by a j giving $\frac{c_{ij}^k}{2} k$. The k s are each repeated twice as they appear twice on each edge. As a result, the product should be a scalar, but $a \neq b$, so $a.b$ is not a scalar and we reach contradiction. \square

Corollary B.2. *Starting at a diagonal metric, gradient descent on loss functions defined in (2.9), (2.10) stays in the diagonal metrics space.*

Proof. Both (2.9), (2.10) are linear combinations of tensor networks, and the $(\det(g) - 1)^2$ term can be easily checked to have the same property under gradient descent. \square

Recall how a subgroup $H < C_n$ defines a degeneracy pattern 2.1.

Theorem B.3. *Let D be diagonal with a degeneracy pattern determined according to the orbits of a subgroup $H < C_n$. The gradient flow of loss functions (2.9), (2.10) preserves not only the diagonal form of the metric, but also the degeneracy pattern modulo the merging of the eigenspaces, e.g $\{d_1, d_2, d_3\}$ could become $\{d_1, d_2 + d_3\}$.*

Proof. We previously showed that D stays diagonal under gradient flow. Next, we need to show that entries corresponding to the same eigenspace stay equal, i.e. have equal gradients. It suffices to prove so for every tensor network T ; as the determinant term $(\det(g) - 1)^2$ has the same property, the statement would follow.

Assume $g_{aa} = g_{bb}$ where there is an $h \in H$ such that $h(a) = \eta_a b$ where $\eta_a \in \{\pm 1\}$. To show $\frac{\partial T}{\partial g_{aa}} = \frac{\partial T}{\partial g_{bb}}$, we use the same picture as in the previous theorem. Every tensor contraction in the summation giving $\frac{\partial T}{\partial g_{aa}}$ is given by fixing both ends of some edge e with label a . This tensor is sent to an equal contraction in $\frac{\partial T}{\partial g_{bb}}$ by simply applying h on the tensor. Indeed, since $g_{xx} = g_{h(x)h(x)}$ for any x , the coefficients involving g entries are equal. As for the structure constants multiplier, since $h(i) = \eta_i h_i$ for some h_i in the Pauli word basis, we need to show that

$$\prod_{i,j,k \text{ selected labels}} c_{ij}^k = \prod_{i,j,k \text{ selected labels}} c_{h_i h_j}^{h_k}$$

This is true for *any* $h \in C_n$. Indeed, as computed in remark A.3,

$$c_{h_i h_j}^{h_k} = \frac{\eta_k}{\eta_i \eta_j} c_{ij}^k,$$

and since in the product $\prod_{i,j,k \text{ selected labels}} c_{ij}^k$ every label appears twice and $\eta_x \in \{\pm 1\}, \forall x$, we are left with the equality of the two products. \square

Remark B.1. While on one hand, this theorem implies that one can find local minima with a penalty degeneracy pattern by simply initializing at such a pattern, it also means that the penalty pattern is not special in that regard, as we can find local minima with many other degeneracy patterns.

C Individual diagrams are convex

Experiments show that the individual tensor networks used in the asymptotic expansion of (1.7) are all positive convex functions with unique minimum at $g = \text{Id}$. We can provide some theoretical evidence in this regard for some of the diagrams.

We start with the Theta diagram $\Theta(c, g)$. Assume $g = UDU^\dagger$ where D is diagonal and U unitary. We wish to show $\min_D \Theta(c, UDU^\dagger) = \Theta(c, \text{Id})$.

We can rewrite the expression above as $\min_D \Theta(c_U, D) = \Theta(c_U, \text{Id}) = \Theta(c, \text{Id})$, where the edges of tensor c with the tensors U (or U^\dagger) are contracted. We have

$$\Theta(c_U, D) = \sum_{i,j,k} |c_{U(i)U(j)}^{U(k)}|^2 d_i d_j d_k^{-1}.$$

Writing $d_i = e^{w_i}$, $\Theta(c_U, D)$ becomes a positive sum of exponentials of affine functions in w_i , hence convex. By using the arithmetic-geometric mean inequality, the minimum can be seen to occur at $g = \text{Id}$. As previously mentioned, experiments have shown that this minimum is unique.

We can prove similar results for other diagrams such as the Tincan, by bringing the contraction in the form written above for $\Theta(c_U, D)$, through successive minimization over some edges, while others are assumed fixed and once contracted give two equal tensors being contracted over the same indices, same as in Θ .

For the Tincan $T(c, g)$, the starting idea is similar, fixing U in the diagonalization of g and letting D be the variable. However, at first, the vertical edges are also assumed to be fixed, meaning the variable is D on the 4 vertical edges. The contraction of tensors along the vertical lines gives two 4-valent tensors A, B connected with D (or D^{-1}) along each edge. The two tensors are equal due to the symmetry of the diagram and the previous contraction being done along the same index. Hence, we obtain an expression similar to the one for $\Theta(c_U, D)$, implying the minimum is attained at $D = \text{Id}$ on the horizontal edges, given the vertical edges. Next step is to consider D on the vertical edges as the variable and the proof follows similarly.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] K.G. Wilson, *The Renormalization Group: Critical Phenomena and the Kondo Problem*, *Rev. Mod. Phys.* **47** (1975) 773 [INSPIRE].
- [2] M.A. Nielsen, *A geometric approach to quantum circuit lower bounds*, [quant-ph/0502070](#).
- [3] M.A. Nielsen, M.R. Dowling, M. Gu and A.C. Doherty, *Quantum computation as geometry*, *Science* **311** (2006) 1133.
- [4] M.R. Dowling and M.A. Nielsen, *The geometry of quantum computation*, *Quant. Inf. Comput.* **8** (2008) 861.
- [5] A.R. Brown, L. Susskind and Y. Zhao, *Quantum Complexity and Negative Curvature*, *Phys. Rev. D* **95** (2017) 045010 [[arXiv:1608.02612](#)] [INSPIRE].
- [6] A.R. Brown and L. Susskind, *Second law of quantum complexity*, *Phys. Rev. D* **97** (2018) 086015 [[arXiv:1701.01107](#)] [INSPIRE].
- [7] A.R. Brown, M. Freedman, H. Lin and L. Susskind, *Complexity Geometry of the Exponential Metric*, to appear.
- [8] D.A. Trunin, *Pedagogical introduction to SYK model and 2D Dilaton Gravity*, [arXiv:2002.12187](#) [INSPIRE].
- [9] S. Helgason, *Differential geometry, Lie groups, and symmetric spaces*, Academic Press (1979).
- [10] J. Milnor, *Curvatures of left invariant metrics on lie groups*, *Adv. Math.* **21** (1976) 293.
- [11] V. Berestovskii, *Curvatures of left-invariant riemannian metrics on lie groups*, *Geometry and Topology of Homogeneous Spaces* (in Russian), Barnaul Univ., Barnaul (1988) 3.
- [12] C.M. Bender and S.A. Orszag, *Advanced mathematical methods for scientists and engineers I: Asymptotic methods and perturbation theory*, Springer Science & Business Media (2013).
- [13] J.D. Bekenstein, *Black holes and entropy*, *Phys. Rev. D* **7** (1973) 2333 [INSPIRE].
- [14] C.A. Egan and C.H. Lineweaver, *A Larger Estimate of the Entropy of the Universe*, *Astrophys. J.* **710** (2010) 1825 [[arXiv:0909.3983](#)] [INSPIRE].
- [15] A. Selberg, *Harmonic analysis and discontinuous groups in weakly symmetric spaces with applications to dirichlet series*, *J. Indian Math. Soc.* **20** (1956) 47.
- [16] G.R. Jensen and S. Chern, *The scalar curvature of left-invariant riemannian metrics*, *Indiana Univ. Math. J.* **20** (1971) 1125.
- [17] D. Bar-Natan, *Perturbative chern-simons theory*, *J. Knot Theor. Ramifications* **4** (1995) 503.
- [18] D.P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, [arXiv:1412.6980](#) [INSPIRE].