JHEP

Published for SISSA by Springer

Received: June 25, 2019
Revised: July 22, 2019
Accepted: July 29, 2019
Published: August 9, 2019

SO(8) supergravity and the magic of machine learning

Iulia M. Comsa, Moritz Firsching and Thomas Fischbacher

Google Research,
Brandschenkestrasse 110, 8002 Zürich, Switzerland

E-mail: iuliacomsa@google.com, firsching@google.com, tfish@google.com

Abstract: Using de Wit-Nicolai $D = 4$ $\mathcal{N} = 8$ SO(8) supergravity as an example, we show how modern Machine Learning software libraries such as Google's TensorFlow can be employed to greatly simplify the analysis of high-dimensional scalar sectors of some M-Theory compactifications. We provide detailed information on the location, symmetries, and particle spectra and charges of 192 critical points on the scalar manifold of SO(8) supergravity, including one newly discovered $\mathcal{N} = 1$ vacuum with SO(3) residual symmetry, one new potentially stabilizable non-supersymmetric solution, and examples for "Galois conjugate pairs" of solutions, i.e. solution-pairs that share the same gauge group embedding into SO(8) and minimal polynomials for the cosmological constant. Where feasible, we give analytic expressions for solution coordinates and cosmological constants.

As the authors' aspiration is to present the discussion in a form that is accessible to both the Machine Learning and String Theory communities and allows adopting our methods towards the study of other models, we provide an introductory overview over the relevant Physics as well as Machine Learning concepts. This includes short pedagogical code examples. In particular, we show how to formulate a requirement for residual Supersymmetry as a Machine Learning loss function and effectively guide the numerical search towards supersymmetric critical points. Numerical investigations suggest that there are no further supersymmetric vacua beyond this newly discovered fifth solution.

Keywords: Supergravity Models, Supersymmetry Breaking, AdS-CFT Correspondence, M-Theory

ArXiv ePrint: 1906.00207

https://doi.org/10.1007/JHEP08(2019)057

JHEP08(2019)057

Contents

JHEP08(2019)057

At the moment, the $\mathcal{N} = 8$ Supergravity Theory is the only candidate in sight. There are likely to be a number of crucial calculations within the next few years that have the possibility of showing that the theory is no good. If the theory survives these tests, it will probably be some years more before we develop computational methods that will enable us to make predictions and before we can account for the initial conditions of the universe as well as the local physical laws. These will be the outstanding problems for theoretical physics in the next twenty years or so.

But to end on a slightly alarmist note, they may not have much more time than that. At present, computers are a useful aid in research, but they have to be directed by human minds. If one extrapolates their recent rapid rate of development, however, it would seem quite possible that they will take over altogether in theoretical physics. So, maybe the end is in sight for theoretical physicists, if not for theoretical physics.

S. Hawking, Conclusion of his 1981 Inaugural lecture [1]
"Is the end in sight for theoretical physics?"

1 Introduction[1]

Google's primary open source library for Machine Learning, TensorFlow [2], has many potential uses beyond Machine Learning. In this article, we want to show how it also is an excellent tool to address one specific technically challenging M-Theory research problem: finding static field configurations of dimensionally reduced models with known but structurally complicated potentials, such as SO(8) supergravity [3, 4], which we study here, as well as determining their stability properties. The underlying computational methods can be readily generalized to other models, including for example maximal five-dimensional supergravity [5].

For the impatient reader, there is an open sourced Google Colab at [6] that runs an efficient search for vacuum candidates of SO(8) supergravity and can be used interactively from within a web browser, alongside additional Python code to analyze numerical solutions at [7].

1.1 On M-theory

"M-Theory", or "The Theory Formerly Known as Strings" [8] is a so far only partially explored and understood unifying framework for studying (some) field theoretic models of quantum gravity. The five known (very likely) consistent ten-dimensional Superstring theories (including compactifications to lower dimensions), as well as 11-dimensional supergravity are understood to be different limits of M-Theory dynamics [9]. If supersymmetry [10] is part of the answer why the observed fundamental laws of physics are the way they are (and it seems to have some good answers to problems that arise in Planck-scale physics), then, due to the existence of gravity, there is no way to escape the conclusion that a viable theory must contain supergravity [11, 12] and in particular a supersymmetric

[1] An expanded introduction that provides more context on M-Theory and Machine Learning to interested readers without a deep background in one of these subjects is available in version 3 of the arXiv preprint of this work at: arXiv:1906.00207v3.

JHEP08(2019)057

partner to the graviton with spin-3/2, the gravitino. While the question is still not settled whether one can construct a theory of supergravity that not only works as an effective low energy theory but is well-behaved at every length scale [13, 14], it is generally thought that problems that arise in simple models of supergravity [15] are ultimately resolved by the notion of "point particles" in quantum field theory breaking down at very high energies [16], i.e. superstring theory. Now, if one accepts superstring theory, there is no way to avoid the conclusion that there also needs to be a way to describe its non-perturbative strong coupling limit, which then inevitably leads us to M-Theory [17].

Despite the remarkable success of the Standard Model (SM) of Elementary Particle Physics [18], which quantitatively describes the properties and interactions of matter and force particles so well that the LHC at the time of this writing did not to come up with clear evidence of "new" (beyond-the-SM) physics, there are a number of unsolved problems, for example non-observation of the particles that constitute Dark Matter [19], or explaining why the neutron's electric dipole moment is too small to be measurable [20]. The most puzzling such problem of theoretical physics currently is perhaps explaining the observed positive — but from a quantum field theory perspective extremely small — vacuum energy density [21] of the universe. M-Theory currently struggles to give an answer to how this could arise naturally. Even if M-Theory ultimately turned out to not be the correct answer to the question how to quantize gravity, it already by now has made major contributions to uncovering interesting hidden connections in pure mathematics, of which we here only want to mention the geometric Langlands correspondence as one example, [22].

1.1.1 Unification

The unification of Quantum Electrodynamics with Quantum Chromodynamics (QCD) and the Weak Force into the SM is highly successful from a theoretical perspective, with both QCD and Electroweak theory individually being afflicted by problems that cancel in the SM [23, 24]. A key property is that in the SM, all forces are described in an uniform way by vector gauge bosons. Since spacetime symmetries (rotations and boosts) affect all vector gauge bosons in the same way, one can consider superposition quantum states between them. Indeed, the SM Photon emerges as a specific quantum superposition of a Weak Force's particle and another "hidden" force's particle termed $\mathrm{U}(1)_Y$ in a way that is governed by properties of the Brout-Englert-Higgs Boson. This also sets — for example — the relative strengths of the Electromagnetic and Weak forces. It is quite plausible that at even (much!) higher energies (but somewhat below the quantum gravity energy scale), the same mechanism also is at play in the form of "Grand Unification" [25, 26], merging the Strong Force with the Electroweak Force.

The main new discoveries described in the current article are also about this "Higgs effect" [27, 28], leading to some particular sets of particles and interactions in "low energy" physics, so in terms of basic mechanisms this is rather similar to what is now well established SM physics. Our setting, however, is that of a model that might actually describe Planck-scale physics well, so if this construction ever turned out to somehow actually be related to the SM (see e.g. [29–31] for some speculation in this direction), interesting Physics that is not well understood yet would need to happen in the gap of ~ 17 orders

JHEP08(2019)057

of magnitude between the Higgs boson energy scale and the quantum gravity energy scale. In particular, there is a discrepancy on gauge groups, cosmological constant, and particle chirality properties.

More likely, this work describes a collection of (stable as well as unstable background-)solutions to the M-Theory field equations in a very different corner than the one that describes the Standard Model. Given our still limited understanding of M-Theory, it nevertheless seems useful to get a better idea about relevant properties, mechanisms, and phenomena by investigating solutions that are accessible with current technology. In the past, studying solutions to the 4-dimensional field theory equations and especially their embedding into M-Theory has taught us some very useful lessons about compactification mechanisms and about 11-dimensional dynamics. One in particular notes that SO(8) supergravity also describes the physics of a stack of M2-branes [32].

1.1.2 Kaluza-Klein supergravity

While the observation, originally by Kaluza [33] and Klein [34], that four-dimensional physics can be understood in the context of dynamics in higher dimension was certainly interesting, the question remained how to find guidance on what higher dimensional dynamics to start from. A major step forward happened in 1979, when Cremmer, Julia, and Scherk set out to construct the four-dimensional field theory with the maximum possible amount of boson-fermion symmetry [35–37]. Due to the complicated structure of the (Higgs-like) scalar boson interactions, this was achieved by first realizing that such a theory should be obtainable via compactification of a higher dimensional ancestor theory, and that there can only be one graviton with helicity ± 2 and no interacting massless particles with higher helicity (due to nonexistence of a suitable source current), and that the highest dimension in which such a symmetry can exist is eleven (since otherwise dimensional reduction to four dimensions would give rise to too many supersymmetry generators that require going to helicities beyond $+2$ that cannot be consistently coupled). Constructing the 11-dimensional model first in [35], the maximally symmetric four-dimensional model in which all matter and force particles are unified succeeded in [36] via Kaluza-Klein reduction on a 7-dimensional torus.

The "auxiliary" 11-dimensional field theory, originally introduced as a mathematical trick, soon was found to be very interesting in itself. For example, it so turns out that symmetry constraints completely determine its structure, and there is no way to adjust its parameters or field content. It almost certainly describes the low energy limit of an as yet unknown 11-dimensional theory of supersymmetric membranes (and perhaps other dynamical degrees of freedom) which, upon dimensional reduction on a circle also wrapped by one direction of the membrane, produces 10-dimensional Superstring Theory [9]. This unknown (perhaps) 11-dimensional theory of (likely) supermembranes has been given the provisional name "M-Theory".

As explained, a key ingredient in the effort to unify force and matter quantum fields is boson-fermion symmetry, or "Supersymmetry", as it is commonly called. This is presently is a somewhat esoteric topic outside of quantum field theory and some branches of differential topology, plus perhaps the theory of stochastic dynamical systems [38].

JHEP08(2019)057

1.1.3 Supergravity in eleven and four dimensions

A supersymmetry transformation changes the helicity of a particle by 1/2 and so — in a theory of gravity — it must connect the helicity-2 graviton with a helicity-3/2 fermionic particle, the "gravitino". It is possible to construct models with more than the minimal amount of supersymmetry that then fuse more helicity states [39–41], allowing one to start with a helicity-2 graviton, apply a supersymmetry transformation to step down to the helicity-3/2 gravitino, and use another independent supersymmetry transformation to further step down to a helicity-1 photon-like particle that couples with gravitational strength (a "graviphoton" [42]). A maximally supersymmetric theory in four dimensions, as obtainable through dimensional reduction of 11-dimensional supergravity, has $\mathcal{N} = 8$ independent supersymmetries that connect all quantum states from the helicity +2 graviton down to the oppositely-polarized helicity −2 graviton, with (according to simple combinatorics) $\binom{8}{k}$ particles of helicity $2 - k/2$, so in total one graviton, eight gravitini, 28 photon- or gluon-like force carriers, 56 spin-1/2 fermions, and 70 Higgs-Boson-like scalars. It was the discovery of this $\mathcal{N} = 8$ supergravity, which manages to unify all particles and interactions starting from only a symmetry principle as input, that made S. Hawking suggest in his 1981 inaugural lecture [1] that within perhaps 20 years, we would know the "Theory of Everything".

A peculiar feature of this construction is that it interprets the 70 Higgs fields as parametrizing a very special 70-dimensional coset manifold. Just as the sphere can be regarded as the manifold of 3-dimensional rotations (read: orthogonal bases) modulo another rotation (around the outward-pointing direction), i.e. $S^2 = \mathrm{SO}(3)/\mathrm{SO}(2)$, and the hyperbolic plane[2] can be regarded as the coset space $\mathrm{SO}(2,1)/\mathrm{SO}(2) = \mathrm{SL}(2)/\mathrm{SO}(2)$, the relevant scalar manifold of $\mathcal{N} = 8$ Supergravity is[3] $E_{7(7)}/\mathrm{SU}(8)$, where SU(8) is the group of complex 8×8 matrices with unit determinant and the maximal compact subgroup of the 133-dimensional non-compact exceptional Lie group $E_{7(7)}$, which we describe in appendix A.

In Cremmer and Julia's original construction [36], which compactified 11-dimensional Supergravity on a 7-dimensional torus, there are 28 photon-like gauge boson particles. Soon after, it was realized that one can also obtain a four-dimensional theory with the maximal amount of supersymmetry by dimensionally reducing on the "round" 7-dimensional sphere[4] instead [46]. Here, one ends up with the 28 vector gauge bosons belonging to the nonabelian gauge symmetry SO(8), which may superficially be thought of as some sort of more complicated Quantum Chromodynamics (but with very differently behaving "quarks" and "gluons", and perhaps without confinement due to vanishing β-function). Clearly, given that such nonabelian gauge symmetries do play an important role in the SM, this looks like a major step in the right direction, but unfortunately, the group SO(8) is too small

[2] For a game that allows one to develop some intuition about living on a hyperbolic plane, see [43].

[3] Strictly speaking, it is actually $E_{7(7)}/(\mathrm{SU}(8)/\mathbb{Z}_2)$, as the SU(8) group element that maps an **8**-vector (or $\bar{\mathbf{8}}$-vector) to minus itself gets represented as the identity operation when acting on the scalar manifold, while this does not happen for any other element from the center of SU(8), apart from the identity itself.

[4] The 7-sphere is rather special, as there are many (28 in total) different spaces that all are topologically 7-spheres, but not diffeomorphic to one another [44, 45].

JHEP08(2019)057

to embed the $SU(3)_{QCD} \times SU(2)_{Weak} \times U(1)_Y$ gauge group of the SM into it. Also, the experimentally observed left-right asymmetry of the SM ("chirality") cannot be obtained by compactifying non-chiral 11-dimensional supergravity on a manifold [47].

The early literature on this topic contemplated scenarios in which the observed particles would emerge as composite, being made of more fundamental "preons", somewhat along the lines of how QCD at lower energies gives rise to baryon and meson bound states. Given that there are in terms of energy perhaps 17 orders of magnitude between the quantum gravity scale and the Higgs boson energy scale, this may not be entirely unplausible. Still, considering in particular the problems associated e.g. with chiral fermions, it is nowadays generally regarded as more promising to investigate scenarios in which the SM's gauge symmetry directly emerges from some large higher-dimensional symmetry. Going from 11-dimensional M-Theory to Superstring Theory first, and then down to four dimensions, there are by now multiple options to directly get a large "Grand Unification" gauge symmetry into which the SM gauge symmetry can be embedded.

The main focus of the current article, however, is not to provide more insights into how experimental particle physics might be related to M-Theory. Rather, we want to allow deeper investigations into the structure of M-Theory by both expanding our knowledge on what possible background solutions to its field equations can look like, and also by providing tools that allow one to come to grips with some of the technical complications that arise in particular when working with high-dimensional scalar manifolds. In the past, we have time and again seen the study of models of quantum gravity produce highly surprising and useful insights even if they were not at all focused on the four-dimensional world we inhabit. Most notably, there was the realization in 1997 [48] that the partition function (/generating functional) of a Conformal Field Theory (CFT) can be the same as the partition function of a supersymmetric theory of gravity with *negative* cosmological constant (hence in "anti de Sitter (AdS) space") in a different spacetime dimension (i.e. with an extra spatial direction), the so-called AdS/CFT correspondence [48–50]. One example for a rather surprising further development of this idea is the insight that Quantum Field Theory may provide a lower bound for the ratio of shear viscosity to entropy density of *any* liquid [51]. Another modern development that could not have been anticipated is the application of this idea of gauge/gravity duality to study superconductivity [52–54]. To give another example, "holographic duality" has been employed to map solution-generating symmetries of the Einstein equations [55, 56] to solution-generating symmetries for the Navier-Stokes equation [57, 58].

Concerning specifically AdS_4/CFT_3 duality, the holographic dual of the SO(8) supergravity studied here is (the $k = 1$ case of) three-dimensional ABJM theory [59], which describes the dynamics of M2-branes. This was used e.g. in [60] to construct new supersymmetric AdS_4 black holes and provide an explanation for their Bekenstein-Hawking entropy, exploiting the relation between mass-deformed ABJM theory with $\mathcal{N} = 2$ supersymmetry and the AdS_4 vacuum with $\mathcal{N} = 2\, SU(3) \times U(1)$ symmetry, which in this work is called solution S0779422.

JHEP08(2019)057

1.2 On machine learning

Artificial Intelligence (AI) is a broad field concerned with crafting algorithms for solving problems that require some form of human-like intelligence. To avoid any misconceptions, we clarify that the main concern of AI is not finding ways to allow algorithms to perform introspective reasoning on par with or exceeding human ability. Indeed, as famously noted by Alan Turing, the question of whether machines can think is ill-posed [61].

The earliest forms of AI consisted of explicit, manually-crafted rules. Machine Learning (ML) introduced a new perspective on creating artificially intelligent algorithms. This field was pioneered by Arthur Samuel, who demonstrated that a computer program can learn to play the game of checkers better than the person who programmed it [62]. Instead of operating with pre-adjusted rules and fixed numeric values, the algorithm would instead tune itself in order to solve the problem. In other words, given a function of an input space that represents the problem data and an output space that represents the problem solution, the challenge becomes to learn the parameters of this function in such a way that its results (the output, or the solution of the algorithm) is as close as possible to the correct solution. Usually, the learned parameters are of numeric form. The field of ML is thus primarily concerned with the pragmatic problem of finding and efficiently refining functions that usually have a large number of adjustable ("learnable") parameters, with the purpose of solving challenging problems that often involve real-world data. ML methods are suitable whenever facing a problem that is difficult to put into words or fixed rules.

In a way, Machine Learning (ML) and physics can be regarded as intellectual antipodes: physics tries to understand fundamental processes and important mechanisms underlying the functioning of a system, while ML tries solve a particular problem as well as possible, while eschewing the need to fully understand it. In fact, the implicit knowledge obtained by an ML algorithm by solving a problem is often difficult to analyze. Understanding how certain highly-successful ML algorithms manage to solve highly difficult problems and visualizing various parts of the learned function in order to produce an intuitive understanding of the problem and the solution space is an active field of research [63].

Example problems that have, sometimes surprisingly so, turned out to be amenable to ML approaches include text [64] or object [65, 66] recognition in images, mapping pictures to textual descriptions of their content [67], machine translation of natural language [68], scoring possible moves in the game of Go [69] and Starcraft [70], and many more. Increasingly, we also see ML methods being applied to problems that do not strictly follow this pattern, such as synthesis of realistic-looking portraits [71].

Concerning direct applications of ML to theoretical physics, it can and indeed has happened in the past that ML demonstrated an ability to predict a system's behavior well beyond what our current thinking would have considered possible, indicating the existence of extra structure that our current models cannot capture well. For example, [20] demonstrated a clever set-up that allows ML to accurately predict the behavior of a chaotic system over eight Lyapunov times.

One particularly successful family of ML algorithms is that of Artificial Neural Networks (ANNs). ANNs are loosely inspired by biological brains, which are made of billions

JHEP08(2019)057

of interconnected neurons working together to control optimally the behaviour of intelligent organisms. A simple model of a neuron, called a Perceptron, was proposed by Frank Rosenblatt in 1958 [72], but the idea of networks composed of multiple layers of Perceptrons only started becoming popular in the ML community in the 1980s [73]. ANNs consist of artificial "neurons", non-linear circuit elements that are interconnected through directed artificial "synapses" that transmit signals with different efficacies, which act like "weights" in directed graphs. The connectivity architecture of such a network is usually layered, the intuition being that each layer builds up more abstract concepts than the previous. A fully connected feedforward network includes connections between every node of a layer and every node of the next layer, but other variations also exist, such as recurrent [74] or convolutional [75] layers.

Such "layered" ANNs are popular as they are known to be universal approximators [76] and have been found to work well for many problems, but it is by no means true that ML is tied to this particular class of architectures. As long as there is a way to model a problem in terms of a function that differentiably depends on many parameters, and parameter-tuning can substantially improve performance, ML techniques are applicable.

Deep learning, which has recently achieved resounding success in solving difficult real-world problems like the ones mentioned earlier, refers to ANNs with a large number of stacked layers especially designed to apply specialized operations on the input. It was not trivial to discover that such deep networks can work at all — until Hinton's seminal 2006 article [77], which sparked the deep learning revolution, common thinking was that networks with more than two layers were essentially impossible to train, and other ML approaches, such as kernelized support vector machines [78], would generally perform better than ANNs. Later progress uncovered a number of general misconceptions and useful tricks on how to train ANNs, for example the superior performance of the "Rectified Linear Unbounded" (ReLU) activation function [79] in comparison to the classical sigmoid non-linear activation function used in earlier research.

What type of problems is ML applicable to? Depending on the amount and type of available data, there are three main paradigms for training an ML model: supervised, unsupervised and reinforcement learning. Supervised learning refers to data where the expected result is known in advance for the data available; for example, given a large set of images of people, the name of the person appearing in each image is also given. This type of learning is often used with classification ("given m labels, pick the correct one") and regression ("predict a value in a continuous domain"), but can often be adapted for other types of problems, for example in assessing the value of each possible next action in a game [69]. Supervised learning with ANNs is currently the most widely-used and successful approach to ML. In contrast, unsupervised learning occurs when no labels exist for the given data; in this case, the aim is to group the data in such a way that items similar to each other belong to the same group, or are close to each other in the output domain [80]. Examples where this approach is useful are fetching web pages, songs or videos similar to the one that an internet user might be viewing or listening to currently. Nonetheless, such problems can also be expressed as a supervised problems [81] Finally, reinforcement learning is applied when no exact labels exist, but there is some knowledge on whether a proposed

JHEP08(2019)057

output is good or bad. This is applicable in particular to automated game playing, where the algorithm acts as agent that chooses to perform a sequence of actions with the aim of winning the game; by playing multiple games, it slowly learns to pick better actions based on whether the past games resulted in wins or losses. Reinforcement learning can also be combined with supervised learning: given a very large number of possible actions, a supervised deep ANN can approximate the value of each possible action [82].

So how does learning in an ML model actually work? A key idea is that learning involves the minimization of a loss (or error) function. This function is designed such that, when applied to any given output, it provides a numerical measure of how far off this output is from an expected answer. In supervised learning, this can be thought of as a distance between the actual output and the target (desired) output of the algorithm. If the value of the loss function is smaller, the error is smaller, and the algorithm is closer to the desired output. Thus, instead of deeming an output as either correct or incorrect, the loss function provides a graded measure of "wrongness". The output of the network can thus be often interpreted as a probability. Crucially, if the loss function is differentiable with respect to the algorithm parameters (for example, in case of an ANN, the network weights), the gradient of the loss function can be used to point towards the direction of the minimum of this function. The gradient of the loss function (which usually is estimated on a random selection of examples, see below) can be used to iteratively tune parameters in order to improve the performance of predictions. For many problems, there are natural choices of loss functions. For classification problems with n possible labels (e.g. "which digit does the image show" with $n = 10$), the predicted probabilities can be regarded as dual to chemical potentials, represented as (linearly) accumulated evidence E_j for or against a particular classification label p_j, i.e. $p_j = exp(-E_j)/\sum_{i=1}^{n} exp(-E_i)$ [83]. However, any type of loss function can be employed as long as it indicates the correct solution to the given problem, is differentiable with respect to the learnable parameters, and has a reasonable shape — for example, not too many 'bad' local minima. One of the surprising insights of the Deep Learning revolution was that a simple non-linear activation function with discontinuous derivative that reduces to the identity in the activation region and to the null function outside that region, i.e. $\mathrm{ReLU}\,(x) := (x + |x|)/2$, allows training deeply nested transformations to extract high-level information such as whether there is a face in an image. In some situations, finding good loss functions to represent important aspects of a problem is less straightforward, and may need some experimenting. In this work, for example, we show how the desire to have unbroken vacuum supersymmetry can be represented concisely through a ML loss function.

A key idea that made ANN-based learning possible is that, when given a computer program that computes a $\mathbb{R}^n \to \mathbb{R}$ function f, it is possible to automatically transform this into another program that computes the n-component gradient ∇f at any given point with relatively small computational effort that is independent of n [84]. The work on reverse-mode automated differentiation pre-dates and provides a more general framework than "error backpropagation" for ANNs as it was rediscovered independently in 1985 in [85]. Interestingly, one can also regard "reverse mode automatic differentiation" as a discretized version of the idea underlying Pontyragin's maximum principle in Optimal Control Theory,

JHEP08(2019)057

i.e. the Hamilton-Jacobi-Bellman (HJB) equation [86] from the 50's, in the sense that applying reverse-mode AD to the most basic ODE solver algorithm directly produces the HJB equation.

Given that not all ML applications use a single straightforward layered ANN architecture, it makes sense for a Machine Learning library like TensorFlow to provide some form of general-purpose reverse-mode automatic differentiation capabilities. In principle, there are three ways to do this, (1) full program analysis, which for a language as complex as C++ or even only Python is a formidable task (this has been done for Scheme with R6RS-AD, [87]), (2) implementing some "domain-specific language (DSL)" for arithmetic graphs, and (3) "Tape-based" AD, where in the forward pass, the sequence of arithmetic operations gets recorded on a "tape", which then is replayed in reverse. TensorFlow 1.x uses approach (2), while TensorFlow 2.x tries to make the tape-based paradigm the default choice. In this article, we will exclusively use a graph-based TensorFlow 1 approach.

1.3 Tensors in machine learning

To give a rough mental picture of what the training process might look like at the level of number-crunching, we give an example problem where the goal is to predict whether a person appears or not in a given image. We assume that a labeled dataset on the order of a million images is available for training. An image could be represented as a 3-index array $X[row, col, c]$, with indices providing row and column pixel coordinates as well as the color channel c. The label for each image is represented by the number 1 of there is a person in the image, and the number 0 if not. One would typically start by grouping example images into sufficiently large (randomized) batches to get reasonable estimates for loss function gradients with respect to model parameters, perhaps $b = 1024$ images per batch. A batch of training images would then be naturally represented as a b-dimensional array of pairs (X_b, Y_b). It has become fashionable to call these higher-rank arrays "tensors" in ML terminology, which indeed is a useful notion for expressing the ANN operations in terms of tensor products and index-splitting operations. However, symmetry groups to this date play a rather minor role in ML (with notable exceptions such as [88]), and if they actually do, one often talks about "equivariant neural networks" to discriminate these from networks with less structure. For a problem such as recognizing whether a picture contains a person, which evidently benefits from utilizing symmetry, the common approach is to factor out translational symmetry by effectively imposing constraints on network parameters relevant for detecting the target entity, or elements of it, at different locations in the image. This is usually done by "convolutional layers" that computes convolutions $C[b, i, j, k] = \sum_{\xi,\eta} X[b, i + \eta, j + \xi, c]S[k, \eta, \xi, c]$ of the example images with a collection $S[\cdot, \cdot, \cdot, \cdot]$ of small stencils represented as an array of trainable parameters. The stencil parameters then get adjusted in the training phase such that they are optimally useful for coming up with a good probability prediction for the image to show a person in any location. Each such stencil will consist of lines that describe typical features associated with a person in the image, such as noses, eyes or ears. Intuitively, one could imagine one of the stencils getting tuned by training to have large inner product with "the average shape of all noses", so getting specialized to a nose-detector. Each such stencil would, in turn, be

JHEP08(2019)057

made of lower-level stencils, such as lines with particular orientations, which, in the right combination, form salient shapes. Subsequent processing layers in the network would then collect and combine different such evidence and in the end produce a Bayesian prediction roughly along the lines of: "We are highly confident to have seen a nose in the picture, and we also have moderate confidence to have seen an eye somewhere, so, with high likelihood, the image shows a person". Krizhevsky's seminal paper [65] explains in detail one such convolutional ANN architecture for image processing. Recent work on feature visualization in ANNs has spectacularly uncovered collections of shapes and patterns that hidden layers in a network learn to recognize [63].

The above example illustrates why numerical higher-rank arrays are so prevalent in modern ML. As hinted earlier in this section, one very common primitive "tensor" operation in such a setting is batched matrix-multiplication. For example, linear conversion of a set of example images from RGB color space to some implicit color space that can be trained to be optimally useful for solving the problem codified by the loss function expressed as

$$X2[b,i,j,d] = \sum_c X[b,i,j,c] * M[d,c]$$

with trainable parameters in the matrix M.

2 $D = 4$ SO(8) supergravity and its scalar sector

Let us briefly review some salient features of supergravity in four and eleven dimensions before we look into finding equilibrium solutions to the equations of motion.

Four-dimensional supersymmetry can at most unify all particle states from the helicity +2 down to the helicity -2 graviton. As there are eight helicity-1/2 steps between these helicities, we can have at most eight times the minimal amount of supersymmetry, and as each of these eight supersymmetry transformations comes as a real (Majorana) four-dimensional spinor, we are looking at a theory with $8 \cdot 4 = 32$ supersymmetry components. The highest spacetime dimension in which we can have a real 32-component spinor is $d = 11$ (or perhaps $d = 12$ if we accepted a second direction of time [89]). A supersymmetric theory of gravity in $D = 11$ dimensions will have $(D-2)(D-1)/2 - 1 = 44$ transversal graviton polarization states (described by a symmetric traceless 9×9 matrix), plus 128 gravitino degrees of freedom. The mismatch in the number of degrees of freedom is compensated by a gauge field with 84 degrees of freedom, describing a higher-dimensional cousin of the photon whose polarization is not given by a 1-dimensional vector, but by a 3-dimensional volume(-form) embedded into 9-dimensional transversal space, A_{MNP}, with associated (4-form) field strength F_{MNPQ}. With the "polarization" being a 3-dimensional object, this (abelian) gauge field can not be sourced by charged particles (the 1-dimensional photon polarization couples to the 1-dimensional worldline of an electron), but by some membrane-like extended object that lives in eleven dimensions. This is now understood to be the M2-brane [90]. It is amazing to see how starting from one of the three possible gauge principles, the vector-spinor, in its very own preferred (maximal) dimension, one automatically obtains a theory that unifies all three of the possible gauge principles, and furthermore turns out to be completely fixed, i.e. not permit any free parameters.

JHEP08(2019)057

The Lagrangian of 11-dimensional Supergravity reads [91, 92]

$$\begin{aligned}\mathcal{L}/e = &\frac{1}{4}R_{MN}{}^{AB}e_M{}^A e_N{}^B - \frac{i}{2}\bar{\Psi}_M\Gamma^{MNP}D_N\left(\frac{1}{2}(\omega+\tilde{\omega})\right)\Psi_P\\ &-\frac{1}{48}F_{MNPQ}F^{MNPQ}\\ &+\frac{2}{12^4}\epsilon^{MNPQRSTUVWX}F_{MNPQ}F_{RSTU}A_{VWX}\\ &+\frac{3}{4\cdot 12^2}\left(\bar{\Psi}_M\Gamma^{MNWXYZ}\Psi_N + 12\bar{\Psi}^W\Gamma^{XY}\Psi^Z\right)\left(F_{WXYZ}+\tilde{F}_{WXYZ}\right)\end{aligned}$$

where

$$\begin{aligned}e &:= \det e_M{}^A\\ D_M(\omega) &:= \partial_m - \frac{1}{4}\omega_M{}^{AB}\Gamma_{AB},\\ \omega_{MAB} &:= \frac{1}{2}\left(\Omega_{ABM} - \Omega_{MAB} - \Omega_{BMA}\right) + K_{MAB}\\ K_{MAB} &:= \frac{i}{4}\left(-\bar{\Psi}_N\Gamma_{MAB}{}^{NP}\Psi_P + 2\left(\bar{\Psi}_M\Gamma_B\Psi_A - \bar{\Psi}_M\Gamma_A\Psi_B + \bar{\Psi}_B\Gamma_M\Psi_A\right)\right)\\ \Omega_{MN}{}^A &:= \partial_N e_M{}^A - \partial_M e_N{}^A\\ \tilde{\omega}_{MAB} &:= \omega_{MAB} + \frac{i}{4}\bar{\Psi}_N\Gamma_{MAB}{}^{NP}\Psi_P\\ F_{MNPQ} &:= 4\delta^{RSTU}_{MNPQ}\partial_R A_{STU}\\ \tilde{F}_{MNPQ} &:= F_{MNPQ} - 3\delta^{RSTU}_{MNPQ}\bar{\Psi}_R\Gamma_{ST}\Psi_U\ .\end{aligned} \tag{2.1}$$

2.1 Compactification to four dimensions

Freund and Rubin noted [93] that this theory preferentially compactifies to four dimensions due to the presence of the four-form field strength F_{ABCD}. Indeed, a "flux" compactification with $F_{ABCD} \sim \epsilon_{\mu\nu\rho\sigma}$, i.e. flux aligned with the submanifold of four-dimensional spacetime, will look isotropic from the four dimensional perspective. Kaluza-Klein compactification to four spacetime dimensions on a 7-sphere that is the surface of an 8-ball gives the Lagrangian of the de Wit-Nicolai model [3, 46].

For our investigations, we are mostly concerned with the scalar sector of this "SO(8) supergravity". Naturally, polarized fields in 11 dimensions give rise to different types of fields in four dimensions, depending on how 11-dimensional polarization is oriented with respect to the split into a seven-dimensional compact manifold and four-dimensional space-time, just like in original Kaluza-Klein theory, where the five-dimensional metric gives rise to the four-dimensional metric (gravitons), vector potential (photons), and scalar field (Higgs boson). Maximal supersymmetry fixes the particle content completely, and so Cremmer and Julia's construction of ungauged four-dimensional maximal supergravity that compactifies on a 7-torus and drops higher Kaluza-Klein modes must give rise to the same particle content as compactification on the surface of an 8-ball (and retaining only massless modes). The rather nontrivial input here is that both constructions actually do lead to

JHEP08(2019)057

maximally supersymmetric models. In Cremmer and Julia's construction, one gets 35 Higgs fields from the 11-dimensional "A_{MNP}-photons" for which the 3-dimensional polarization is parallel to the direction of the 7-dimensional compactification manifold. Since reversing the handedness of three-dimensional space can be expressed as an 11-dimensional rotation that also reverses the handedness of the 7-dimensional compactification manifold, which is experienced by a 3-form field as a sign reversal, these $7 \cdot 6 \cdot 5/3! = 35$ scalar fields are pseudoscalars, i.e. odd under a parity transformation. Correspondingly, we get $7 \cdot 8/2 = 28$ scalars from those polarization states of the 11-dimensional graviton g_{MN} that are parallel to the embedding manifold. *However*, we also get seven four-dimensional-2-form potentials $A_{\mu\nu P}$ for which only one of the three 11-dimensional A_{MNP} polarization directions is parallel to the compactification manifold. These give rise to four-dimensional 3-form field strengths $\sim F_{\mu\nu\rho}$, which can be dualized to 1-form field strengths $G_\lambda \sim g_{\lambda\sigma}\epsilon^{\sigma\mu\nu\rho}F_{\mu\nu\rho}$, which in turn come from scalar potentials $G = \partial A_G$. So, dualization [94, 95] of these 2-forms produces another seven scalar fields which, like the 28 from the graviton, are parity-even, so (proper) scalars. One finds that these indeed combine into one irreducible representation of eight-dimensional rotations, so, in this compactification, we have 35 scalars (35^+) as well as 35 pseudoscalars (35^-) from the (lowest-energy Kaluza-Klein ("Fourier") modes of the) 11-dimensional degrees of freedom. One indeed finds that these $35 + 35$ scalar fields can be understood at parametrizing the coset space $E_{7(7)}/(\mathrm{SU}(8)/\mathbb{Z}_2)$.

Subtly, despite ungauged maximal supergravity and SO(8) supergravity having equivalent particle spectra, and the latter also having a smooth limit in which the gauge coupling constant is taken to zero,[5] one can *not* readily identify the 70 scalars of one construction with the 70 scalars of the other [96]. Rather, when compactifying on a 7-sphere, one has to work out fluctuations around a compactification background geometry, as explained e.g. in [97–100]. In the latter case, there is a straightforward way for the rotational symmetry of the 8-ball to act on these fluctuations, so all four-dimensional fields should form SO(8) irreducible representations. In the former case, the seven two-forms which one gets from A_{MNP} clearly do not form an irreducible representation of SO(8), so the symmetry enlargement is an emergent phenomenon.

In general, determining the low-energy field content of Kaluza-Klein type compactifications of M-Theory will require carefully analyzing the spectra of generalized Laplace operators which act not on scalar but tensor-valued fields (see e.g. [92], esp. chapters 4, 5, 9), whose eigenfunctions can be thought of as generalized spherical harmonics that live on the compactification manifold rather than the surface of a 3-dimensional ball (as the spherical harmonics do). On other compactification manifolds, the low energy particle content of the theory may be rather different, it may even contain *more* Higgs-like fields than this SO(8) supergravity, as in the construction discussed in [101], which in total has $67 + 67 = 134$.

The algebra $\mathfrak{so}(8)$ of eight-dimensional rotations is very special in that it allows an S_3 group of outer algebra automorphisms which permute the roles of the three different real

[5] This does not hold in general, and in particular not for maximal supergravity in five or seven dimensions, see e.g. [5].

JHEP08(2019)057

eight-dimensional irreducible representations, the vectors, spinors, and co-spinors. Due to this phenomenon of "triality", we have a choice in how to attach the "vector", "spinor" and "co-spinor" label to the different eight-dimensional representations. While it is physically reasonable and common in the literature to associate the 35^+ with the symmetric traceless matrices over the $\mathfrak{so}(8)$ vectors $\mathbf{35}_v$ (given that they contain the graviton polarization states), we deviate from this convention in the present work and instead associate the scalars with the symmetric traceless matrices over the spinors $35^+ \equiv \mathbf{35}_s$, while associating (now in alignment with the literature) the pseudoscalars with the symmetric traceless matrices over the co-spinors, $35^- \equiv \mathbf{35}_c$. The advantage of this approach is that it aligns the defining **8**-representation of the maximal compact subalgebra $\mathfrak{su}(8)$ of the $\mathfrak{e}_{7(7)}$ algebra with the $\mathfrak{so}(8)$ vector representation, as well as the $\mathbf{35}_{s,c}$ with the self-dual/anti-self-dual four-forms of $\mathfrak{su}(8)$, i.e. we can use the geometric $\mathfrak{so}(8)$-invariants $\gamma^{ijkl}_{\dot\alpha\dot\beta}$, $\gamma^{ijkl}_{\alpha\beta}$ to translate between (anti-)self-dual and symmetric-traceless-matrix language. We heavily rely on this property to give simple expressions for the locations of all critical points.

While it would be tempting to give the full general $\mathcal{N}=8$ Lagrangian in the general unifying form presented in [102] that uses the gauge group embedding tensor framework to also include some alternative constructions in which the gauge group is a non-compact group[6] such as a different real form of SO(8), i.e. SO$(p,8-p)$, or a contraction thereof [103, 104], or a "dyonic" variant [105], it is more straightforward for this work to instead refer to the "classical" de Wit-Nicolai Lagrangian in order to explain the physical role of some key objects for which this work provides extensive data.

The Lagrangian of SO(8) supergravity reads [46]:

$$
\begin{aligned}
\mathcal{L}/e = &-\frac{1}{2}R(e,\omega) - \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}\left(\bar\psi^i_\mu\gamma_\nu D_\rho\psi_{\sigma i} - \bar\psi^i_\mu\overleftarrow{D}_\rho\gamma_\nu\psi_{\sigma i}\right)\\
&-\frac{1}{12}\left(\bar\chi^{ijk}\gamma^\mu D_\mu\chi_{ijk} - \bar\chi^{ijk}\overleftarrow{D}_\mu\gamma^\mu\chi_{ijk}\right) - \frac{1}{96}\mathcal{A}^{ijk\ell}_\mu\mathcal{A}^\mu_{ijk\ell}\\
&-\frac{1}{8}\left(F^+_{\mu\nu\,IJ}\left(2S^{IJ,KL}-\delta^{IJ}_{KL}\right)F^{+\,\mu\nu}{}_{KL} + \text{h.c.}\right)\\
&-\frac{1}{2}\left(F^+_{\mu\nu\,IJ}\left(S^{IJ,KL}O^{+\,\mu\nu\,KL}\right) + \text{h.c.}\right)\\
&-\frac{1}{4}\left(O^{+\,IJ}_{\mu\nu}\left(S^{IJ,KL}+u^{ij}{}_{IJ}v_{ijKL}\right)O^{+\,\mu\nu\,KL} + \text{h.c.}\right)\\
&-\frac{1}{24}\left(\bar\chi_{ijk}\gamma^\nu\gamma^\mu\psi_{\nu\ell}\left(\hat{\mathcal{A}}^{ijk\ell}_\mu + \mathcal{A}^{ijk\ell}_\mu\right) + \text{h.c.}\right)\\
&-\frac{1}{2}\delta^{ij}_{i'j'}\bar\psi^{i'}_\mu\psi^{j'}_\nu\bar\psi^\mu_i\psi^\nu_j\\
&+\frac{\sqrt{2}}{4}\left(\bar\psi^i_\lambda\sigma^{\mu\nu}\gamma^\lambda\chi_{ijk}\bar\psi^j_\mu\psi^k_\nu + \text{h.c.}\right)\\
&+\left(\frac{1}{144}\eta\epsilon_{ijk\ell mnpq}\bar\chi^{ijk}\sigma^{\mu\nu}\chi^{\ell mn}\bar\psi^p_\mu\psi^q_\nu + \frac{1}{8}\bar\psi^i_\lambda\sigma^{\mu\nu}\gamma^\lambda\chi_{ik\ell}\bar\psi_{\mu j}\gamma_\nu\chi^{jk\ell} + \text{h.c.}\right)
\end{aligned}
$$

[6] While one would be inclined to outright reject the idea of noncompact gauge groups, it turns out that at least the obvious unitarity problems are avoided in supergravity as the vector kinetic term has a "mass matrix" like factor involving the scalars that actually fixes the signs for non-compact directions [103], and any concerns about renormalizability of such theories are not much different from standard supergravity.

JHEP08(2019)057

$$
\begin{aligned}
&+\frac{\sqrt{2}\eta}{6\cdot 144}\left(\epsilon^{ijk\ell mnpq}\bar{\chi}_{ijk}\sigma^{\mu\nu}\chi_{\ell mn}\bar{\psi}^{r}_{\mu}\gamma_{\nu}\chi_{pqr}+\text{h.c.}\right)\\
&+\frac{1}{32}\,\bar{\chi}^{ik\ell}\gamma^{\mu}\chi_{jk\ell}\bar{\chi}^{jmn}\gamma_{\mu}\chi_{imn}\\
&-\frac{1}{96}\,\bar{\chi}^{ijk}\gamma^{\mu}\chi_{ijk}\bar{\chi}^{\ell mn}\gamma_{\mu}\chi_{\ell mn}\\
&+\sqrt{2}g\,A_1^{ij}\bar{\psi}_{\mu\,i}\sigma^{\mu\nu}\psi_{\nu\,j}+\frac{1}{6}\,g\,A_2{}^{i}{}_{jk\ell}\bar{\psi}_{\mu\,i}\gamma^{\mu}\chi^{jk\ell}\\
&+g\,A_3{}^{ijk\,\ell mn}\bar{\chi}_{ijk}\chi_{\ell mn}+\text{h.c.}\\
&+g^2\left(\frac{3}{4}\,A_1^{ij}A_{1\,ij}-\frac{1}{24}\,A_2{}^{i}{}_{jk\ell}A_2{}_{i}{}^{jk\ell}\right).
\end{aligned}
\tag{2.2}
$$

In the above Lagrangian, $\mathcal{O}^{\pm}_{\mu\nu\,KL}$ is a bilinear function of the fermionic fields ψ and χ, $S^{IJ,KL}$ is a function of the Higgs-fields, $u^{ij}{}_{IJ}$ and v_{IJKL} are pieces of the E_7 "vielbein" in the $\mathbf{56}\times\mathbf{56}$ representation that describes a point on the Higgs scalar manifold, while the $\mathcal{A}^{ijk\ell}_{\mu}$ are Higgs-scalar kinetic velocities. For details cf. [46, 102, 106].

This Lagrangian is a *consistent truncation* of 11-dimensional supergravity [100], i.e. the Kaluza-Klein modes retained here do not source higher modes, and so any solution of the four-dimensional field equations can be uplifted to an exact (non-linear) solution of the equations of motion of 11-dimensional supergravity. This is a "miraculous" property of the S^7 compactification for which the $F\wedge F\wedge A$-term in the Lagrangian plays an essential role. The gauge coupling constant g here is proportional to the inverse radius of the compactification manifold S^7 which, in Kaluza-Klein Supergravity, is not determined.

2.2 The scalar potential

In this work, we are mainly concerned with the $\propto g$ and $\propto g^2$ terms in the Lagrangian. At order g^1, we see Yukawa couplings that provide the (naive) gravitino and spin-1/2 fermion mass terms via their coupling to the Higgs-like scalars, $\sim gA_1\bar{\psi}\sigma\psi$, $\sim gA_3\bar{\chi}\sigma\chi$, and $\sim A_2\bar{\psi}\gamma\chi$. Here, the "spin 1/2 fermion mass matrix" $A_3^{ijk\,\ell mn}$, is given in terms of the gravitino-fermion Yukawa matrix A_2 as

$$
A_3^{ijk\,\ell mn}=\frac{\sqrt{2}}{144}\epsilon^{ijkpqr\ell' m'}A_2{}^{n'}{}_{pqr}\delta^{\ell mn}_{\ell' m' n'}.
\tag{2.3}
$$

At order g^2, we have the scalar potential

$$
V(\phi):=-g^2 e\left\{\frac{1}{24}A_{2i}{}^{jk\ell}A_{2\,jk\ell}^{i}-\frac{3}{4}A_1^{ij}A_{1\,ij}\right\}.
\tag{2.4}
$$

Since we are restricting ourselves in this work to the single case of the *compact* gauge group SO(8) of the original de Wit-Nicolai model [3], we can ignore a number of subtle aspects of electric/magnetic duality in four-dimensional supergravity that become relevant when trying to generalize our investigations to other gaugings in four dimensions, for details see [105–107]. The problem at hand then consists of finding critical points of the scalar potential $V(\phi_0,\ldots,\phi_{69})$, parametrized by 70 scalar coefficients of non-compact generators of the $\mathfrak{e}_{7(7)}$ algebra. In detail, the computation of the potential looks as follows, using

JHEP08(2019)057

the notational conventions of [108], *apart from index-counting always starting at* 0 *in this work*, in order to make the correspondence between tensor arithmetic and numerical code published alongside it even more straightforward.

$$V/g^2 = -\frac{3}{4} A_1^{ij} \left(A_1^{ij}\right)^* + \frac{1}{24} A_2{}^i{}_{jkl} \left(A_2{}^i{}_{jkl}\right)^* \tag{2.5}$$

with:

$$\begin{aligned}
A_1{}^{ij} &= -\frac{4}{21} T_m{}^{ijm} \\
A_{2\,\ell}{}^{ijk} &= -\frac{4}{3} T_\ell{}^{i'j'k'} \delta^{ijk}_{i'j'k'} \\
T_\ell{}^{kij} &= \left(u^{ij}{}_{IJ} + v^{ijIJ}\right) \left(u_{\ell m}{}^{JK} u^{km}{}_{KI} - v_{\ell mJK} v^{kmKI}\right) \\
\mathcal{V}^A{}_B &= \exp\left(\sum_{\mathtt{n}} \phi_{\mathtt{n}} g^{(\mathtt{n})}\right)^A{}_B \\
u_{ij}{}^{IJ} &= 2\,\mathcal{V}^A{}_B\, \delta^{\mathtt{m}}_A \delta^B_{\mathtt{n}} \delta^{\mathtt{ab}}_{ij} \delta^{IJ}_{\mathtt{cd}} \\
&\qquad \text{for} A < 28,\ B < 28,\ (\mathtt{a},\mathtt{b}) = Z(\mathtt{m}),\ (\mathtt{c},\mathtt{d}) = Z(\mathtt{n}) \\
u^{kl}{}_{KL} &= 2\,\mathcal{V}^A{}_B\, \delta^{\mathtt{m}}_A \delta^B_{\mathtt{n}} \delta^{kl}_{\mathtt{ab}} \delta^{\mathtt{cd}}_{KL} \\
&\qquad \text{for} A \geq 28,\ B \geq 28,\ (\mathtt{a},\mathtt{b}) = Z(\mathtt{m}-28),\ (\mathtt{c},\mathtt{d}) = Z(\mathtt{n}-28) \\
v_{ijKL} &= 2\,\mathcal{V}^A{}_B\, \delta^{\mathtt{m}}_A \delta^B_{\mathtt{n}} \delta^{\mathtt{ab}}_{ij} \delta^{\mathtt{cd}}_{KL} \\
&\qquad \text{for} A < 28,\ B \geq 28,\ (\mathtt{a},\mathtt{b}) = Z(\mathtt{m}),\ (\mathtt{c},\mathtt{d}) = Z(\mathtt{n}-28) \\
v^{klIJ} &= 2\,\mathcal{V}^A{}_B\, \delta^{\mathtt{m}}_A \delta^B_{\mathtt{n}} \delta^{kl}_{\mathtt{ab}} \delta^{IJ}_{\mathtt{cd}} \\
&\qquad \text{for} A \geq 28,\ B < 28,\ (\mathtt{a},\mathtt{b}) = Z(\mathtt{m}-28),\ (\mathtt{c},\mathtt{d}) = Z(\mathtt{n})
\end{aligned}$$

Here, we are using the auxiliary function Z to translate integer indices for the adjoint representation of $\mathfrak{so}(8)$ to ordered pairs of indices in the defining representation, with index-counting starting at zero,

$$\begin{aligned}
&Z(i \cdot 8 + j - (i+1)(i+2)/2) = (i,j), \\
&\text{i.e. } Z(0) = (0,1), \quad Z(1) = (0,2), \ldots, \quad Z(27) = (6,7).
\end{aligned} \tag{2.6}$$

The "input data" are the 70 $\phi_{\mathtt{n}}$ coefficients of non-compact $\mathfrak{e}_{7(7)}$ generators $g^{(\mathtt{n})}$. Even as in this work, we only use the non-compact and $\mathfrak{so}(8)$ generators of $\mathfrak{e}_{7(7)}$, we give a complete construction of the 133 56×56 generator matrices in appendix A, mostly to ensure that all subsequent investigations into alternative gaugings can all use the same definitions.

2.2.1 Equilibria of the equations of motion

When looking for viable 11-dimensional field configurations of supergravity that correspond to vacua of a four-dimensional theory, one is asking for solutions to the dynamical equations of motion in which, from the four dimensional perspective, all directional quantities are zero (since a "vacuum" should not have a preferred spatial direction) — so, we can set all four-dimensional gauge boson field strengths to zero, i.e. we are here not interested in "electrovacuum" [109] type solutions. Also, in this analysis, we set all fermionic (spin-1/2

JHEP08(2019)057

matter and spin-3/2 gravitino) fields to zero. We do not consider fermion condensates here. This leaves us with the need to pick a ground state on the 70-dimensional manifold parameterized by the Higgs-Boson-like scalars of the theory. Conceptually, one would want to look for minima of the scalar potential, but the actual story is slightly more involved here [110].

2.2.2 Vacuum stability

While the equations of motion for the scalar fields (and fields coupling to them) require the gradient of the potential to vanish in a vacuum configuration, it so turns out that viable vacuum states correspond not just to minima, but also some saddle points (and even a maximum at the origin!) in the potential. This is due to the value of the scalar potential playing the role of a cosmological constant in these models. So, for a negative cosmological constant, our vacuum will have the geometry of a space of constant negative curvature — an Anti-de Sitter (AdS) space. When studying stability with respect to small localized scalar field perturbations of finite total energy, one has to take into account that the spatial variation of such a perturbation can not be made arbitrarily small in an AdS background geometry. So, if a localized perturbation of a spatially constant background scalar field at a saddle point (or maximum) can decrease potential energy, the spatial gradient will lead to an increase in kinetic energy that cannot be made arbitrarily small. One finds that, overall, one can have (perturbative) stability even at a non-minimum critical point (i.e. $\nabla V(\phi_0) = 0$) as long as there is no direction $\delta\phi$ for which the 2nd derivative of the scalar potential (in a parametrization that gives a "conventionally normalized" kinetic term $\mathcal{L}_{kin} = \frac{1}{2}(\delta\phi)^2$) is smaller than a threshold known as the Breitenlohner-Freedman (BF) bound [111]:

$$m^2L^2 = -\frac{1}{2}(d-1)(d-2)\frac{V''(\phi_0)}{V(\phi_0)} \geq -\frac{1}{4}(d-1)^2, \tag{2.7}$$

which for $d = 4$ is $-9/4 = -2.25$. Here, L is the AdS radius, $L^2 = m_0^{-2} = -3/V(\phi_0)$ [100]. Loosely speaking, "masslessness" does not correspond to zero eigenvalues of the mass matrix in the curved AdS background. For a representation theoretic perspective and explanation, cf. [107].

In fact, it so turns out that for standard SO(8) supergravity (and many other Kaluza-Klein models), the potential does not seem to have any minima at all, but there are saddle points that give rise to AdS backgrounds in which this bound is satisfied. In particular, any background geometry with some residual supersymmetry will be stable and not violate this bound. To date, there is only a single known critical point of the scalar potential of SO(8) supergravity that corresponds to a stable non-supersymmetric AdS background [110, 112, 113]. While even this detailed investigation, which presents many more critical points, did not manage to reveal any other stable non-supersymmetric solutions, and there are good reasons to believe that they are indeed rare [114], there are indications that the method used here to search for solutions tends to (unfortunately) somewhat avoid parameter space regions that do correspond to stable critical points. This is, after all, how the new $\mathcal{N} = 1$ SO(3) vacuum escaped discovery in earlier investigations. So, the authors consider it possible (but unlikely) that there still are other such solutions that hide very well.

JHEP08(2019)057

2.2.3 Finding solutions

Historically, the most powerful approach to find critical points of supergravity potentials before a more effective strategy was presented in [115] was to introduce "Euler angle style" coordinate parameterizations of interesting submanifolds of the scalar manifold that have been selected according to group-theoretical considerations in such a way that critical points on the submanifold also will be critical points on the full manifold. While a full coordinate parameterization of $E_{7(7)}/(\mathrm{SU}(8)/\mathbb{Z}_2)$ is easily seen to be well outside computational reach, it is indeed feasible to consider the subgroup SU(3) of SO(8) in an $\mathrm{SU}(3) \subset \mathrm{SU}(4) \subset \mathrm{U}(4) \subset \mathrm{SO}(8)$ embedding and parameterize the six-dimensional manifold of SU(3)-invariant scalars. When Taylor expanding the full 70-dimensional potential around a point that is a critical point on such a subgroup-invariant submanifold, the linear term has to vanish, as the gradient then also decomposes into irreducible representations of the selected subgroup, but cannot carry any contributions that are not invariant under the chosen subgroup (since each term in the Taylor expansion is). This strategy was used in [116] to find all[7] critical points with residual symmetry at least SU(3), and led to the general belief that going substantially beyond this analysis by picking a smaller subgroup of SO(8) would be possible in principle, but technically very much infeasible, with perhaps only a few possible exceptions. This is due to the combinatorial explosion in algebraic complexity of explicit forms of coordinate-parametrized potentials as the number of coordinates increases.

Now that we know many critical points that have very little or even no continuous unbroken gauge symmetry at all, hindsight tells us that insisting on a fully analytic approach to solve a "discovery"-type problem limited our view. While an analytic approach easily becomes extremely complicated, all that complexity is eliminated by instead working with numerically evaluated quantities, and focusing on the use of backpropagation rather than analytic expressions in order to obtain gradients. Once one has good numerical data, one can start looking for corresponding exact expressions.

Critical points of the scalar potential correspond to (true or false) vacuum solutions, i.e. field configurations for which all directed quantities vanish, and the scalar fields do not experience any acceleration. While false vacua are unstable with respect to some small localized fluctuations that violate the BF bound, and the vast majority of critical points of SO(8) supergravity are indeed observed to be of this type, they are nevertheless interesting to study. In the past, we have learned much from such solutions. For example, the study of the SO(7) critical point S0698771 in [117] revealed the need to generalize the Freund-Rubin ansatz to include a warp factor, while some of the new solutions from [108] have been useful to identify and resolve subtleties in the uplifting from four to eleven dimensions in [100]. For some of the new solutions presented here, a deeper investigation into the nature of accidental (i.e. unrelated to any obvious symmetry) degeneracies in the mass spectra would seem appropriate.

So, while using the AdS/CFT correspondence to study e.g. condensed matter phenomena is doubtful if the AdS side, when embedded into M-Theory, has unstable modes

[7] There is a second way to embed SU(3) into SO(8), but this does not come with an invariant submanifold of scalars.

JHEP08(2019)057

(which may even be invisible in the truncation, as is the case for the SU(4) solution), one would nevertheless want to at least come to a deeper understanding of the 11-dimensional origin(s) of instability(-ies), perhaps even looking for ways of stabilization, cf. e.g. [118].

The scalars transform as a (reducible, nontrivial) representation of the gauge group SO(8), and a critical point with nonzero vacuum expectation values for the scalars will hence break the gauge symmetry to some subgroup of SO(8) via the Higgs effect. As the scalar potential has an overall SO(8) rotational symmetry, a shift in the scalar fields obtained by applying a small SO(8) rotation that actually moves the critical point on the scalar manifold, i.e. some generator of the SO(8) symmetry that is broken by the particular choice of the solution on its SO(8) orbit, corresponds to a flat direction in the potential. In the particle spectrum, these shifts would hence correspond to massless scalar ("Goldstone") particles, which however for a broken local (gauge) symmetry get absorbed ("eaten") by the gauge field to form the extra ("longitudinal") spin-1 polarization state that a massive vector boson has over a massless helicity-1 vector boson. Likewise, massless fermions get absorbed by the gravitinos to produce missing gravitino polarization states through the super-Higgs effect.

2.3 TensorFlow to the rescue

While we cannot use the supergravity potential directly as a ML loss function (since we are looking for saddle points, and not minima), it is possible to derive an expression that conceptually can serve as the length-squared of the gradient, $S := |\nabla V(\phi)|^2$, which can be used as a loss function and is reasonably easy to compute, cf. (2.21) in [119]:

$$\begin{aligned} S &:= |Q^{ijkl}_{(+)}|^2, \\ \text{where} \qquad Q^{ijkl}_{(+)} &= Q^{ijkl} + \frac{1}{24}\epsilon^{ijklmnpq}Q_{mnpq}, \quad \text{and} \\ Q^{ijkl} &= \left(\frac{3}{4}A_{2m}{}^{ni'j'}A_{2n}{}^{k'l'm} - A_1{}^{mi'}A_{2m}{}^{j'k'l'}\right)\delta^{ijkl}_{i'j'k'l'} \,. \end{aligned} \tag{2.8}$$

The Q^{ijkl}_{+} is the (self-dual) change of the value of the potential under an infinitesimal variation of the vielbein when multiplying with an infinitesimal E_7 element from the left, i.e. we are not considering $\delta V = V(\mathcal{V}(\phi+\delta\phi)) - V(\mathcal{V}(\phi))$, which would be the gradient of the potential with respect to the Higgs fields ϕ, but use the E_7 structure of the potential and rather consider

$$\delta V = V((1+\delta\mathcal{V})\cdot\mathcal{V}(\phi)) - V(\mathcal{V}(\phi)) \tag{2.9}$$

i.e. the change of the potential with respect to a small E_7 rotation applied to the vielbein matrix $\mathcal{V}$ from the left. As the 70 parameters of $\delta\mathcal{V}$ transform as self-dual 4-forms under the SU(8) subgroup of $E_{7(7)}$, the self-dual part of the tensor Q^{ijkl} that multiplies this variation to give the change to the potential has to vanish at a critical point. (This is also the variation one has to perform to get second derivatives at a critical point that correspond to actual particle masses, i.e. where the normalization of the kinetic term is the conventional one.) Since we want to compute the tensors A_1, A_2 anyway as part of the search procedure, e.g. to add a supersymmetry-encouraging term to the loss function as discussed later, this is

JHEP08(2019)057

straightforward to implement. A slightly less efficient strategy would be to ask TensorFlow for the length-squared of the gradient, which would (in "classic" TensorFlow graph-mode) perform a backpropagating transformation on the computational graph, costing roughly twice the memory, and six times the computation time.

From the ML perspective, minimizing the "stationarity violation" S of the potential then is a problem of just tuning 70 "learnable" parameters so that the stationarity condition is satisfied. While it is indeed possible to use the rotational SO(8) symmetry of the potential to further reduce this 70-dimensional optimization problem to a $70 - 28 = 42$-dimensional one, performing the search in the full 70 dimensions instead seems to make sense, as it is not very clear what a "good" random distribution to sample starting points from would be. Furthermore, even if one chooses to use SO(8) symmetry to (say) diagonalize the 35 pseudo-scalars, it may well happen that a critical point discovered in this way is more easily understood in a presentation that diagonalizes the 35 scalars. So, one should anyway always be able to diagonalize any solution for any of these two representations.

Performing numerical optimization in some 70-dimensional space looks like an unusually easy ML problem. Yet, there are some peculiarities:

- We are not interested in one minimum of the loss function, but (ultimately) want to know all inequivalent ones.
- The idea of "stochastic gradient descent" does not make sense in this setting: there is a well-defined gradient, but there are no "examples to perform well on", and therefore also no human-provided labels to tune towards.
- The loss function takes a highly uncommon form. In particular, its computation involves exponentiating a complex matrix (in a differentiable way).
- We are actually interested in high numerical accuracy in our numerically tuned "training parameters".

Our problem, then, is to:

- numerically find solutions to the $S = 0$ stationarity condition (2.8),
- canonicalize them to a form with few parameters, and obtain highly accurate numerical data, and
- extract information about physical properties (such as particle charges and masses) as well as (if possible) analytic expressions for the location of the solution.

Ideally, one would like the last step to at the very least produce sufficiently accurate numerical data to leave little doubt about the actual existence of a critical point — even if its location and properties are only approximately known. In the authors' view, seeing that the stationarity condition is satisfied numerically to better than 10^{-100} (as was achievable for most of the new solutions) is rather convincing.

Of the above steps, the first "discovery" step, when attempted without an efficient computational framework that can do automated backpropagation, would ask for manually

JHEP08(2019)057

re-writing Ricci-calculus code. While this is certainly doable by hand (as has been demonstrated with [108] and especially [115], which was published including hand-backpropagated code), it requires both effort and practice, and it certainly would be useful if this mechanical transformation were automated — especially when computations involve steps such as matrix exponentiation. Also, debugging hand-written gradient backpropagation code is often tedious, but at least straightforward, since one can always check the claimed sensitivitites in the backward pass by ad-hoc injecting an ϵ change into the associated quantity in the forward pass and observing the actual sensitivity.

Here, TensorFlow can help in these ways:

- We only need to write code for the computation of the loss function. All code that then computes the gradient efficiently is generated automatically.
- It becomes almost trivial to do exploration that requires computing gradients for scalar(!) quantities that are themselves defined in terms of gradients.
- Tensor arithmetic be executed on hardware that has been optimized to perform well on such tasks, such as in particular GPUs.
- Google Colab sandbox notebooks [120] simplify TensorFlow based code sharing and collaboration.

While TensorFlow also allows executing code on specialized Machine Learning hardware, such as Google's Tensor Processing Units (TPUs) [121], this is at present not an interesting option for this research here, since ML applications generally can work with much lower numerical accuracy than what is needed in this work, and so there is not a strong economic incentive towards high numerical precision for TPUs. Similarly, while quantum field theoretic problems often involve somewhat sparse tensors (in particular due to sparsity of Gamma matrices), the general trend in ML seems to be away from designs that rely on sparsely populated tensors, and so trying to exploit sparseness to improve computational efficiency when solving field theory problems like the ones studied here with TensorFlow may often not be worthwhile.

The second point above is interesting. As is known from the general theory of reverse-mode automatic differentiation of algorithms [84], it is always possible to compute the gradient of a scalar function that is described by an algorithm in a way that needs no more than some small constant k times the effort for evaluating the original function, *independent of the number of components of the gradient!* In practice, k somewhat depends on e.g. cache performance, and one typically finds $k \sim 5$, but *never* $k \geq 10$.

2.3.1 Simplifying basic analysis

For this work, masses of the scalars had to be determined in order to check whether any modes violate the BF bound (2.7). Still, no code had to be written to implement the mass

JHEP08(2019)057

matrix formula, eq. (2.25) from [119],

$$\begin{aligned}\mathcal{L}(\Sigma^2) = &-\frac{1}{96}\, g^{\mu\nu}\partial_\mu\Sigma_{ijkl}\,\partial_\nu\Sigma^{ijkl} - \frac{g^2}{96}\left(\left(\frac{2}{3}\,V + \frac{13}{72}\left|{A_{2\ell}}^{ijk}\right|^2\right)\Sigma_{ijkl}\,\Sigma^{ijkl}\right.\\ &+\left(6\,{A_{2k}}^{mni}{A_2}^{j}{}_{mn\ell} - \frac{3}{2}\,{A_{2n}}^{mij}{A_2}^{n}{}_{mk\ell}\right)\Sigma_{ijpq}\Sigma^{klpq}\\ &\left.-\frac{2}{3}\,{A_2}^{i}{}_{mnp}{A_{2q}}^{jk\ell}\Sigma^{mnpq}\Sigma_{ijkl}\right)\,. \end{aligned} \tag{2.10}$$

Rather, scalar masses were computed directly by just left-multiplying the vielbein matrix with an exponentiated e_7 generator Taylor-expanded to 2nd order only, and then using TensorFlow's `tf.hessians()` function to obtain the mass matrix. This performs 70 gradient computations each no more than six times as expensive as one evaluation of the potential starting from the unperturbed vielbein, rather than $\sim 70^2$ evaluations of the potential. In this sense, this work provides an independent confirmation for the correctness of (2.10), given that masses match values from the literature for critical points known earlier.

As the potential is exactly known, our gradients are not noisy estimates (as they usually are in ML), and it makes sense to employ an optimization method that can utilize this, i.e. conjugate-gradient optimization or BFGS optimization [122], which both try to use subsequent gradient evaluations to estimate the 2nd-order structure of the objective function. One convenient way to use TensorFlow as a "gradient machine" for various such higher order optimization methods is provided by the `tf.contrib.opt.ScipyOptimizerInterface()` helper function. One must be aware, however, that for degenerate minima of the objective function, these optimization methods are not expected to always perform well very close to the minimum, and given the rather special structure of the problem at hand, we may well encounter such degenerate minima.

Starting at randomly chosen locations on the 70-dimensional scalar manifold over and over again produces different critical points. For this work, the authors solved about 390 000 numerical minimization problems, each producing a critical point, that afterwards were deduplicated. Two solutions were considered equivalent if both the cosmological constant as well as the eigenvalue spectrum of the ${A_1}_{IJ}{A_1}^{JK}$ tensor were compatible to within the estimated numerical accuracy of a solution candidate. There are some cases of critical points with very similar cosmological constant, but no degeneracies arise at the finesse provided by the *Snnnnnnn* naming scheme that is used in this work for solutions.

Given location information for a solution-candidate that is good to more than about five decimal digits, the discovery problem can be considered solved, and one then has to deal with the subsequent problem of finding a highly accurate — ideally, analytic — form. In some cases, one finds that the geometry of a critical point is rather special, making it hard for a higher order optimizer to produce an accurate location. In such situations, it typically helps to run basic gradient descent (still with hardware floating point accuracy) as a post-processing step, which also can be done very efficiently with TensorFlow.

Using this approach, different critical points of the scalar potential get re-discovered many times over. One finds that the relative sizes of "basins of attraction" for different solutions are very different. While details do of course somewhat depend on the probability

JHEP08(2019)057

distribution used to generate starting points, one observes (for example) that the likelihood to end up at critical point $S1400056$ is about $100\times$ higher than the likelihood to end up at the $S1400000$ vacuum. Indeed, some of the solutions presented here were seen only *once*.[8] This makes it rather likely that just increasing the effort by another factor 10 would produce further solutions. Figuratively speaking, we suffer from some vacua strongly vacuuming in (pardon the pun) a large region of search space.

2.3.2 Loss function design

Given this situation, one naturally would like to have alternative approaches to investigate the structure of the scalar manifold. One idea — inspired by Morse Theory[9] [123] — is to look not for the minima of the scalar function that measures stationarity violation, but its saddle points, and then determine how following the gradient when starting from small perturbations along special unstable directions (such as the principal axes of the Hessian) carries one into different critical points of the potential. As it is plausible that a critical point with a small basin of attraction when minimizing stationarity violation may actually be reachable by walking down from a saddle that has a large basin of attraction in the search for such saddles, this change of perspective may offer a way to improve the efficiency of the search for overlooked critical points.

It turns out that implementing this idea in the most naive way is very easy with TensorFlow, requiring only very little coding while (due to backpropagation) still offering very good numerical performance. In order to give an impression of how little effort this is indeed, we show example code in appendix C. One notes that the corresponding calculations involve *third* derivatives of the potential (as the stationarity condition is a function of the gradient, and in order to determine its saddle points, we look at its gradient-squared, as well as the 2nd derivative (Hessian) of the stationarity condition). Still, as long as these derivatives get combined into intermediate scalar quantities (such as: length-squared of a gradient), the basic insight of reverse mode automatic differentiation holds, i.e. an extra derivative only multiplies the computational effort by a factor of about six (but retaining high numerical accuracy).

While naively following the gradient disrespects the underlying symmetry of our 70-dimensional space,[10] this may actually help rather than harm the search, with an eye on the intended purpose, by breaking up degeneracies in principal axes. With this "naive" saddle point approach, one observes that minimization is much more likely to run into a saddle than a minimum of the stationarity condition. Inspection makes it plausible that knowing the height of the saddle as well as the value of the potential to three digits after the point suffice to (mostly) deduplicate saddles, and with this, one can produce a "subway

[8] Specifically: $S2503105$, $S2547536$.

[9] One needs to keep in mind that critical points of the (un-adulterated) potential may well be degenerate, and not of the generic form required by Morse Theory. For example, even when removing the SO(8) degeneracy $S0800000$ has extra flat-to-2nd-order directions.

[10] The gradient is an element of the cotangent space, and, with our parametrization of the e_7 algebra, already at the origin, taking a step in the corresponding coordinate-direction is conceptually wrong, as it needs to be mapped back to an element of tangent space with the inverse scalar product of the non-orthogonal basis used here.

JHEP08(2019)057

map" of how one can cross from one critical point to another via some saddle. Irrespective of whether one uses the "physically correct" geometry on the scalar manifold or not, a (mostly) complete map is too complex to be fully visualized. Figure 1 provides a glimpse on what a tiny part of the graph looks like.

Generating 600 (non-unique) near-origin saddle points and then analyzing their 12 000 unstable principal axes did indeed confirm that some critical points which are hard to find by minimizing the stationarity condition are easier to obtain by this saddle point method. In particular, the odds for hitting the non-supersymmetric stable point raise from about 1 : 20 000 to about 1 : 600. This (limited) analysis did, however, not produce any new critical points in the near-origin region where the search was performed. In the authors' opinion, observing that a somewhat independent method only reproduces the solutions found with a straightforward random search, but fails to discover new ones, suggests that the list presented here likely is the near-complete answer to the question what the critical points of SO(8) supergravity are, at least in the near-origin region. That is, the authors expect the long list to likely still miss a few cases, perhaps even rather interesting ones,[11] but not to list only a small selection of critical points that happen to be strongly attractive in a random search.

Likewise, TensorFlow makes it very simple to tweak loss functions in order to search for points on the scalar manifold with specific desired properties. Clearly, one would like to know whether the current work now gives a complete list of the supersymmetric vacua of SO(8) supergravity. While the methods employed here are insufficient to stringently prove this, it is very easy to tune the search to strongly favor supersymmetric critical points. A straightforward way to do this is to replace the length-squared-of-the-gradient loss function $L_0 = |\nabla V|^2$ with a loss function that includes another term which is zero for supersymmetric solutions only. The obvious idea here is that, for a supersymmetric solution, there needs to be a massless gravitino, i.e. some vector η_K such that

$$L_S := \left| A_1{}^{IJ} A_{1IK} \eta^K + \frac{V}{6g^2} \eta^K \right|^2 = 0. \tag{2.11}$$

Due to the SO(8) symmetry of the potential, we do not have to compute an eigenvector in a differentiable way here, but can simply fix $\eta_K = \delta_{K0}$ without loss of generality. Using not L_0 but $L_0 + \lambda L_s$ with a BFGS optimizer is indeed observed to be extremely effective for finding supersymmetric solutions. Taking $\lambda \sim 10$, and starting from a randomly picked 70-dimensional vector (with coordinates drawn from a normal distribution), with uniform distribution of a length multiplier, one observes that numerical optimization occasionally does get caught in a new local minimum with $L_0 > 0$ (i.e. not a critical point), but otherwise manages to find each of the known supersymmetric vacua multiple times with in less than an hour of computing time on moderately recent hardware. This approach also unearths one additional supersymmetric vacuum (which also is found many times

[11] Soon after the first release of a preprint of this article, follow-up work [124] as a by-product indeed gave early evidence for the existence of two further unstable critical points not listed here, $S2096313$ with $\mathrm{SO}(3) \times \mathrm{U}(1)$ symmetry, and $S2443607$ with SO(3) symmetry. These solutions will be discussed in the upcoming article.

JHEP08(2019)057

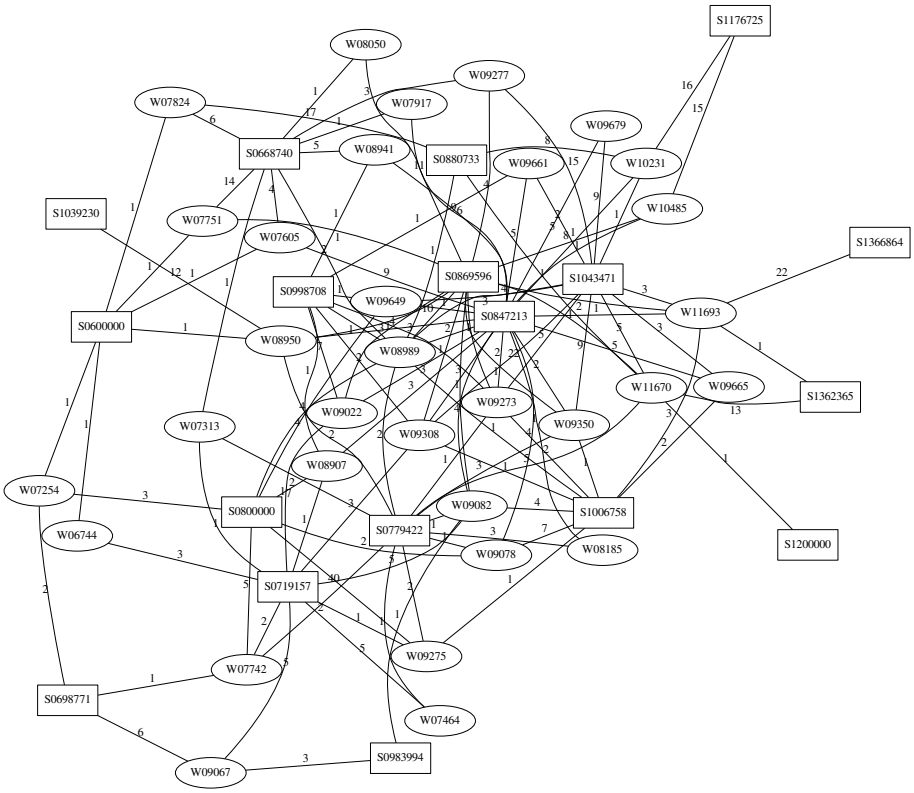

Figure 1. A tiny part of the "watershed map" of critical points of the stationarity condition. Minima of the stationarity condition (at value zero) are (true and false) vacua of the potential, and labeled using the naming scheme introduced in [125]. Saddles are also labeled by cosmological constant, e.g. the saddle at $-V/g^2 \approx -7.605$ corresponds to $W07605$. Edge labels indicate how many gradient-parallel (in naive geometry) paths along a principal axis run into a particular critical point.

over) that has $\mathcal{N} = 1$ supersymmetry and breaks SO(8) to SO(3). This is the solution named S1384096, see section 3 and the appendices for properties. Running this search for a day on a single computer produced 7150 supersymmetric solutions, with each of the now five solutions (SO(8), G_2, SU(3) × U(1), U(1) × U(1), SO(3)) being discovered many times over, the lowest count being 318 for S0600000.

Is it also possible to directly encode BF-stability as a ML loss function, and hence directly search for stable vacua in a similar way? In principle, this can be implemented e.g. by adding to the stationarity-violation loss L_0 another non-negative contribution L_{BF} that can only be zero if the scalar mass matrix S with all eigenvalues shifted up by the BF bound is positive semidefinite, i.e.

$$L_{\text{BF}} := \sum_{A,B} \left| S_{AB} + \frac{9}{4} I_{AB} - \Lambda_{CA}\Lambda_{CB} \right|^2 , \tag{2.12}$$

where one introduces a lower triangular matrix Λ of $70 \cdot 71/2 = 2485$ trainable parameters that will, when minimizing the loss, try to (Cholesky-)factorize the shifted mass matrix. Unfortunately, the associated cost that comes with this large increase in the number of

JHEP08(2019)057

training parameters means that loss minimization becomes (in comparison) painfully slow. One notes that the mass matrix in general unfortunately is not $35^+, 35^-$-block-diagonal. We anticipate that this technique might become useful for problems with smaller scalar sectors, such as perhaps maximal gauged $D = 5$ supergravity, but not for SO(8) supergravity in $D = 4$. Still, there are other minor (fixable) annoyances with the basic form of this loss contribution, such as bad behavior in the $V > 0$ region, and the new term driving the search too fast towards the origin.

There are many more ways in which being able to effortlessly engineer loss functions might help. For example, it might be feasible to multiply the stationarity-violation with an extra factor that increases as search approaches known strong attractors, effectively reducing the size of their basin of attraction. One obvious way in which this could be realized would be to add factors of the form

$$|Q_{ijkl}|^2 \cdot \left(\sum_{n,p} f\left(\mathrm{tr}\, \left(A_{1\,ij} A_1{}^{ik} \right)^p - c_{n,p} \right) \right) \tag{2.13}$$

where the sum over p runs over (the first few) powers of the gravitino mass matrix that we use to "fingerprint" solutions, the $c_{n,p}$ is the corresponding known fingerprint-value for the n-th known too-attractive solution that should be punished in the search, and f is some function with $f(x) \approx 1$ away from 0 and $f(x) \gg 1$ near 0. Some experimenting will be needed to find an approach that does not create many new nonzero minima of the loss function.

2.4 Canonicalization

For any critical point obtained by a numerical search, the SO(8) symmetry of the scalar potential allows us to freely pick an arbitrary point on its SO(8) orbit as an equivalent presentation. Naturally, one would want to use a form that allows describing the solution with a minimal number of parameters. This is not only desirable for typographic compactness, but also establishes the connection with simple exact analytic descriptions of these critical points. Setting an additional coordinate on $E_{7(7)}/(\mathrm{SU}(8)/\mathbb{Z}_2)$ to zero corresponds to imposing an extra algebraic constraint on the solution, and using sufficiently many such constraints to eliminate all freedom to rotate a solution produces a 56-bein matrix with only algebraic entries, since the defining properties of the 56-bein, i.e. belonging to $E_{7(7)}/(\mathrm{SU}(8)/\mathbb{Z}_2)$, can also be expressed through algebraic constraints. Specificaly, the 56-bein respects the symplectic invariant of Sp(56) as well as Cartan's quartic invariant of $E_{7(7)}$ (e.g. (B.4b) in [37]),

$$\begin{aligned} \mathbf{56} \to\ & (\mathbf{28}, \overline{\mathbf{28}}) : (x_{ij}, y^{kl}) \\ I_4 \ =\ & x_{ij} y^{jk} x_{kl} y^{li} - \frac{1}{4} \left(x_{ij} y^{ij} \right)^2 \\ & + \frac{1}{96} \left(\epsilon_{ijklmnpq} y^{ij} y^{kl} y^{mn} y^{pq} + \epsilon^{ijklmnpq} x_{ij} x_{kl} x_{mn} x_{pq} \right) . \end{aligned} \tag{2.14}$$

At the Lie algebra level (i.e. prior to exponentiation), this then means that there are exact analytic expressions for the coordinate-parameters describing a given solution, typically

JHEP08(2019)057

of the form {algebraic number} × log{algebraic number}. Given that the actual 56-bein entries may well be determined by rather complicated intersections of many algebraic varieties, actually finding algebraic forms may well be computationally out of reach in some cases (i.e. one may well imagine to encounter zeros of irreducible polynomials of degrees well beyond 1000). Still, for some of the new solutions described here, the authors were able (with reasonable computational effort) to determine analytic expressions from high-precision numerics alone. Each such expression is correct with overwhelming likelihood.

This procedure starts with first obtaining high-precision (hundreds to thousands of correct digits) numerical data for quantities that are known to be algebraic (i.e. vielbein entries and derived quantities, such as the cosmological constant). Unfortunately for us, as extremely high numerical accuracy is generally not very relevant for ML, TensorFlow does not support tensor arithmetics with higher numerical precision than what common hardware can provide, i.e. IEEE-754 double precision floating point. In that sense, from the perspective of M-Theory research, TensorFlow perhaps is best thought of as a "discovery machine" and not a "precision machine", carrying over terminology from accelerator physics (e.g. [126, 127]).

In principle, it would be doable to run already the "discovery" computation with adjustable accuracy, computing e.g. the algebraic entries of $\mathcal{V}$ to hundreds of decimal digits. This technique has partly been employed in [108], bases on highly performant compiled Common Lisp code in conjunction with an adapter library that allows a common generic limited-precision numerical optimizer to work in an high precision setting, but all this generally involves carefully performing the code transformations needed for sensitivity backpropagation by hand.[12] This technique that produces algebraic expressions in a fully automatized way should hence not (yet) be considered as being easy to apply and widely accessible. It hence makes sense to aim for a clearer separation of the "discovery" and "precision" steps.

2.5 Parameter-reducing heuristics

An approximate location of a solution obtained by the "discovery" step, once suitably rotated to be coordinate-aligned to the largest possible extent (using the procedure described further on), gives us an idea about what coordinates on the scalar manifold can be set to zero, and what others can likely be set to identical values (or simple rational multiples of one another). One finds that most solutions are very non-generic and allow very many such simplifying linear identities. In terms of automated processing of many solution-candidates, this then requires code that tries some basic heuristics (and automatically abandons them when they turn out to not actually hold at high precision). The basic process is to go through all coordinates, check if an observed coordinate is close to another one seen earlier (or a simple rational multiple thereof, or zero) within some tolerance limit τ such as $10^{-3}, 10^{-4}, 10^{-5}$, etc. If so, the observed coincidence is assumed to hold, and codified in a linear model-parameters-to-solution-coordinates matrix map. If

[12] At the time of this writing, trying to combine the "mpmath" [128] and "autograd" [129] Python libraries in order to achieve this does not work.

JHEP08(2019)057

the attempt to improve accuracy based on such a model map runs into a dead end, the process is restarted with a less permissive tolerance limit τ. This "automated heuristic modeling" step typically reduces a 70-dimensional optimization problem to a much more manageable problem in 2–20 or so parameters, for which obtaining high-accuracy data is very often feasible even without having fast gradient computation available. The most important techniques here are using a multidimensional Newton solver (as provided by the "mpmath" [128] package), and high-precision Nelder-Mead optimization, which is feasible for up to about 14 parameters. If both these techniques fail, it is sometimes useful to use basic fixed-scaling (TensorFlow-based) gradient descent with hardware numerics to turn a solution that is good to eight digits into one that is good to at least twelve. One finds that basic gradient descent with some simple heuristic to make learning rates adaptive indeed seems to work better for this problem than any of the more advanced minimization methods that are currently popular in ML applications, e.g. Adam, RMSProp, AdaGrad, FTRL. In some situations, the parameter reducing heuristic produced a problematic canonical form, and one has to start over with canonicalization after applying a random SO(8) rotation to the solution.

This "distillation" step produces a high degree of evidence for the existence of a particular critical point (the length of the potential's gradient having been shown numerically to reach values typically below 10^{-20}), as well as a first highly accurate location described by only a few numbers, and also information on whether the solution is sufficiently well-behaved (i.e. non-degenerate) for the multidimensional Newton method to allow quick determination of coordinates and physical properties such as the cosmological constant to an accuracy of hundreds of digits.

2.6 Coordinate-aligning rotations

As a point on the scalar manifold can be described by giving two symmetric traceless matrices, one carrying the $\mathbf{35}_s$ and one carrying the $\mathbf{35}_c$ representation of $\mathfrak{so}(8)$, and we can always use a SO(8) rotation to diagonalize one of these, the effective dimension of the scalar manifold relevant for finding critical points is reduced to $70 - 28 = 42$. Still, even if one exploited this symmetry from the start and only looked for critical points for which one of $\mathbf{35}_s, \mathbf{35}_c$ is diagonal, this does not eliminate the need to numerically canonicalize a solution, as any degeneracy in the entries of the diagonal matrix would leave some residual rotational symmetry that can be used to reduce the number of non-zero entries in the other diagonal matrix. For example, if the parameters in $\mathbf{35}_s$ can be brought into the form $\mathrm{diag}(3A, 3A, -A, -A, -A, -A, -A, -A)$, this still leaves a residual symmetry of $\mathrm{SO}(2) \times \mathrm{SO}(6)$ which must be fixed by imposing algebraic constraints on the $\mathbf{35}_c$ in order to make the entries of the 56-bein matrix algebraic.

As it hence is difficult, in a numerics-based search, to avoid the need for a "canonicalization" step that eliminates residual rotational freedom, we may just as well fish for solutions in full 70-dimensional parameter space. While our particular choice of e_7 generators leaves us with a non-diagonal scalar product on the 70-dimensional manifold of

JHEP08(2019)057

scalars,[13] we nevertheless start the search with a 70-vector picked at random from a distribution that is isotropic with respect to the coordinate-basis, not the restricted e_7 Killing form. This choice is apparently "good enough" to find many new solutions.

For elements of a 35-dimensional irreducible representation of SO(8), it is easy to numerically find a rotation G_d that diagonalizes the corresponding symmetric traceless matrix. The orthonormal eigenbasis serves this purpose if we multiply the last eigenvector with ± 1 in order to ensure a positive determinant. Knowing how to diagonalize, say, $\mathbf{35}_s$, how do we then find the corresponding action on the $\mathbf{35}_c$ (and vice-versa)? Logarithmizing the group element to obtain the algebra element is out of the question, but we can employ a higher-dimensional generalization of Davenport chained rotations [130, 131] to write G_d as a product of a sequence of up to 28 rotations in coordinate-aligned planes,

$$G_d = R_{67}(\alpha_{67}) \cdots R_{57}(\alpha_{57}) \cdots R_{56}(\alpha_{56}) \cdots \cdot R_{02}(\alpha_{02}) \cdot R_{01}(\alpha_{01}), \tag{2.15}$$

each of which, when moved to the left size, cancels another off-diagonal entry of G_d without destroying earlier such reductions.[14] Using this presentation, we can then proceed to logarithmize each factor

$$R_{\gamma\delta}(\alpha_{\gamma\delta}) = \exp\left(r_{\gamma\delta} \cdot \alpha_{\gamma\delta}\right) \tag{2.16}$$

and find the corresponding Lie algebra action on the 8_c representation by employing the $\gamma^{\alpha\beta}_{\dot\gamma\dot\delta}$ invariant. Lifting this to an action on $\mathbf{35}_c \subset (\mathbf{8}_c \times \mathbf{8}_c)$ and exponentiating, we can readily determine the action of G_d on $\mathbf{35}_c$. An alternative approach would be to run numerical minimization starting from Lie algebra elements that get exponentiated to obtain group actions on the $\mathbf{35}_s$ and $\mathbf{35}_c$ with an objective function that punishes off-diagonal elements for the matrix carrying the $\mathbf{35}_s$ representation.

Once the $\mathbf{35}_s$ has been diagonalized by any such method that also gives us the effect of the rotation which was employed on the $\mathbf{35}_c$, we can proceed to determine the residual subalgebra of $\mathfrak{so}(8)$ that keeps the diagonalized $\mathbf{35}_s$ unchanged. We can employ this subalgebra to reduce the number of off-diagonal entries for the matrix carrying the $\mathbf{35}_c$ representation, but in general not completely. Also, this residual symmetry group will typically be rather small — such as $\mathrm{SO}(3) \times \mathrm{SO}(2)$. It hence makes sense to consider this step as a somewhat low-dimensional numerical optimization problem. In general, problems that maximize the number of zero entries in a matrix often are hard, but here, a common sparseness-encouraging ML technique works reasonably well: we use the L_1-norm of the off-diagonal matrix entries as a loss function.

After this sparseness-encouraging rotation, which will in general have put many more than 28 coefficients to zero, we proceed by making an automated guess for the form of the symmetric matrices carrying the $\mathbf{35}_s$ and $\mathbf{35}_c$ representations as described.

2.7 "Algebraization"

Once an highly-accurate numerical value for a known-to-be-algebraic parameter has been found, one can use an integer relation algorithm such as PSLQ [132] to find a polynomial

[13] It is orthonormal up to an overall factor, and the non-orthogonal inner products of the basis vectors that correspond to the $\mathrm{diag}(1,-1,0,\ldots,0)$, $\mathrm{diag}(0,1,-1,0,\ldots,0)$ parts of the symmetric traceless matrices.

[14] For better numerical stability, one should re-order processing of row-entries by absolute magnitude.

JHEP08(2019)057

of which this is a zero and that seems plausible, i.e. the total information content of its coefficients is much smaller than the information content of the known digits of the parameter. This works well for up to a few thousand decimal digits. More specifically, one scans for a set of integer coefficients $c_j, j \in \{0, 1, \ldots, N-1\}, |c_j| < 10^d$ such that for a number ξ known to D (perhaps $D \sim 300$) digits after the point, with $D > N \cdot d + 30$, we have $|\sum_j c_j \xi^j| < 10^{-(Nd+10)}$. If such a polynomial is found, we can easily determine its actual zero to D digits after the point, and check if this polynomial correctly predicts many further digits of ξ that were known but not used to find the c_j. Naturally, if there is a single candidate polynomial that was found by using 300 known good digits of precision manages to predict the next 50 digits, we would expect this to happen purely by chance to be $\sim 1 : 10^{50}$. So, by using a large enough reservoir of extra precision, we can make the likelihood to accidentally predict an incorrect algebraic expression fantastically small. While this still does not constitute a strict mathematical proof, it would be rather unreasonable to disbelieve the result. Of course, it will in many cases then be possible to independently establish the validity of a claimed exact expression, but this will generally require ingenuity and effort beyond what can easily be automated to process scores of solutions.

2.8 Tweaks to the basic procedure

In general, it makes sense to use moderately refined numerical data (e.g. known to be good to 14 digits of accuracy) and then repeat the entire procedure that starts with finding a coordinate-aligning $\mathfrak{so}(8)$ rotation from cleaned up numerical data, as limitations on numerical precision of input data from the ML library data may have caused the (imperfect) heuristic that suggests a low-dimensional model to miss some reduction opportunities. In particular, starting from an already partially parameter-reduced model will lead to rather short products of Davenport chained rotations, which are numerically much better behaved than those seen in the initial reduction step.

Once highly-accurate and nicely coordinate-aligned numerical data are available, inverse symbolic computation methods can automatically search for exact expressions for physically relevant properties such as the cosmological constant, coordinates, particle masses and charges (and hence also residual supersymmetry), and others.

2.9 Extracting the physics

Unbroken continuous gauge symmetries can be determined numerically by mostly straightforward methods, and Lie group theory then allows an automatic classification. In a first step, one determines the space h of $\mathfrak{so}(8)$ generators that leave a given solution invariant and splits it into orthogonal pieces w.r.t. the Cartan-Killing metric $h = [h, h] + h'$. This separates off the generators of the $\mathrm{U}(1)^N$ part of the residual symmetry. For all new solutions described here, the dimension of the semisimple $[h, h] =: \tilde{h}$ part is 0, 3, or 6, so the corresponding non-abelian symmetry algebra can only be $k \cdot \mathfrak{so}(3), k \in \{0, 1, 2\}$, but perhaps embedded into $\mathfrak{so}(8)$ in different ways. The authors' automated analysis handles this case by looking for a maximal set of p orthogonal and commuting $\tilde{h}$-elements $\tilde{h}_{c,j}, j \in \{0, \ldots, p-1\}$ (which thanks to $\mathfrak{so}(8)$ being the algebra of a compact Lie group are simultaneously diagonalizable in the adjoint representation), then splitting $\tilde{h}$ into subspaces that are simulta-

JHEP08(2019)057

neous eigenspaces for these p generators, labeled by p-dimensional vectors of eigenvalues. After defining a subspace of positive roots and identifying simple roots, taking the commutators of raising and lowering operators associated with simple roots produces a basis for the Cartan subalgebra that is useful for numerically identifying angular momentum spectra, which for $k \cdot \mathfrak{so}(3)$ are straightforward to map to the irreducible representation content. For $\mathrm{U}(1)^N$ generators, the $N = 2$ case will in general require finding a rotation that coordinate-aligns the charge lattice. In any case, we scale every U(1) generator u such that $\exp(2\pi u)$ is the identity on all particle states, while no λu with $0 < \lambda < 1$ also has this property. For solutions whose gauge group has a single U(1), we use superscripts to indicate electric charge of particle states, while for solutions with $\mathrm{U}(1) \times \mathrm{U}(1)$ abelian gauge symmetry, we use super- and sub-scripts for the two different types of charges.

Having split the unbroken symmetry in this way, we can proceed to present branching rules for the $\mathbf{8}_{v,s,c}$ as well as $\mathbf{28}$, and decompose particle mass-eigenstates for the spin-1/2 fermions, spin-3/2 gravitinos, and spin-0 scalars in terms of irreducible representations of the residual gauge algebra. For the scalars, we first try to split mass eigenspaces into orthogonal subspaces of the parity operator P that is $+\mathbf{I}$ on $\mathbf{35}_s$ and is $-\mathbf{I}$ on $\mathbf{35}_c$. However, the analysis presented here only splits off the pure $P = 1$ and $P = -1$ subspaces and subsequently merges their joint orthogonal complement into one subspace. Mass eigenstates are then determined on these three subspaces, and marked with a superscript of (s), (c), or (m) for the "mixed" case, respectively. This step is numerically slightly problematic, as it so turns out that, for some critical points, we find $P = 1 - \epsilon$ eigenstates with $\epsilon < 10^{-3}$. Hence, auto-identification of some mass eigenstate as a pure scalar or pure pseudoscalar state may be somewhat unreliable. It is nevertheless reassuring that decomposition of weights into irreducible representations is observed to always succeed, which it would not if some P-eigenvectors were wrongly assigned to subspaces. All particle masses are known with an accuracy of more than eight digits, but it may happen that different particle masses look the same when truncated to three digits after the point for presentation. This explains why some solutions list the same mass as belonging to more than one mass-subspace. This happens e.g. for $S1039624$, which lists scalars with m^2/m_0^2 of $-2.417^m_{\mathbf{1}^{+\!+\!+\!+}\mathbf{+1}^{-\!-\!-\!-}}, -2.417^m_{\mathbf{1}}$, rather than $-2.417^m_{\mathbf{1}^{+\!+\!+\!+}\mathbf{+1}^{-\!-\!-\!-}\mathbf{+1}}$ i.e. there is a two-dimensional subspace of mass eigenstates with electric charges ± 4 and another one-dimensional mass eigenstate with a too-close-to-be-discriminated-at-presentation-accuracy mass. For a solution to be perturbatively stable, no scalar mass-eigenstate must have m^2/m_0^2 that violates the BF bound of $-9/4$. Unstable mass-eigenstates that violate this bound are marked with an asterisk $(*)$ in the tables.

As none of the new solutions has a residual gauge group that contains a simple group of rank ≥ 2, and so apparently all such solutions already have been identified and studied in detail in the 80's [116], there is no compelling reason to automate assignment of quantum numbers to such larger symmetry groups.

Residual supersymmetry can be identified [133] numerically by using Singular Value Decomposition (SVD) to look for solutions of

$$\eta^i A_{2\,i}^{jkl} = 0\,. \tag{2.17}$$

The corresponding gravitino states are marked with an asterisk in the tables.

JHEP08(2019)057

3 A guide to the new solutions

Detailed data on all the solutions obtained in a large TensorFlow based cluster search are presented in appendix D.

The structure of the location data presented in these tables is explained in section 2.4. Given the sheer amount of data (masses and charges for 26k particles!), it makes sense to include a lookup table that lists the most important properties. This is presented in appendix D. Let us in this section highlight some specific examples with interesting properties.

S1384096 is a new stable vacuum with $\mathcal{N} = 1$ supersymmetry. This is clearly the most remarkable new discovery. As the 10-parameter form presented here resisted all attempts to increase numerical accuracy by employing a multidimensional Newton solver, a computationally rather expensive Nelder-Mead optimization had to be performed to extract 500+ digits of numerical accuracy. With this, PSLQ based analysis (using 400 digits of accuracy to predict 120 further digits) was able to determine the cosmological constant algebraically as a root of the following (rather remarkable!) polynomial:

$$\begin{aligned}
v &:= -V/g^2 \approx -13.84096, \qquad w := v^4 \\
2^{20}\cdot 3^{30} &= 5^{15}\,w^3 - 2^8\cdot 3^4\cdot 7\cdot 53\cdot 107\cdot 887\cdot 1567\,w^2 + 2^{15}\cdot 3^{17}\cdot 210719\,w \\
\Longrightarrow \quad v &= \frac{2}{3125}\cdot 13500^{1/4}\cdot 5^{1/2} \\
&\quad \cdot\{2^{1/3}\cdot(1173197938392473578665161132812 5\cdot 129^{1/2} \\
&\qquad +48821817290865578054293157348181 79)^{1/3} \\
&\qquad +22836248051085301205852\cdot 2^{2/3}\cdot D^{-1/3} + 220704046052\}^{1/4} \\
\text{where} \quad D &:= 1173197938392473578665161132812 5\cdot 129^{1/2} \\
&\quad +48821817290865578054293157348181 79. \qquad (3.1)
\end{aligned}$$

Further, the gravitino masses are roots of these algebraic expressions:

$$\begin{aligned}
\mu := m^2/m_0^2[\psi] = \{&\text{A zero of}\dots\} \\
&\mu - 1 && (\mu = 1.0), \\
&9\,\mu^3 - 48\,\mu^2 + 80\,\mu - 48 && (\mu \approx 2.90620272), \\
&4\,\mu^3 - 28\,\mu^2 + 64\,\mu - 49 && (\mu \approx 3.17965204), \\
&36\,\mu^3 - 216\,\mu^2 + 412\,\mu - 321 && (\mu \approx 3.41145711).
\end{aligned} \qquad (3.2)$$

The gauge group SO(3) is embedded in a triality-invariant way as the diagonal subgroup of a $\mathrm{SO}(3)\times\mathrm{SO}(3) \subset \mathrm{SO}(8)$. There are ten SO(3)-invariant scalars which all have different masses apart from a pair of two with $m = 0$. A detailed study of the analytic properties of this new solution is in progress [124]. Remarkably, **S1384135**, which has only a minimally lower cosmological constant has a Gravitino state that *almost* would satisfy the Killing spinor equation (no other Gravitino in the long list comes this close) and also has its gauge group — here, U(1) — embedded in a triality-invariant way. The mass spectra of these two solutions are very similar.

JHEP08(2019)057

One notes that all the cosmological constant polynomials that have been identified have rather special form which, given their 2-3-5-factorizations, somehow seems to be suggestive of an underlying E_8 structure.

S1600000 and **S1800000** extend the known list of solutions with rational (even integer) cosmological constant by two new entries, the others being **S0600000** with $\mathcal{N} = 8$ SO(8) symmetry, **S0800000** with SU(4) symmetry, **S1200000** with $\mathcal{N} = 1$ $\mathrm{U}(1) \times \mathrm{U}(1)$ symmetry, and the only known stable non-supersymmetric point **S1400000**, which has $\mathrm{SO}(3) \times \mathrm{SO}(3)$ gauge symmetry (listed as SO(4) in the summary table). Despite both **S1600000** and **S1800000** having no residual gauge symmetry, they have rather remarkable and currently unexplained degeneracies in particular in the fermion mass spectra. Such "accidental degeneracies" also occur for some other critical points, such as **S1046018** or **S1176725**.

The **S0847213** solution is the only "modern" critical point with residual gauge symmetry for which the $\mathbf{8}_{v,s,c}$ branching does *not* have any residual triality symmetry. Also, it is the only case with a 3-component gauge group.

The **S1039230** solution has been first described in [134]. In total, there are three known critical points with residual gauge symmetry SO(4). The gauge group embedding has $V \leftrightarrow S$ triality symmetry for this solution, while it has $V \leftrightarrow C$ triality symmetry for the **S0880733** solution and $S \leftrightarrow C$ symmetry for the (BF-stable!) **S1400000** solution.

S2099419 and **S2099422** are a pair of critical points with extremely similar but nevertheless different particle properties and also cosmological constant. There are a few other examples for critical points with very small difference in the cosmological constant and similar properties, such as the pair **S2511744**-**S2512364**, or **S3254262**-**S3254576**.

For almost every solution, the number of scalar modes with $m^2/m_0^2 = 0$ matches the number of broken $\mathfrak{so}(8)$ generators, as it has to according to Goldstone's theorem. Two solutions have extra massless (w.r.t. "naive mass", not AdS-massless) modes that survive being eaten by gauge bosons: **S0800000** (i.e. the SU(4) critical point) and **S1200000** (i.e. the $\mathcal{N} = 1$ $\mathrm{U}(1) \times \mathrm{U}(1)$ vacuum).

Solutions **S1200000**, **S2279257** and **S2279859** are listed in the summary table with symmetry $\mathrm{U}(1)_1$, since in these cases, the charge of all U(1)-charged particles is the same in magnitude. Still, these particle-states are listed as having charges $\mathbf{1}^{++}$ etc. in the tables for typographic reasons, due to the $\mathbf{8}_{s,c}$ representations branching to $4 \times \mathbf{1}^{+} + 4 \times \mathbf{1}^{-}$, but the spectrum not having any particles transforming under these representations.

While this expanded list did not uncover any other stable non-supersymmetric critical points beyond the known $\mathrm{SO}(3) \times \mathrm{SO}(3)$ solution **S1400000**, we observe a another instance of the phenomenon that instabilities can become invisible when truncating a solution to the scalar submanifold that is invariant under the unbroken gauge group. This was first observed and discussed in detail for **S0800000** in [118]. There, the unstable scalars come as a single 20-dimensional irreducible representation of SU(4), and the question was raised whether one could project out these unstable modes with an orbifold construction which would have to use a non-abelian discrete subgroup of SU(4) other than the Weyl group (since in these cases, the **20'** decomposes with singlets). This phenomenon also happens for the SO(7) critical points **S0668740** and **S0698771**, where the unstable modes trans-

JHEP08(2019)057

form as a **27**, as well as for **S1424025**, where the unstable modes form a single **5** of SO(3). Clearly, the analysis of all viable discrete subgroups of the gauge group under which this irreducible representation branches without singlets is greatly simplified by going from SU(4) to SO(3), where there is an obvious candidate symmetry (namely the icosahedral group) under which $\mathbf{5} \to \mathbf{5}$.

Remarkably, **S2416856** has *no* residual gauge symmetry, but six unstable scalar modes with 1+2+3 mass-degeneracy. While trying to find a way to stabilize this solution seems hopeless here, it would be interesting to understand what symmetry is responsible for the accidental mass degeneracies in the unstable modes.

The solutions **S1195898** and **S2503105** feature a peculiar SO(3) gauge symmetry which is embedded in a triality-invariant way, with branching $\mathbf{8}_{v,s,c} \to \mathbf{3}+\mathbf{5}$. In both cases, there are only two SO(3)-invariant scalars, but unfortunately, one of them is unstable. This is also the only instability of these solutions. Remarkably, the minimal polynomial for the cosmological constant is *identical* for these two solutions, so this seems to be an example of the algebraic equations for a vanishing gradient allowing a pair of real solutions that are Galois conjugates. It is noteworthy that even in the large numerical search performed here, **S2503105** was a chance hit that only showed up *once*. Had it been missed, it is conceivable that it ultimately might have been found in a detailed study of **S1195898** as an algebraically equivalent solution. So, looking for Galois conjugates seems to be one new method to fill possible gaps in the list of solutions.

The same Galois-doubling phenomenon occurs also for the pair **S1039624**–**S1402217**, which both also have identically embedded gauge group. One might speculate that Galois conjugates may well occur more often, in particular with solutions without residual symmetry and a cosmological constant that is not known algebraically. Perhaps then, it may be possible to extract algebraic numbers that are easier to identify from pairs of numerically known cosmological constants of critical points with identically embedded gauge group.

A rather unique property of the **S4168086**-solution will be the topic of a short follow-up article.

4 Conclusions and outlook

Some problems in the world are very amenable to ML based approaches, others not so much. As we are, during the current ML revolution, working on finding out which is which, we sometimes encounter pleasant surprises. Being able to present an elegant way to address a fundamental need in quantum gravity research — the need to be able to analyze potentials on complicated high-dimensional scalar manifolds — certainly is one of these.

More work remains to be done that analyzes other relevant cases using the methods presented here, in particular the scalar sector of maximal $D = 5$ gauged supergravity [5], i.e. the AdS side of the best studied case of the AdS/CFT correspondence, as well as $C\mathrm{SO}(p, q, r)$-gaugings of four-dimensional maximal supergravity, plus their dyonic variants [105], and other gaugings discussed in [102], as well as gauged supergravities in three and two dimensions.

JHEP08(2019)057

Also, while the present article is a major step forward towards a proven-exhaustive classification of the critical points of SO(8) supergravity, this challenging problem is still out of reach. It is quite likely that the rather peculiar form of the minimal polynomials of the cosmological constant that could be obtained for about three dozen critical points (including all solutions that can be described with up to six parameters) is a very major clue which the authors currently do not know how to utilize. Nevertheless, having an algebraic expression for the cosmological constant typically means that it is also feasible to find algebraic expressions for the entries of the 56-bein matrix, and hence quite a bit of the work required to uplift a critical point to a solution of the equations of motion of 11-dimensional supergravity can be automated. Not surprisingly, giving exact expressions for coordinates (which are not algebraic, but typically $\{\text{algebraic}\} \cdot \log\{\text{algebraic}\}$) is quite a bit harder (in fact, with the techniques employed here, this could only be achieved for a few coordinates), but might actually be unnecessary to answer most questions as long as the vielbein entries are known exactly and one can show that the vielbein indeed is an element of $E_{7(7)}$. Clearly, it would be very interesting to know in particular for all the supersymmetric vacua what 11-dimensional geometries these solutions correspond to — and also what the CFT renormalization group flows on M2-branes between the corresponding fixed points look line.

For all solutions presented here, numerical data on their location are available in the source package of the arXiv.org preprint of this article. In many cases, the authors have been able to numerically reduce the violation of the stationarity condition to much less than 10^{-1000}, and for each claimed critical point, it would be unreasonable to doubt its existence. Unfortunately, going from machine accuracy as obtained with TensorFlow to high accuracy is a somewhat messy process as the heuristics to guess a low-dimensional model as well as the attempt to use a multi-dimensional Newton solver do occasionally fail (for quite a few independent reasons) and require manual intervention. Due to limited time and a rather large number of cases to analyze, the quality of numerical data is rather uneven, typically providing 100+ good digits, but sometimes only providing as few as 16. With a bit of numerical experimenting, it typically is possible to obtain 1000+ good digits for any given solution within about two days of work.

Quite a few of the solutions feature (occasionally large) accidental degeneracies in the spectrum that should be understood and may point to exploitable symmetry properties. It may well be that the accidental $\mathrm{U}(1) \not\subset \mathrm{SO}(8)$ "background round S^7 diffeomorphism" symmetry of the $S1400000$ solution discussed in [113], is a first example of such an extra symmetry. Also, it is interesting to observe that, out of all the solutions newly discovered by employing numerical methods, there is a single one for which the residual symmetry group is *not* embedded in a way that has triality symmetry. Clearly, triality seems to play a rather important role, and so re-phrasing the problem in octonionic language might reveal additional structure.

In this work, we have only scratched the surface on cleverly engineering loss functions in order to extract additional information from the scalar potential. We expect that, with some ingenuity, much more is possible here, perhaps even allowing an efficient direct scan for BF-stable critical points.

JHEP08(2019)057

Given that even this long list will still have some gaps, what are the most promising approaches to fill them? One strategy still is to parametrize and study interesting submanifolds, such as the manifold of scalars that are invariant under the SO(3) gauge group of the new supersymmetric vacuum. Also, assuming that there indeed are perfect algebraic "Galois twin" solutions (i.e. not only the cosmological constant, but also all particle masses are related by root flipping), it may be possible to discover new solutions by obtaining algebraic expressions for some at present only numerically known solutions and then check whether root-flipping can produce new real solutions. A very promising further approach may be to study the extremal structure of the "dyonic" variants described in [105] and study the fate of critical points when changing the new ω-parameter. In [134], this method found the third SO(4) critical point $S1039230$ before it was independently rediscovered here, and it may well reveal further solutions.

Historically, the potential of SO(8) supergravity was first written down in [3]. About fourty years later, we now have what is likely the almost-complete list of critical points, and very likely the complete list of supersymmetric vacua of this theory. Given that some of the ideas that enabled this analysis were about as old as the original problem but largely unknown in the String Theory community, and were identified as useful and had to be brought together as part of one of the authors' personal journey, it seems likely that other technically hard problems in String Theory also would benefit from more exchange with other fields of research. We hope that by aiming to keep the introduction in this article accessible to readers with a technical background who are not experts in field theory, we might attract the attention of experts who can contribute missing puzzle pieces that we are not even aware of yet, perhaps even allowing a completeness proof of the list of solutions (perhaps after filling the last gaps).

Acknowledgments

It is a pleasure to thank our managers, first and foremost Jyrki Alakuijala, Rahul Sukthankar, and Jay Yagnik for support on this rather exotic side project, which turned into a rather unusual Machine Learning adventure. T.F. would like to thank Hermann Nicolai and the Albert Einstein Institute in Potsdam for hospitality and useful discussions during the final phase of this project, and also Krzysztof Pilch, Nikolay Bobev, Adolfo Guarino, and Michael Duff for feedback.

A E_7 conventions

Spin(8) triality provides some intuitive insights into the question why the exceptional Lie group E_7 exists. Let us start from the observation that one can, for example, understand the algebra rotations in nine spatial dimensions, $\mathfrak{so}(9)$, as an expansion of the group of rotations in eight spatial dimensions, $\mathfrak{so}(8)$, by eight extra elements r_{k8} that, when scaled and exponentiated, rotate each of the coordinate axes of eight-dimensional space against

JHEP08(2019)057

the ninth axis,

$$\begin{aligned} R_{k8}(\alpha)_{mn} &= \exp(\alpha r_{k8})_{mn} \\ &= (\delta_{km}\delta_{kn} + \delta_{8m}\delta_{8n})\cos\alpha + (\delta_{km}\delta_{8n} - \delta_{8m}\delta_{kn})\sin\alpha \end{aligned} . \tag{A.1}$$

and themselves transform under $\mathfrak{so}(8)$ as eight-dimensional vectors, while giving rise to other $\mathfrak{so}(8)$ rotations in their commutator, $[r_{08}, r_{18}] = -2r_{01}$ so that we get an overall structure of $[g,g] \sim g$, $[g,h] \sim h$, $[h,h] \sim g$. Likewise, one can obtain the 63-dimensional algebra $\mathfrak{sl}(8)$ from the 28-dimensional algebra of $\mathfrak{so}(8)$ by adding 35 generators that transform as symmetric traceless matrices under $\mathfrak{so}(8)$. One easily sees how for the 8-dimensional vector representation of $\mathfrak{sl}(8)$, the generators can be expressed in a basis of symmetric and antisymmetric traceless matrices, where the latter form the subalgebra of $\mathfrak{so}(8)$ and the former are the generators that extend this to $\mathfrak{sl}(8)$. One notes that one also gets a 63-dimensional real algebra (over complex matrices) that closes if one multiplied each of the extra 35 generators with i, so that commutator relations are of the form $[g,g] \sim g$, $[g,ih] \sim ih$, $[ih,ih] \sim -g$. The extra minus sign here also shows up in the signature of the quadratic invariant that can be formed from the generators, the Killing form, $G_{mn} := \mathrm{Tr}\,(g_m g_n)$. Applying this "Weyl unitarity trick" to $\mathfrak{sl}(8)$ produces the Lie algebra of the *compact* Lie group SU(8), i.e. $\mathfrak{su}(8)$.

Now, the $\mathfrak{so}(8)$ (or "$\mathfrak{spin}(8)$") algebra is special in that there is a non-abelian S_3 symmetry that acts on its irreducible representations. This symmetry permutes the roles of the three inequivalent eight-dimensional irreducible representations, the "vectors", "spinors", and "co-spinors". Let us consider extending $\mathfrak{so}(8)$ with an eight-dimensional vector representation v as before in order to get $\mathfrak{so}(9)$, but simultaneously with eight-dimensional spinors s and co-spinors c in such a way that we get commutator relations of the form $[g,g] \sim g$, $[g,v] \sim v$, $[v,v] \sim g$, $[v,s] \sim c$ plus the corresponding cyclically triality-rotated forms such as $[s,s] \sim g$, $[s,c] \sim v$, etc, by employing the invariant tensor $\gamma^i_{\alpha\dot{\beta}}$ of $so(8)$ that implements a non-degenerate generalized product of eight-dimensional representations. There is a unique way to work out commutator relations such that the Jacobi identity $[a,[b,c]] + \{\text{cyclic}\} = 0$ holds, and this gives the Lie algebra of the $28 + 3 \cdot 8 = 52$-dimensional exceptional group F_4. Performing a similar construction that extends $\mathfrak{so}(8)$ in a triality-symmetric way now with the symmetric traceless matrices over the vectors, spinors and co-spinors produces the $28 + 35_v + 35_s + 35_c = 133$-dimensional Lie algebra of $\mathfrak{e}_7$. It is possible to choose signs such that one of the 35-dimensional representations extends $\mathfrak{so}(8)$ to $\mathfrak{su}(8)$, while the other two each extend $\mathfrak{so}(8)$ to $\mathfrak{sl}(8)$. This then gives the noncompact real form $\mathfrak{e}_{7(7)}$ that shows up in four-dimensional $N = 8$ supergravity. So, in a sense, the $\mathfrak{e}_7$ algebra can be thought of as a generalized algebra of rotations with "siamese triplet" structure, three organisms co-joined at the $\mathfrak{so}(8)$ heart, but functioning as one whole body. In order to make this work self-contained, we spell out the conventions underlying our construction of $\mathfrak{e}_{7(7)}$ in detail. These match [108], apart from index counting always being 0-based the present article.

We start from the $\mathfrak{so}(8)$ "Pauli-matrices" $\gamma^i_{\alpha\dot{\beta}}$ in the conventions of Green, Schwarz, and Witten [135], but with all indices shifted down by 1, as it makes much more sense

JHEP08(2019)057

in a computational setting to consistently start index counting at zero. The nonzero entries of $\gamma^i_{\alpha\dot\beta}$ are all ±1, and we list them in compact form $i\alpha\dot\alpha_\pm$, so e.g. 357_- translates as $\gamma^{i=3}_{\alpha=5,\dot\beta=7} = -1$, etc.

$$\begin{array}{cccccccc}
007_+ & 016_- & 025_- & 034_+ & 043_- & 052_+ & 061_+ & 070_- \\
101_+ & 110_- & 123_- & 132_+ & 145_+ & 154_- & 167_- & 176_+ \\
204_+ & 215_- & 226_+ & 237_- & 240_- & 251_+ & 262_- & 273_+ \\
302_+ & 313_+ & 320_- & 331_- & 346_- & 357_- & 364_+ & 375_+ \\
403_+ & 412_- & 421_+ & 430_- & 447_+ & 456_- & 465_+ & 474_- \\
505_+ & 514_+ & 527_+ & 536_+ & 541_- & 550_- & 563_- & 572_- \\
606_+ & 617_+ & 624_- & 635_- & 642_+ & 653_+ & 660_- & 671_- \\
700_+ & 711_+ & 722_+ & 733_+ & 744_+ & 755_+ & 766_+ & 777_+
\end{array} \tag{A.2}$$

The Spin(8) traceless symmetric matrices over the vectors and spinors are "bosonic" objects, i.e. cannot discriminate between a 360-degrees rotation and the identity, so they must be expressible in terms of SO(8) representations alone. It so turns out that the $\mathbf{35_s}$ and $\mathbf{35_c}$ are equivalent to the self-dual, respectively anti self-dual four-forms of SO(8), and the corresponding Spin(8)-invariant dictionaries are the tensors

$$\gamma^{ijkl}_{\alpha\beta} = \gamma^m_{\alpha\dot\gamma}\gamma^n_{\gamma\dot\gamma}\gamma^p_{\gamma\dot\epsilon}\gamma^q_{\beta\dot\epsilon}\delta^{ijkl}_{mnpq} \tag{A.3}$$

$$\gamma^{ijkl}_{\dot\alpha\dot\beta} = \gamma^m_{\gamma\dot\alpha}\gamma^n_{\gamma\dot\gamma}\gamma^p_{\epsilon\dot\gamma}\gamma^q_{\epsilon\dot\beta}\delta^{ijkl}_{mnpq}. \tag{A.4}$$

We choose the basis for $e_{7(7)}$ such that the last 28 generators (elements #105 to #132) form the $so(8)$ subalgebra, elements #0 to #34 transform as symmetric traceless matrices over the spinors $\mathbf{35}_s$, elements #35 to #69 as symmetric traceless matrices over the cospinors $\mathbf{35}_c$, and elements #70 to #104 as symmetric traceless matrices over the vectors. Thus, the last 63 elements form the sub-algebra $\mathfrak{su}(8)$, and the first 70 elements the 70 noncompact directions of the coset manifold of supergravity scalars. An E_7 adjoint index $\mathcal{A}$ splits as (with the underline representing a single index that can be associated with a pair of $\mathfrak{so}(8)$-indices):

$$\mathcal{A} \to \underline{(\alpha\beta)} + \underline{(\dot\gamma\dot\delta)} + \underline{(ij)} + \underline{[kl]} \tag{A.5}$$

We furthermore choose the basis for $\mathfrak{so}(8)$ in such a way that element $105+n$, when acting on the vector representation of $\mathfrak{so}(8)$, would be represented as the rotation matrix $(r_{\underline{[jk]}})^m{}_p = \delta^j_m\delta^k_p - \delta^j_p\delta^k_m$, i.e. the rotation that takes the k-direction into the j-direction, with $0 \le j \le k \le 7$, where $n = j\cdot 8 + k - (j+1)(j+2)/2$. Hence, $\mathfrak{e}_7$ basis element #105 corresponds to the rotation $r_{\underline{[01]}}$, basis element #106 $= r_{\underline{[02]}}$, etc. (lexicographically ordered).

For the three 35-dimensional symmetric traceless irreducible $\mathfrak{so}(8)$ representations, we use the convention that the first 7 basis elements correspond to the diagonal matrices $\mathrm{diag}(1,-1,0,\ldots,0)$, $\mathrm{diag}(0,1,-1,0,\ldots,0)$, $\mathrm{diag}(0,\ldots,0,1,-1)$ (in that order), while element $7+n$ corresponds to the matrix $(S_{\underline{(jk)}})^m{}_p = \delta^j_m\delta^k_p + \delta^j_p\delta^k_m$, again with $0 \le j \le k \le 7$, and also $n = j\cdot 8 + k - (j+1)(j+2)/2$ — and with a likewise lexicographical order for the corresponding non-diagonal parts of $S_{\underline{(\alpha\beta)}}$ and $S_{\underline{(\dot\alpha\dot\beta)}}$. With these conventions, the e_7 symmetric bilinear form obtained from the fundamental representation, $g_{\mathcal{AB}} = T_{\mathcal{AC}}{}^D T_{\mathcal{BD}}{}^C$ is

JHEP08(2019)057

almost diagonal, with entries $+96$ for $B = C < 70$, entries -96 for $B = C \geq 70$, entries -48 for the non-orthogonal generators corresponding to the diagonal parts of the symmetric traceless matrices over the spinors and co-spinors (i.e. $g_{\mathcal{A}=0\,\mathcal{B}=1}$, $g_{\mathcal{A}=3\,\mathcal{B}=2}$ $g_{\mathcal{A}=35\,\mathcal{B}=36}$, etc.), and entries $+48$ for $G_{\mathcal{A}=70\,\mathcal{B}=71}$, etc. for the diagonal part of the $\mathbf{35}_v$ representation.

In order to define the 56-bein, we need explicit generators for the pseudoreal 56-dimensional fundamental representation of $\mathfrak{e}_7$. In the expressions below, the Einstein summation convention does not apply for "technical" auxiliary indices that are set in typewriter font and do not belong to irreducible representations).

Given $T^{(E7)}{}_{\mathcal{A}B}{}^{C}$, the $\mathfrak{e}_7$ generator matrices g used to define the scalar potential are $\left(g^{(\mathtt{n})}\right)^{C}{}_{B} = \left(T^{(E7)}{}_{\mathcal{A}=\mathtt{n}}\right)_{B}{}^{C}$ for $n = 0, \ldots, 69$. These 56×56 matrices $\left(g^{(\mathtt{n})}\right)$ look as follows (with $Z(n)$ given by (2.6)):

$$T^{(\mathrm{SU}(8))}{}_{Aj}{}^{k} = \begin{cases} +i \text{ for } A = j = k < 7 \\ -i \text{ for } A + 1 = j = k < 7 \\ +i \text{ for } 7 \leq A < 35,\ (\mathtt{m},\mathtt{n}) = Z(A-7), j = \mathtt{m}, k = \mathtt{n} \\ +i \text{ for } 7 \leq A < 35,\ (\mathtt{m},\mathtt{n}) = Z(A-7), j = \mathtt{n}, k = \mathtt{m} \\ +1 \text{ for } 35 \leq A,\ (\mathtt{m},\mathtt{n}) = Z(A-35), j = \mathtt{m}, k = \mathtt{n} \\ -1 \text{ for } 35 \leq A,\ (\mathtt{m},\mathtt{n}) = Z(A-35), j = \mathtt{n}, k = \mathtt{m} \end{cases}$$

$$S^{(\mathrm{SO}(8))\,abcd}_{\underline{(\alpha\beta)}} = \begin{cases} \gamma^{abcd}_{\alpha\beta}(\delta^{\alpha}_{\mathtt{m}}\delta^{\beta}_{\mathtt{n}} - \delta^{\alpha}_{\mathtt{n}}\delta^{\beta}_{\mathtt{n}}) \text{ for } \underline{(\alpha\beta)} = \mathtt{m} = \mathtt{n} - 1 < 7 \\ \gamma^{abcd}_{\alpha\beta}(\delta^{\alpha}_{\mathtt{m}}\delta^{\beta}_{\mathtt{n}} + \delta^{\alpha}_{\mathtt{n}}\delta^{\beta}_{\mathtt{m}}) \text{ for } \underline{(\alpha\beta)} \geq 7,\ (\mathtt{m},\mathtt{n}) = Z(\underline{(\alpha\beta)} - 7) \end{cases}$$

$$C^{(\mathrm{SO}(8))\,abcd}_{\underline{(\dot\alpha\dot\beta)}} = \begin{cases} \gamma^{abcd}_{\dot\alpha\dot\beta}(\delta^{\dot\alpha}_{\mathtt{m}}\delta^{\dot\beta}_{\mathtt{n}} - \delta^{\dot\alpha}_{\mathtt{n}}\delta^{\dot\beta}_{\mathtt{n}}) \text{ for } \underline{(\dot\alpha\dot\beta)} = \mathtt{m} = \mathtt{n} - 1 < 7 \\ \gamma^{abcd}_{\dot\alpha\dot\beta}(\delta^{\dot\alpha}_{\mathtt{m}}\delta^{\dot\beta}_{\mathtt{n}} + \delta^{\dot\alpha}_{\mathtt{n}}\delta^{\dot\beta}_{\mathtt{m}}) \text{ for } \underline{(\dot\alpha\dot\beta)} \geq 7,\ (\mathtt{m},\mathtt{n}) = Z(\underline{(\dot\alpha\dot\beta)} - 7) \end{cases}$$

$$T^{(E7)}{}_{\mathcal{A}B}{}^{C} = \begin{cases} \frac{1}{8} S^{(\mathrm{SO}(8))\,abcd}_{\underline{(\alpha\beta)}}(\delta^{a}_{\mathtt{m}}\delta^{b}_{\mathtt{n}} - \delta^{a}_{\mathtt{n}}\delta^{b}_{\mathtt{m}})(\delta^{c}_{\mathtt{p}}\delta^{d}_{\mathtt{q}} - \delta^{c}_{\mathtt{q}}\delta^{d}_{\mathtt{p}}) & \text{for} \\ \quad \mathcal{A} < 35,\ \underline{(\alpha\beta)} = \mathcal{A}, \\ \quad B \geq 28,\ (\mathtt{p},\mathtt{q}) = Z(B-28), \\ \quad C < 28,\ (\mathtt{m},\mathtt{n}) = Z(C) \\ \hline \frac{1}{8} S^{(\mathrm{SO}(8))\,cdab}_{\underline{(\alpha\beta)}}(\delta^{c}_{\mathtt{m}}\delta^{d}_{\mathtt{n}} - \delta^{c}_{\mathtt{n}}\delta^{d}_{\mathtt{m}})(\delta^{a}_{\mathtt{p}}\delta^{b}_{\mathtt{q}} - \delta^{a}_{\mathtt{q}}\delta^{b}_{\mathtt{p}}) & \text{for} \\ \quad \mathcal{A} < 35,\ \underline{(\alpha\beta)} = \mathcal{A}, \\ \quad B < 28,\ (\mathtt{p},\mathtt{q}) = Z(B), \\ \quad C \geq 28,\ (\mathtt{m},\mathtt{n}) = Z(C-28) \\ \hline \frac{i}{8} C^{(\mathrm{SO}(8))\,abcd}_{\underline{(\dot\alpha\dot\beta)}}(\delta^{a}_{\mathtt{m}}\delta^{b}_{\mathtt{n}} - \delta^{a}_{\mathtt{n}}\delta^{b}_{\mathtt{m}})(\delta^{c}_{\mathtt{p}}\delta^{d}_{\mathtt{q}} - \delta^{c}_{\mathtt{q}}\delta^{d}_{\mathtt{p}}) & \text{for} \\ \quad 35 \leq \mathcal{A} < 70,\ \underline{(\dot\alpha\dot\beta)} = \mathcal{A} - 35, \\ \quad B \geq 28,\ (\mathtt{p},\mathtt{q}) = Z(B-28), \\ \quad C < 28,\ (\mathtt{m},\mathtt{n}) = Z(C) \end{cases}$$

JHEP08(2019)057

$$
T^{(E7)}{}_{\mathcal{A}B}{}^{C} = \begin{cases}
-\frac{i}{8} C^{(\mathrm{SO}(8))\,cdab}_{\qquad\quad\underline{(\dot\alpha\dot\beta)}} (\delta^c_{\mathtt{m}}\delta^d_{\mathtt{n}} - \delta^c_{\mathtt{n}}\delta^d_{\mathtt{m}})(\delta^a_{\mathtt{p}}\delta^b_{\mathtt{q}} - \delta^a_{\mathtt{q}}\delta^b_{\mathtt{p}}) & \text{for} \\
\quad 35 \le \mathcal{A} < 70,\ \underline{(\dot\alpha\dot\beta)} = \mathcal{A} - 35, & \\
\quad B < 28,\ (\mathtt{p},\mathtt{q}) = Z(B), & \\
\quad C \ge 28,\ (\mathtt{m},\mathtt{n}) = Z(C-28) & \\
\hline
2\, T^{(\mathrm{SU}(8))}{}_{Aj}{}^{k} \left(\delta^{\mathtt{q}}_{\mathtt{n}}\delta^j_{\mathtt{m}}\delta^{\mathtt{p}}_k + \delta^{\mathtt{p}}_{\mathtt{m}}\delta^j_{\mathtt{n}}\delta^{\mathtt{q}}_k - \delta^{\mathtt{p}}_{\mathtt{n}}\delta^j_{\mathtt{m}}\delta^{\mathtt{q}}_k - \delta^{\mathtt{q}}_{\mathtt{m}}\delta^j_{\mathtt{n}}\delta^{\mathtt{p}}_k \right) & \text{for} \\
\quad \mathcal{A} \ge 70,\ A = \mathcal{A} - 70 & \\
\quad B < 28,\ (\mathtt{p},\mathtt{q}) = Z(B), & \\
\quad C < 28,\ (\mathtt{m},\mathtt{n}) = Z(C) & \\
\hline
2\, \left(T^{(\mathrm{SU}(8))}{}_{Aj}{}^{k}\right)^* \left(\delta^{\mathtt{q}}_{\mathtt{n}}\delta^j_{\mathtt{m}}\delta^{\mathtt{p}}_k + \delta^{\mathtt{p}}_{\mathtt{m}}\delta^j_{\mathtt{n}}\delta^{\mathtt{q}}_k - \delta^{\mathtt{p}}_{\mathtt{n}}\delta^j_{\mathtt{m}}\delta^{\mathtt{q}}_k - \delta^{\mathtt{q}}_{\mathtt{m}}\delta^j_{\mathtt{n}}\delta^{\mathtt{p}}_k \right) & \text{for} \\
\quad \mathcal{A} \ge 70,\ A = \mathcal{A} - 70 & \\
\quad B \ge 28,\ (\mathtt{p},\mathtt{q}) = Z(B-28), & \\
\quad C \ge 28,\ (\mathtt{m},\mathtt{n}) = Z(C-28) &
\end{cases}
\tag{A.6}
$$

B The octonions and the $\mathfrak{spin}(8)$ invariant $\gamma^i_{\alpha\dot\beta}$

This section provides a simple and fully self-contained instructive example for using TensorFlow to numerically solve tensorial algebraic constraints with very little mental effort. The basic techniques are essentially the same as the ones used for the main part of this work. Also, this section provides a detailed answer to the question how two common conventions for Octonions and $\mathfrak{spin}(8)$ gamma matrices are related.

The Lie group Spin(8) has three inequivalent irreducible eight-dimensional representations, the vectors $\mathbf{8}_v$, for which we use indices $i, j, k, \ldots$, the spinors $\mathbf{8}_s$, indexed with $\alpha, \beta, \ldots$, and the co-spinors $\mathbf{8}_c$, indexed with $\dot\alpha, \dot\beta, \ldots$. The $\mathfrak{spin}(8)$ invariant $\gamma^i_{\alpha\dot\beta}$ provides a unique way to map spinors and co-spinors to vectors, $\phi(\cdot,\cdot) : \mathbf{8}_s \times \mathbf{8}_c \to \mathbf{8}_v : (s, c) \to v$ such that (e.g.) for any non-zero element S of $\mathbf{8}_s$, the map $\mathbf{8}_c \to \mathbf{8}_v : c \mapsto \phi(S, c)$ is nondegenerate.

This means that we can use $\gamma^i_{\alpha\dot\beta}$ to define an invertible 8-dimensional product, that is, an 8-dimensional real division algebra. Now, using e.g. an explicit form of the $(8, 8, 8)$ tensor $\gamma^i_{\alpha\dot\beta}$ such as the one given in [135], eq. (5.B.3), which has no reason to know about the division algebra interpretation, some arbitrary choice has been made for the vector space bases of the $\mathbf{8}_v$, $\mathbf{8}_s$, $\mathbf{8}_c$ representations. Without loss of generality, we can identify the chosen basis of $\mathbf{8}_v$ with the basis of the "output" vector space of the octonionic product as it is defined in [136], with the octonionic imaginary units $e_1 \ldots e_7$ satisfying:

$$
\begin{gathered}
e_j e_j = -1, \quad e_j e_k = -e_k e_j \,(j \neq k), \quad e_1 \cdot e_2 = e_4 \\
e_i \cdot e_j = e_k \implies e_{i+1} \cdot e_{j+1} = e_{k+1} \text{ and } e_{2i} \cdot e_{2j} = e_{2k} \pmod 7.
\end{gathered}
\tag{B.1}
$$

Then, as we want to retain orthonormality (but not necessarily handedness), we can try to find one $O(8)$ element that changes the basis of $\mathbf{8}_s$, and another $O(8)$ element that

JHEP08(2019)057

changes the basis of $\mathbf{8}_c$ in order to precisely align the entries of the $(8,8,8)$-tensor $\gamma^i_{\alpha\dot{\beta}}$ with the $(8,8,8)$-tensor of the octonionic multiplication table. This objective provides $8^3 = 512$ constraints, while we only have $2 \cdot 28$ continuous parameters (two eight-dimensional rotations) to do the alignment. Hence, it is quite a nontrivial statement that the stated objective can indeed be achieved. (It so turns out that there are discrete choices for this problem that differ in the way how signs are distributed.) For this instructional example, we are extra lazy and do not even try to employ a proper parametrization of $O(8)$ elements (such as e.g. via a Cayley transform $M = (I - A)(I + A)^{-1}$). Rather, we design our optimization problem such that we in principle allow arbitrary elements of GL(8), but introduce a term in our objective function that punishes deviations from orthonormality.

The important point about the piece of code shown below is that, while this solves a numerical optimization problem in $2 \cdot 64 = 128$ parameters both with very good performance and accuracy, nowhere did the need arise for us to provide any code that computes gradients. This is all handled by the TensorFlow framework.

```
import numpy
import tensorflow as tf


def get_gamma_vsc():
  """Computes SO(8) gamma-matrices."""
  # Conventions match Green, Schwarz, Witten's.
  entries = (
      "007+ 016- 025- 034+ 043- 052+ 061+ 070- "
      "101+ 110- 123- 132+ 145+ 154- 167- 176+ "
      "204+ 215- 226+ 237- 240- 251+ 262- 273+ "
      "302+ 313+ 320- 331- 346- 357- 364+ 375+ "
      "403+ 412- 421+ 430- 447+ 456- 465+ 474- "
      "505+ 514+ 527+ 536+ 541- 550- 563- 572- "
      "606+ 617+ 624- 635- 642+ 653+ 660- 671- "
      "700+ 711+ 722+ 733+ 744+ 755+ 766+ 777+")
  ret = numpy.zeros([8, 8, 8])
  for ijkc in entries.split():
    ijk = tuple(map(int, ijkc[:-1]))
    ret[ijk] = +1 if ijkc[-1] == '+' else -1
  return ret


def get_octonion_mult_table():
  """Computes the octonionic multiplication table"""
  # Cf. diagram at: http://math.ucr.edu/home/baez/octonions/
  ret = numpy.zeros([8, 8, 8])
  fano_lines = "124 156 137 235 267 346 457"
  for n in range(1, 8):
      ret[0, n, n] = -1
      ret[n, n, 0] = ret[n, 0, n] = 1
  ret[0, 0, 0] = 1
  for cijk in fano_lines.split():
    ijk = map(int, cijk)
    for p, q, r in ((0, 1, 2), (1, 2, 0), (2, 0, 1)):
      # Note that we have to 'go against the direction of the arrows'
      # to make the correspondence work.
      ret[ijk[r], ijk[p], ijk[q]] = -1
      ret[ijk[r], ijk[q], ijk[p]] = +1
  return ret
```


JHEP08(2019)057

```
def find_transforms():
  with tf.Graph().as_default():
    # Ensure reproducibility by seeding random number generators.
    tf.set_random_seed(0)
    transforms = tf.get_variable('transforms', shape=(2, 8, 8),
                                 dtype=tf.float64,
                                 trainable=True,
                                 initializer=tf.random_normal_initializer())
    id8 = tf.constant(numpy.eye(8), dtype=tf.float64)
    gamma = tf.constant(get_gamma_vsc(), dtype=tf.float64)
    otable = tf.constant(get_octonion_mult_table(),
                         dtype=tf.float64)
    # Transform gamma matrices step-by-step, since tf.einsum() does not
    # do SQL-like query planning optimization.
    rotated_gamma = tf.einsum(
        'vAb,bB->vAB', tf.einsum('vab,aA->vAb', gamma, transforms[0]),
        transforms[1])
    delta_mult = rotated_gamma - otable
    delta_ortho_s = tf.einsum('ab,cb->ac',
                              transforms[0], transforms[0]) - id8
    delta_ortho_c = tf.einsum('ab,cb->ac',
                              transforms[1], transforms[1]) - id8
    # This 'loss' function punishes deviations of the rotated gamma matrices
    # from the octonionic multiplication table, and also deviations of the
    # spinor and cospinor transformation matrices from orthogonality.
    loss = (tf.nn.l2_loss(delta_mult) +
            tf.nn.l2_loss(delta_ortho_s) + tf.nn.l2_loss(delta_ortho_c))
    opt = tf.contrib.opt.ScipyOptimizerInterface(
        loss, options=dict(maxiter=1000))
    with tf.Session() as sess:
      sess.run(tf.global_variables_initializer())
      opt.minimize(session=sess)
      return sess.run([loss, transforms])


loss, transforms = find_transforms()
print('Loss: %.6g, Transforms:\n%r\n' % (
    loss, numpy.round(transforms, decimals=5)))

## Prints:
# Loss: 2.19694e-11, Transforms:
# array([[[ 0.5, -0. , -0. ,  0.5,  0.5,  0. ,  0.5,  0. ],
#         [-0.5,  0. ,  0. ,  0.5, -0.5,  0. ,  0.5,  0. ],
#         [-0. , -0.5,  0.5, -0. , -0. ,  0.5,  0. ,  0.5],
#         [ 0. ,  0.5, -0.5, -0. , -0. ,  0.5, -0. ,  0.5],
#         [-0.5,  0. , -0. , -0.5,  0.5,  0. ,  0.5,  0. ],
#         [ 0.5, -0. ,  0. , -0.5, -0.5, -0. ,  0.5, -0. ],
#         [ 0. , -0.5, -0.5,  0. ,  0. , -0.5, -0. ,  0.5],
#         [-0. ,  0.5,  0.5,  0. , -0. , -0.5, -0. ,  0.5]],
#
#        [[-0. ,  0.5,  0.5,  0. ,  0. , -0.5,  0. ,  0.5],
#         [ 0. ,  0.5,  0.5, -0. , -0. ,  0.5,  0. , -0.5],
#         [ 0.5, -0. ,  0. ,  0.5,  0.5,  0. , -0.5,  0. ],
#         [ 0.5,  0. , -0. , -0.5,  0.5,  0. ,  0.5, -0. ],
#         [-0. , -0.5,  0.5, -0. ,  0. , -0.5, -0. , -0.5],
#         [-0. , -0.5,  0.5, -0. ,  0. ,  0.5,  0. ,  0.5],
#         [ 0.5,  0. ,  0. ,  0.5, -0.5, -0. ,  0.5, -0. ],
#         [ 0.5, -0. , -0. , -0.5, -0.5, -0. , -0.5, -0. ]]])
```


JHEP08(2019)057

C TensorFlow code for watershed analysis

Finding saddles in the stationarity condition (2.8) as well as their principal axes can in principle be done symbolically, at the level of tensor equations (requiring substantial effort), or by manually performing the backpropagation code transformation, and then once again on the backpropagated code (requiring substantial effort). The code below illustrates how TensorFlow allows one to achieve this objective with minimal additional coding effort (27 lines of code) if one already has code to compute the stationarity condition violation. The code shown here sets up a TensorFlow session context (roughly: an association between a graph describing tensor arithmetic operations and hardware resources) and then executes a sequence of independent minimization-searches as specified, serializing results to the filesystem.

```
def do_watershed_descent(seeds_scales=[(0, 0.1)],
                         out='watershed_{}.pickle'):
  graph = tf.Graph()
  with graph.as_default():
    tf_scalar_evaluator = scalar_sector.get_tf_scalar_evaluator()
    t_input = tf.placeholder(tf.float64, shape=[70])
    t_v70 = tf.Variable(
        initial_value=numpy.zeros([70]), trainable=True, dtype=tf.float64)
    op_assign_input = tf.assign(t_v70, t_input)
    sinfo = tf_scalar_evaluator(tf.cast(t_v70, tf.complex128))
    t_potential = sinfo.potential
    t_stationarity = sinfo.stationarity
    t_grad_stationarity = tf.gradients(t_stationarity, t_v70)[0]
    t_loss = tf.tensordot(t_grad_stationarity, t_grad_stationarity, 1)
    t_stationarity_hessian = tf.hessians([t_stationarity], [t_v70])[0]
    optimizer = tf.contrib.opt.ScipyOptimizerInterface(t_loss)
    with tf.Session() as sess:
      sess.run([tf.global_variables_initializer()])
      for n, (seed, scale) in enumerate(seeds_scales):
        rng = numpy.random.RandomState(seed=seed)
        v70 = rng.normal(size=70, scale=scale)
        sess.run([op_assign_input], feed_dict={t_input: v70})
        optimizer.minimize(sess)
        n_info = sess.run([t_potential, t_stationarity, t_loss,
                           t_v70, t_stationarity_hessian])
        with open(out.format(n), 'w') as h:
          pickle.dump(n_info, h)
```


JHEP08(2019)057

D Overview over the solutions

In this table, we list all known critical points of SO(8) supergravity ordered by negative cosmological constant $-V/g^2$. The table's columns are:

N	Number of the solution
S	Tag of the solution, based on the truncated integer part of $-V/g^2 \cdot 10^5$.
$\mathcal{N}$	Number of unbroken supersymmetries. Empty for unstable vacua.
H	The residual gauge symmetry Lie group H.
T	Triality-invariance of the $H \subseteq \mathrm{SO}(8)$ embedding.
M	The dimension of the H-invariant submanifold of the scalar manifold.
P	Number of parameters of the solution.
D	Degree of the cosmological constant's minimal polynomial.
A	Numerical accuracy.
C	Citations, major articles that covered this solution.

The tag in the S-column uses the truncated, rather than rounded, value of the potential, in alignment with earlier articles that use this naming scheme — which has become necessary due to the large number of solutions without any unbroken gauge symmetry. The $\mathcal{N}$-column is empty for (BF-)unstable critical points, and shows the number of unbroken supersymmetries for stable critical points. Supersymmetric solutions are automatically stable. The H-columns lists residual gauge symmetry (ignoring extra discrete factors), where we make an effort here to be specific about the actual group. So, if a $\mathfrak{so}(3)$ subalgebra is embedded into $\mathfrak{so}(8)$ in such a way that some particle state transforms as a spinor, we call the gauge group "Spin(3)", otherwise "SO(3)". For $\mathrm{SO}(3) \times \mathrm{SO}(3) \equiv \mathrm{SO}(4)$, we use the name "SO(4)" if all particle states can be arranged into representations of SO(4) (which would not hold e.g. for an isolated $(\mathbf{3}, \mathbf{1})$ representation), and $\mathrm{SO}(3) \times \mathrm{SO}(3)$ otherwise. For U(1) factors, we indicate with a subscript the largest observed particle charge once the generator has been re-scaled to the minimal length that makes all charges integral. So, for example, the gauge group of **S0880733** that is associated with the Lie algebra $\mathfrak{so}(3)+\mathfrak{so}(3)$ is "SO(4)", while the gauge group of **S1075828** is "$\mathrm{Spin}(3) \times \mathrm{U}(1)_4$". The T-column indicates what subgroup of the "triality" outer automorphism group S_3 of $\mathfrak{so}(8)$ the embedding of H is invariant under. Here, "VSC" means full triality invariance, while "SC" means invariance under a $S \leftrightarrow C$ exchange, etc. The M, P, D-columns provide different rough measures of the complexity of the solution. The M-column shows the dimension of the H-invariant submanifold of the scalar manifold, if there is some residual gauge symmetry H. If one wanted to find exact expressions for the location and properties of a solution without resorting to inverse symbolic computation, one would want to coordinate-parametrize this manifold M. A critical point of the restricted potential on M is then guaranteed to also be a critical point on the whole scalar manifold [116]. The P-column lists the number of different numerical parameters used in this work to describe the solution. It is possible that, in some cases, one can use a SO(8) rotation to find an alternative form with even fewer parameters, so this value only gives an upper bound on how many coordinate-parameters are necessary. The D-column shows the degree of the smallest polynomial with integer

JHEP08(2019)057

coefficients that could be found which has a zero at $-V/g^2$ — if algebraic identification of the cosmological constant was successful given the number of available digits. An entry a indicates an a-th order polynomial, while an entry a^b indicates an a-th order polynomial in x^b. The A-column shows the decimal logarithm of the residual value of the stationarity condition $|Q_{ijkl}|^2$. So, an entry of 100 means that $|Q_{ijkl}|^2 < 10^{-100}$ for the numerical location data that have been made available alongside the preprint of this work. The p column shows the page number on which the solution can be found, and the C-column lists major articles in which the corresponding critical point is discussed. A "*" indicates a new discovery.

Given the sheer number of almost 200 critical points, and observing that detailed data on each of these typically fills more than half a page, the total amount of data is too large to be included in full in a journal publication. Hence, the authors decided to make both raw data and typeset summaries available as ancillary files to this article's preprint on arXiv.org. The URL for a typeset PDF document that describes the properties of critical point `Snnnnnnn` is:

`https://arxiv.org/src/1906.00207v4/anc/extrema/Snnnnnnn/physics.pdf`

while the URL for raw position data is:

`https://arxiv.org/src/1906.00207v4/anc/extrema/Snnnnnnn/location.py.txt`

In online versions of this article, the overview table should have working clickable links to the corresponding detailed physics summaries and high-accuracy scalar parameters.[15].

For each solution, the detailed table shows the location in the form of two symmetric traceless matrices, $M_{\alpha\beta}$ and $M_{\dot\alpha\dot\beta}$, which have been rotated to maximize (but in some cases perhaps not globally) the number of zero entries. The normalization of these matrices is such that the $\mathfrak{e}_7$ generator G in the fundamental **56**-representation that is parameterized by the matrices $M_{\alpha\beta}, M_{\dot\alpha\dot\beta}$ satisfies

$$\operatorname{tr} G \cdot G = 48 \cdot \Big(M_{\alpha\beta} M_{\alpha\beta} + M_{\dot\alpha\dot\beta} M_{\dot\alpha\dot\beta} \Big) \,. \tag{D.1}$$

In the literature, locations of critical points are typically given in ϕ_{ijkl} four-form language, e.g. (4.7) in [133] for the solution S1400000,

$$\phi_{ijkl} = \operatorname{artanh}(2/\sqrt{5}) \Big(\big(\delta^{1234}_{ijkl} + \delta^{5678}_{ijkl}\big) + i \, \big(\delta^{1235}_{ijkl} - \delta^{4678}_{ijkl}\big) \Big). \tag{D.2}$$

The dictionary to translate between ϕ_{ijkl} and $M_{\alpha\beta}$, $M_{\dot\alpha\dot\beta}$ reads:[16]

$$\sqrt{2} \cdot \phi_{ijkl} = M_{\alpha\beta} \gamma^{ijkl}_{\alpha\beta} + i \cdot M_{\dot\alpha\dot\beta} \gamma^{ijkl}_{\dot\alpha\dot\beta} \tag{D.3}$$

[15] An expanded version of this work with a consolidated table of detailed properties of critical points is available as a pre-publication preprint of this work on arXiv.org at https://arxiv.org/abs/1906.00207v4

[16] For $\mathfrak{spin}(8)$ with the obvious basis choices for representation spaces, there is no difference between upper and lower indices.

JHEP08(2019)057

The detailed location data on arXiv.org allow checking all claims made here about the existence of particular critical points to numerical machine precision with the code that has been made available alongside [137]. For some critical points, the numerical data provided are accurate to beyond 1000 digits, which should in some cases be sufficient to obtain algebraic expressions for the 56-bein matrix entries via inverse symbolic computation techniques, hence allowing automation(!) of the uplifting to 11 dimensions along the lines of the construction presented in [100, 113]. Unfortunately, for about 40 critical points, the authors were not able to obtain data that in accuracy go substantially beyond numerical hardware (i.e. IEEE-754 double float) precision.

For the scalars ϕ, the tables list masses $m^2/m_0^2[\phi] = m^2L^2$ relative to the AdS radius $L^2 = -3/V$ according to eq. (2.7) and (2.10). The "naive" gravitino masses $m^2/m_0^2[\psi]$ listed are the eigenvalues of $2\,g^2L^2A_{1\,ij}A_1{}^{ik}$, so unbroken supersymmetry shows as a mass-squared eigenvalue of $+1$. Likewise, fermion masses $m^2/m_0^2[\chi]$ are eigenvalues of $2\,g^2L^2\cdot\frac{1}{6}A_{3\,ijklmn}A_3{}^{ijk\,pqr}$.

The code that was written to automate gauge group representation assignment for the particle spectra only processes gauge groups whose Lie algebra is of the form $\mathfrak{so}(3)^a+\mathfrak{u}(1)^b$, which must hold for all new solutions, as all the solutions with symmetry at least $\mathfrak{su}(3)$ have been classified a long time ago [116], and there are no critical points with residual symmetry $\mathfrak{su}(3) \not\supseteq \mathfrak{s} \subseteq \mathfrak{so}(5)$. So, data on the spectrum of these points were added by hand, based on [118] and [30].

N	S	$\mathcal{N}$	H	T	M	P	D	A	C
1	*S0600000*	8	$SO(8)$	*VSC*	0	0	1	∞	[46][110]
2	*S0668740*		$SO(7)$	*VC*	1	1	1^4	2399	[116]
3	*S0698771*		$SO(7)$	*VS*	1	1	1^2	2400	[138]
4	*S0719157*	1	G_2	*VSC*	2	2	1^4	2399	[116]
5	*S0779422*	2	$SU(3)\times U(1)_6$	–	4	2	1^2	2400	[116]
6	*S0800000*		$SU(4)$	*VSC*	2	1	1	2400	[116]
7	*S0847213*		$Spin(3)\times U(1)_2\times U(1)_2$	–	8	3	2	1978	[125]
8	*S0869596*		$SO(3)\times U(1)_4$	*VS*	4	2	2^2	2398	[125]
9	*S0880733*		$SO(4)$	*VC*	6	2	2^2	2398	[125]
10	*S0983994*		$SO(3)\times U(1)_4$	*VC*	4	3	1^4	2399	[125]
11	*S0998708*		$U(1)_4$	*VSC*	14	9		2399	[108]
12	*S1006758*		$U(1)_2$	*VSC*	26	10		257	[125]
13	*S1039230*		$SO(4)$	*VS*	6	2	1^2	2398	[134]
14	*S1039624*		$U(1)_6$	*VC*	20	6	5^4	227	[125]
15	*S1043471*					10		2399	[108]
16	*S1046017*					6		219	[125]
17	*S1067475*		$U(1)_2\times U(1)_4$	*VC*	12	5	3^4	2398	[108]
18	*S1068971*		$U(1)_4\times U(1)_4$	*VS*	18	7	5^2	2399	[125]
19	*S1075828*		$Spin(3)\times U(1)_4$	*VS*	10	5	1^2	256	[125]
20	*S1165685*		$U(1)_4\times U(1)_4$	*VS*	18	1	2	2398	[108]
21	*S1176725*					4	4^4	2399	[125]
22	*S1195898*		$SO(3)$	*VSC*	2	3	3^4	2398	[125]
23	*S1200000*	1	$U(1)_1\times U(1)_1$	*SC*	14	2	1	2399	[108]
24	*S1212986*		$U(1)_4$	*VSC*	14	6		257	[125]
25	*S1271622*		$Spin(3)$	*VS*	20	3	3	2398	[125]
26	*S1301601*		$U(1)_4\times U(1)_4$	*VC*	18	4	3^4	2397	[125]
27	*S1360892*		$U(1)_6\times U(1)_2$	*SC*	18	5	3^2	238	[108]
28	*S1362365*		$U(1)_2$	*VSC*	26	10		241	[125]
29	*S1363782*					18		230	[125]
30	*S1366864*					9		24	[125]

JHEP08(2019)057

N	S	$\mathcal{N}$	H	T	M	P	D	A	C
31	$S1367611$					13		38	[108]
32	$S1379439$					10		2398	[125]
33	$S1384096$	1	$\mathrm{SO}(3)$	vsc	10	6	3^4	522	*
34	$S1384135$		$\mathrm{U}(1)_2$	vsc	26	11		227	[125]
35	$S1400000$	0	$\mathrm{SO}(4)$	sc	4	1	1	2397	[110]
36	$S1400056$					18		149	[125]
37	$S1402217$		$\mathrm{U}(1)_6$	vc	20	6	5^4	245	[125]
38	$S1424025$		$\mathrm{SO}(3)$	vsc	10	6		232	[125]
39	$S1441574$		$\mathrm{U}(1)_4$	vc	36	11		148	[125]
40	$S1442018$		$\mathrm{U}(1)_4$	vc	36	12		228	[125]
41	$S1443834$		$\mathrm{U}(1)_6$	vs	20	10		23	[125]
42	$S1464498$					7	4^2	2398	[125]
43	$S1465354$					19		24	[125]
44	$S1469693$		$\mathrm{U}(1)_2$	vsc	26	6	1^2	2398	[125]
45	$S1470986$					8		2397	*
46	$S1473607$					8	7^4	2398	*
47	$S1474271$					11		2398	*
48	$S1477609$					6		228	*
49	$S1497038$		$\mathrm{U}(1)_2$	vsc	26	12		17	[108]
50	$S1514242$					8		226	*
51	$S1571627$		$\mathrm{U}(1)_4$	vs	36	5	11^2	2397	*
52	$S1586253$					15		237	*
53	$S1588533$					12		38	*
54	$S1596185$					11		216	*
55	$S1600000$					3	1	2398	*
56	$S1603495$					10		2398	*
57	$S1609267$					16		2397	*
58	$S1611549$		$\mathrm{U}(1)_6$	vs	20	6		2398	*
59	$S1624005$					13		596	*
60	$S1624009$					10		204	*
61	$S1627388$					13		2398	*
62	$S1637792$					11		17	*
63	$S1637802$					7		257	*
64	$S1641445$					18		24	[108]
65	$S1650530$					16		146	*
66	$S1650772$					14		256	*
67	$S1652031$					18		200	*
68	$S1652212$					8		2397	*
69	$S1671973$		$\mathrm{U}(1)_4$	vs	36	7		2398	*
70	$S1691171$					12		216	*
71	$S1775878$					10		26	*
72	$S1775885$					16		204	*
73	$S1787646$					13		26	[108]
74	$S1800000$					4	1	258	*
75	$S1805269$					9		26	[108]
76	$S1810641$		$\mathrm{U}(1)_2$	vsc	26	8		2396	*
77	$S1880177$		$\mathrm{U}(1)_4$	vc	36	12		146	*
78	$S1889269$		$\mathrm{U}(1)_4$	vc	36	10	4^4	220	*
79	$S2006988$		$\mathrm{U}(1)_4$	vsc	14	9		2397	*
80	$S2040656$					26		21	*
81	$S2043647$					23		22	*
82	$S2043965$					19		21	*
83	$S2045402$		$\mathrm{U}(1)_4$	vsc	14	9		2397	*
84	$S2054312$					17		257	*
85	$S2054714$		$\mathrm{U}(1)_4$	vsc	14	6		2397	*

JHEP08(2019)057

N	S	$\mathcal{N}$	H	T	M	P	D	A	C
86	*S*2074862					13		837	*
87	*S*2095412					15		2397	*
88	*S*2099077		$U(1)_2 \times U(1)_4$	*vs*	12	6		256	*
89	*S*2099419					16		195	*
90	*S*2099421					12		22	*
91	*S*2101265					13		21	*
92	*S*2118474					16		243	*
93	*S*2121742					13		235	*
94	*S*2126597					9		74	[108]
95	*S*2135328					26		224	*
96	*S*2135978					26		19	*
97	*S*2140848					9		24	[108]
98	*S*2144361					27		18	*
99	*S*2144749					27		18	*
100	*S*2145935					18		149	*
101	*S*2147655					18		222	*
102	*S*2153574					17		2397	*
103	*S*2154972					12		2397	*
104	*S*2160231					18		226	*
105	*S*2171187					13		236	*
106	*S*2178669					11		257	*
107	*S*2183495					18		234	*
108	*S*2206714					10		2396	*
109	*S*2215486		$U(1)_4$	*vsc*	14	9		2397	*
110	*S*2240836					18		21	*
111	*S*2241557					11		2397	*
112	*S*2262631					13		2397	*
113	*S*2279257		$U(1)_1$	*sc*	30	12		17	*
114	*S*2279859		$U(1)_1$	*sc*	30	7		216	*
115	*S*2291945		$U(1)_6$	*vs*	20	6		241	*
116	*S*2308861					18		153	*
117	*S*2309630					12		2397	*
118	*S*2335620					18		244	*
119	*S*2348985					12		28	*
120	*S*2389014		$U(1)_4$	*vs*	36	15		24	*
121	*S*2389433		$U(1)_4$	*vs*	36	9		2396	*
122	*S*2395245					13		233	*
123	*S*2411233					13		2396	*
124	*S*2416856					6		193	*
125	*S*2416940					9		226	*
126	*S*2420275					17		836	*
127	*S*2423138					13		2396	*
128	*S*2435197		$U(1)_4$	*vsc*	14	10		2397	*
129	*S*2435907					11		207	*
130	*S*2440304					16		230	*
131	*S*2447241					18		226	*
132	*S*2457396		$U(1)_2$	*vsc*	26	10		225	*
133	*S*2476056					26		17	*
134	*S*2482062		$U(1)_2$	*vsc*	26	13		219	*
135	*S*2484339					26		18	*
136	*S*2488182					17		214	*
137	*S*2488241					26		203	*
138	*S*2497037					18		17	*
139	*S*2502436					18		19	*
140	*S*2503105		$SO(3)$	*vsc*	2	3	3^4	2397	*

JHEP08(2019)057

N	S	$\mathcal{N}$	H	T	M	P	D	A	C
141	*S*2503173					18		18	*
142	*S*2511744		$U(1)_4$	*vs*	36	9		2397	*
143	*S*2512364					10		17	*
144	*S*2514936					18		223	[108]
145	*S*2519962		$U(1)_2 \times U(1)_4$	*vs*	12	6		238	*
146	*S*2522870		$U(1)_4$	*vsc*	14	9		836	*
147	*S*2535465		$U(1)_4$	*vc*	36	15		17	*
148	*S*2535919					17		218	*
149	*S*2547536		$U(1)_2 \times U(1)_4$	*vs*	12	6		256	*
150	*S*2550397					16		223	*
151	*S*2551136					14		23	*
152	*S*2566527					17		20	*
153	*S*2584003					13		245	*
154	*S*2595657		$U(1)_4$	*vc*	36	12		2396	*
155	*S*2625940					13		245	*
156	*S*2644794					17		19	*
157	*S*2648510					27		20	*
158	*S*2650165		$U(1)_4$	*vsc*	14	6		2396	*
159	*S*2652206					18		225	*
160	*S*2652377					26		16	*
161	*S*2653753					17		235	*
162	*S*2655334		$U(1)_2$	*vsc*	26	14		76	*
163	*S*2655918		$U(1)_4$	*vsc*	14	9		227	*
164	*S*2666807					13		225	*
165	*S*2697505		$U(1)_4$	*vs*	36	11		218	*
166	*S*2702580					7	3^2	2396	*
167	*S*2705605					16		116	*
168	*S*2707528		$U(1)_4$	*vs*	36	9		255	*
169	*S*2751936		$U(1)_6$	*vs*	20	10		2396	*
170	*S*2803140					8		2396	*
171	*S*2849861					18		18	*
172	*S*2855585		$U(1)_4$	*vs*	36	11		2396	*
173	*S*2874576					26		226	*
174	*S*2952191		$U(1)_6$	*vs*	20	6		2396	*
175	*S*2980210		$U(1)_4$	*vs*	36	15		75	*
176	*S*3108383		$U(1)_4$	*vs*	36	11		2396	*
177	*S*3155402					18		230	*
178	*S*3166153					12		2396	*
179	*S*3254262					13		17	*
180	*S*3254576					15		220	*
181	*S*3254929					10		2396	*
182	*S*3305153					9		2395	*
183	*S*3498681					18		234	*
184	*S*3560884					13		241	*
185	*S*3640089		$U(1)_4$	*vs*	36	11		233	*
186	*S*3777270					16		226	*
187	*S*3908381					18		234	*
188	*S*4009882					19		236	*
189	*S*4045562					18		232	*
190	*S*4145121					18		21	*
191	*S*4168086					9	4^2	227	*
192	*S*4599899					12		235	*

JHEP08(2019)057

References

[1] S.W. Hawking, *Is the end in sight for theoretical physics?*, *Phys. Bull.* **32** (1981) 15.

[2] M. Abadi et al., *TensorFlow: a system for large-scale machine learning*, talk given at the 12th *USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, November 2–4, Savannah, U.S.A. (2016).

[3] B. de Wit and H. Nicolai, $N = 8$ *supergravity with Local* SO(8) × SU(8) *invariance*, *Phys. Lett.* **B 108** (1982) 285 [INSPIRE].

[4] B. de Wit and H. Nicolai, *Local* SO(8) × SU(8) *invariance in* $\mathcal{N} = 8$ *supergravity and its implication for superunification*, technical report CM-P00062104 (1981).

[5] M. Günaydin, L.J. Romans and N.P. Warner, *Gauged* $N = 8$ *supergravity in five-dimensions*, *Phys. Lett.* **154B** (1985) 268 [INSPIRE].

[6] https://colab.sandbox.google.com/github/google-research/google-research/blob/master/m_theory/dim4/so8_supergravity_extrema/colab/so8_supergravity.ipynb

[7] https://github.com/google-research/google-research/tree/master/m_theory

[8] M.J. Duff, *The theory formerly known as strings*, *Sci. Amer.* **278** (1998) 64 [math/9608117].

[9] E. Witten, *String theory dynamics in various dimensions*, *Nucl. Phys.* **B 443** (1995) 85 [hep-th/9503124] [INSPIRE].

[10] J. Wess and B. Zumino, *Supergauge transformations in four-dimensions*, *Nucl. Phys.* **B 70** (1974) 39 [INSPIRE].

[11] D.Z. Freedman, P. van Nieuwenhuizen and S. Ferrara, *Progress toward a theory of supergravity*, *Phys. Rev.* **D 13** (1976) 3214 [INSPIRE].

[12] S. Deser and B. Zumino, *Consistent supergravity*, *Phys. Lett.* **B 62** (1976) 335 [INSPIRE].

[13] Z. Bern et al., *Ultraviolet behavior of* $\mathcal{N} = 8$ *supergravity at four loops*, *Phys. Rev. Lett.* **103** (2009) 81301.

[14] Z. Bern, *Ultraviolet surprises in gravity*, talk given at *Bay Area Particle Theory Seminar (BAPTS)*, October 9, San Francisco, U.S.A (2015).

[15] S. Deser, J.H. Kay and K.S. Stelle, *Renormalizability properties of supergravity*, *Phys. Rev. Lett.* **38** (1977) 527 [arXiv:1506.03757] [INSPIRE].

[16] E. Witten, *What every physicist should know about string theory*, in *Foundations of mathematics and physics one century after Hilbert*, J. Kouneiher ed., Springer, Germany (2018).

[17] S. Weinberg, *The quantum theory of fields. Vol. 3: supersymmetry*, Cambridge University Press, Cambridge U.K. (2013).

[18] M. Tanabashi et al., *Review of particle physics*, *Phys. Rev.* **D 98** (2018) 030001.

[19] G. Bertone, D. Hooper and J. Silk, *Particle dark matter: evidence, candidates and constraints*, *Phys. Rept.* **405** (2005) 279.

[20] R. D. Peccei, *The strong CP problem and axions*, in *Axions*, M. Kuster et al. eds., Springer, Germany (2008).

[21] S.M. Carroll, *The cosmological constant*, *Liv. Rev. Rel.* **4** (2001) 1.

[22] E. Witten, *Geometric Langlands from six dimensions*, arXiv:0905.2720 [INSPIRE].

JHEP08(2019)057

[23] S.L. Adler, *Axial-vector vertex in spinor electrodynamics*, *Phys. Rev.* **177** (1969) 2426.

[24] A. Bilal, *Lectures on anomalies*, arXiv:0802.0634 [INSPIRE].

[25] H. Georgi and S.L. Glashow, *Unity of all elementary particle forces*, *Phys. Rev. Lett.* **32** (1974) 438 [INSPIRE].

[26] J.C. Pati and A. Salam, *Lepton number as the fourth "color"*, *Phys. Rev.* **D 10** (1974) 275.

[27] F. Englert, *Nobel lecture: the BEH mechanism and its scalar boson*, *Rev. Mod. Phys.* **86** (2014) 843.

[28] P.W. Higgs, *Nobel lecture: evading the Goldstone theorem*, *Rev. Mod. Phys.* **86** (2014) 851.

[29] H. Nicolai and N.P. Warner, *The* SU(3) × U(1) *invariant breaking of gauged* $N = 8$ *supergravity*, *Nucl. Phys.* **B 259** (1985) 412 [INSPIRE].

[30] K. A. Meissner and H. Nicolai, *Standard model fermions and* $\mathcal{N} = 8$ *supergravity*, *Phys. Rev.* **D 91** (2015) 65029.

[31] A. Kleinschmidt and H. Nicolai, *Standard model fermions and* $K(E_{10})$, *Phys. Lett.* **B 747** (2015) 251 [arXiv:1504.01586] [INSPIRE].

[32] N. Bobev, N. Halmagyi, K. Pilch and N.P. Warner, *Holographic,* $N = 1$ *supersymmetric RG flows on* $M2$ *branes*, *JHEP* **09** (2009) 043 [arXiv:0901.2736] [INSPIRE].

[33] T. Kaluza, *On the unification problem in physics*, *Int. J. Mod. Phys.* **D 27** (2018) 1870001.

[34] O. Klein, *Quantentheorie und fünfdimensionale Relativitätstheorie*, *Z. Phys.* **37** (1926) 895.

[35] E. Cremmer, B. Julia and J. Scherk, *Supergravity theory in eleven-dimensions*, *Phys. Lett.* **B 76** (1978) 409 [INSPIRE].

[36] E. Cremmer and B. Julia, *The* $N = 8$ *supergravity theory. 1. The lagrangian*, *Phys. Lett.* **B 80** (1978) 48 [INSPIRE].

[37] E. Cremmer and B. Julia, *The* SO(8) *supergravity*, *Nucl. Phys.* **B 159** (1979) 141 [INSPIRE].

[38] G. Parisi and N. Sourlas, *Supersymmetric field theories and stochastic differential equations*, *Nucl. Phys.* **B 206** (1982) 321.

[39] S. Ferrara, J. Scherk and B. Zumino, *Algebraic properties of extended supergravity theories*, *Nucl. Phys.* **B 121** (1977) 393 [INSPIRE].

[40] E.S. Fradkin and M.A. Vasiliev, *Minimal set of auxiliary fields in* SO(2) *extended supergravity*, *Phys. Lett.* **85B** (1979) 47 [INSPIRE].

[41] P. Fayet, *Fermi-Bose hypersymmetry*, *Nucl. Phys.* **B 113** (1976) 135 [INSPIRE].

[42] J. Scherk, *Antigravity: a crazy idea?*, *Phys. Lett.* **88B** (1979) 265 [INSPIRE].

[43] E. Kopczyński, D. Celińska and M. Čtrnáct, *HyperRogue: playing with Hyperbolic Geometry*, in the proceedings of the *Bridges Conference*, July 27–31, Ontario, Canada (2017).

[44] J. Milnor, *On manifolds homeomorphic to the 7-sphere*, *Ann. Math.* **64** (1956) 399.

[45] E.V. Brieskorn, *Examples of singular normal complex spaces which are topological manifolds*, *Proc. Natl. Acad. Sci. U.S.A.* **55** (1966) 1395.

[46] B. de Wit and H. Nicolai, $N = 8$ *supergravity*, *Nucl. Phys.* **B 208** (1982) 323 [INSPIRE].

JHEP08(2019)057

[47] E. Witten, *Fermion quantum numbers in Kaluza-Klein theory*, in the proceedings of *Quantum field theory and the fundamental problems of physics*, June 1–3, Shelter Island, U.S.A. (1983).

[48] J. Maldacena, *The large* N *limit of superconformal field theories and supergravity*, *Int. J. Theor. Phys.* **38** (1999) 1113.

[49] S.S. Gubser, I.R. Klebanov and A.M. Polyakov, *Gauge theory correlators from noncritical string theory*, *Phys. Lett.* **B 428** (1998) 105 [hep-th/9802109] [INSPIRE].

[50] E. Witten, *Anti de Sitter space and holography*, *Adv. Theor. Math. Phys.* **2** (1998) 253.

[51] P. Kovtun, D.T. Son and A.O. Starinets, *Viscosity in strongly interacting quantum field theories from black hole physics*, *Phys. Rev. Lett.* **94** (2005) 111601 [hep-th/0405231] [INSPIRE].

[52] S.S. Gubser, *Breaking an Abelian gauge symmetry near a black hole horizon*, *Phys. Rev.* **D 78** (2008) 065034.

[53] S.A. Hartnoll, C.P. Herzog and G.T. Horowitz, *Holographic superconductors*, *JHEP* **12** (2008) 015.

[54] S.A. Hartnoll, C.P. Herzog and G.T. Horowitz, *Building a Holographic Superconductor*, *Phys. Rev. Lett.* **101** (2008) 031601 [arXiv:0803.3295] [INSPIRE].

[55] J. Ehlers, *Konstruktionen und Charakterisierung von Losungen der Einsteinschen Gravitationsfeldgleichungen*, Ph.D. thesis, Hamburg University, Hamburg, Germany (1957).

[56] R. Geroch, *A method for generating solutions of Einstein's equations*, *J. Math. Phys.* **12** (1971) 918.

[57] J. Berkeley and D.S. Berman, *The Navier-Stokes equation and solution generating symmetries from holography*, *JHEP* **04** (2013) 092 [arXiv:1211.1983] [INSPIRE].

[58] S. Bhattacharyya, S. Minwalla, V. E. Hubeny and M. Rangamani, *Nonlinear fluid dynamics from gravity*, *JHEP* **02** (2008) 045.

[59] O. Aharony, O. Bergman, D.L. Jafferis and J. Maldacena, $\mathcal{N} = 6$ *superconformal Chern-Simons-matter theories, M2-branes and their gravity duals*, *JHEP* **10** (2008) 091.

[60] N. Bobev, V.S. Min and K. Pilch, *Mass-deformed ABJM and black holes in* AdS_4, *JHEP* **03** (2018) 050 [arXiv:1801.03135] [INSPIRE].

[61] A.M. Turing, *Computing machinery and intelligence*, *Mind* **49** (1950) 433.

[62] A.L. Samuel, *Some studies in machine learning using the game of checkers*, *IBM J. Res. Dev.* **3** (1959) 210.

[63] C. Olah, A. Mordvintsev and L. Schubert, *Feature visualization*, *Distill* **2** (2017) e7.

[64] D. Genzel and A. Popat, *Paper to digital in* 200+ *languages*, (2015).

[65] A. Krizhevsky, I. Sutskever and G.E. Hinton, *ImageNet classification with deep convolutional neural networks*, in *Advances in neural information processing systems 25*, F. Pereira et al. eds., Curran Associates Inc., U.S.A. (2012).

[66] C. Szegedy et al., *Going deeper with convolutions*, in the proceedings of the 28th *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 7–12, Boston, U.S.A. (2014) [arXiv:1409.4842].

JHEP08(2019)057

[67] P. Sharma, N. Ding, S. Goodman and R. Soricut, *Conceptual captions: a cleaned, hypernymed, image alt-text dataset for automatic image captioning*, in the proceedings of the 56th *Annual Meeting of the Association for Computational Linguistics*, July 15–20, Melbourne, Australia (2018).

[68] A. Vaswani et al., *Attention is all you need*, in *Advances in neural information processing systems 30*, I. Guyon et al. eds., Curran Associates Inc., U.S.A. (2017).

[69] D. Silver et al., *Mastering the game of Go with deep neural networks and tree search*, *Nature* **529** (2016) 484.

[70] O. Vinyals et al., *AlphaStar: mastering the real-time strategy game StarCraft II*, https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii (2019).

[71] T. Karras, S. Laine and T. Aila, *A style-based generator architecture for generative adversarial networks*, arXiv:1812.04948.

[72] F. Rosenblatt, *The perceptron: a probabilistic model for information storage and organization in the brain*, *Psychol. Rev.* **65** (1958) 386.

[73] D.E. Rumelhart et al., *A general framework for parallel distributed processing*, in *Parallel distributed processing: Explorations in the microstructure of cognition*, D.E. Rumelhart and J.L. McClelland eds., MIT Press, U.S.A. (1986).

[74] S. Hochreiter and J. Schmidhuber, *Long short-term memory*, *Neural Comput.* **9** (1997) 1735.

[75] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, *Gradient-based learning applied to document recognition*, *Proc. IEEE* **86** (1998) 2278.

[76] E.J. Hartman, J.D. Keeler and J.M. Kowalski, *Layered neural networks with gaussian hidden units as universal approximations*, *Neural Comput.* **2** (1990) 210.

[77] G.E. Hinton, S. Osindero and Y.-W. Teh, *A fast learning algorithm for deep belief nets*, *Neural Comput.* **18** (2006) 1527 [arXiv:1111.6189].

[78] C. Cortes and V. Vapnik, *Support-vector networks*, *Machine Learn.* **20** (1995) 273.

[79] X. Glorot, A. Bordes and Y. Bengio, *Deep sparse rectifier neural networks*, in the proceedings of the 14th *International Conference on Artificial Intelligence and Statistics*, April 11–13, Ft. Lauderdale, U.S.A. (2011).

[80] T. Kohonen, *Self-organized formation of topologically correct feature maps*, *Biol. Cybernet.* **43** (1982) 59.

[81] P. Covington, J. Adams and E. Sargin, *Deep neural networks for YouTube recommendations*, in the proceedings of the 10th *ACM Conference on Recommender Systems (RecSys'16)*, September 15–19, Boston, U.S.A. (2016).

[82] V. Mnih et al., *Human-level control through deep reinforcement learning*, *Nature* **518** (2015) 529.

[83] R. Dunne and N. Campbell, *On the pairing of the Softmax activation and cross-entropy penalty functions and the derivation of the Softmax activation function*, in the proceedings of the 8th *Australian Conference on Neural Networks (ACNN97)*, Australia (1997).

[84] B. Speelpenning, *Compiling fast partial derivatives of functions given by algorithms*, Ph.D. thesis, University of Illinois Urbana-Champaign, Champaign, U.S.A. (1980).

JHEP08(2019)057

[85] D.E. Rumelhart, G.E. Hinton and R.J. Williams, *Learning internal representations by error propagation*, technical report, US Dept of the Navy, Cambridge, U.S.A. (1985).

[86] R. Bellman, *Dynamic programming and a new formalism in the calculus of variations*, *Proc. Natl. Acad. Sci.* **40** (1954) 231.

[87] *R6RS-AD*, https://github.com/qobi/R6RS-AD.

[88] R. Kondor et al., *Covariant compositional networks for learning graphs*, [arXiv:1801.02144].

[89] I. Bars, *Supersymmetry, p-brane duality and hidden space-time dimensions*, *Phys. Rev.* **D 54** (1996) 5203 [hep-th/9604139] [INSPIRE].

[90] E. Bergshoeff, E. Sezgin and P.K. Townsend, *Supermembranes and Eleven-Dimensional Supergravity*, *Phys. Lett.* **B 189** (1987) 75 [INSPIRE].

[91] E. Cremmer, B. Julia and J. Scherk, *Supergravity theory in* 11 *dimensions*, in *Supergravities in Diverse Dimensions*, A. Salam and E. Sezgin eds., World Scientific Publishing Company, Singapore (1989).

[92] M.J. Duff, B.E.W. Nilsson and C.N. Pope, *Kaluza-Klein Supergravity*, *Phys. Rept.* **130** (1986) 1 [INSPIRE].

[93] P.G.O. Freund and M.A. Rubin, *Dynamics of dimensional reduction*, *Phys. Lett.* **B 97** (1980) 233 [INSPIRE].

[94] E. Cremmer, B. Julia, H. Lü and C.N. Pope, *Dualization of dualities. 1.*, *Nucl. Phys.* **B 523** (1998) 73 [hep-th/9710119] [INSPIRE].

[95] E. Cremmer, B. Julia, H. Lü and C.N. Pope, *Dualization of dualities. 2. Twisted self-duality of doubled fields and superdualities*, *Nucl. Phys.* **B 535** (1998) 242 [hep-th/9806106] [INSPIRE].

[96] M.J. Duff, *Ultraviolet divergences in extended supergravity*, talk given at the *First School on Supergravity*, April 22–May 6, Trieste, Italy (1981), arXiv:1201.0386 [INSPIRE].

[97] B. Biran, F. Englert, B. de Wit and H. Nicolai, *Gauged* $N = 8$ *supergravity and its breaking from spontaneous compactification*, *Phys. Lett.* **B 124** (1983) 45.

[98] M.J. Duff and C.N. Pope, *Kaluza-Klein supergravity and the seven sphere*, in the proceedings of the *September School on Supergravity and Supersymmetry*, September 6–18, Trieste, Italy (1982).

[99] B. de Wit and H. Nicolai, *The consistency of the* S^7 *truncation in* $D = 11$ *supergravity*, *Nucl. Phys.* **B 281** (1987) 211 [INSPIRE].

[100] H. Nicolai and K. Pilch, *Consistent truncation of* $d = 11$ *supergravity on* $AdS_4 \times S^7$, *JHEP* **03** (2012) 099 [arXiv:1112.6131] [INSPIRE].

[101] M.J. Duff, B.E.W. Nilsson and C.N. Pope, *Compactification of* $d = 11$ *Supergravity on* $K(3) \times \mathrm{U}(3)$, *Phys. Lett.* **B 129** (1983) 39 [INSPIRE].

[102] B. de Wit, H. Samtleben and M. Trigiante, *The maximal* $\mathcal{D} = 4$ *supergravities*, *JHEP* **06** (2007) 049.

[103] C.M. Hull and N.P. Warner, *The structure of the gauged* $N = 8$ *supergravity theories*, *Nucl. Phys.* **B 253** (1985) 650 [INSPIRE].

JHEP08(2019)057

[104] C.M. Hull, *Non-compact gaugings of $\mathcal{N} = 8$ supergravity*, in *Supergravities in Diverse Dimensions*, A. Salam and E. Sezgin eds., World Scientific Publishing Company, Singapore (1989).

[105] G. Dall'Agata, G. Inverso and M. Trigiante, *Evidence for a family of* SO(8) *gauged supergravity theories*, *Phys. Rev. Lett.* **109** (2012) 201301 [arXiv:1209.0760] [INSPIRE].

[106] B. de Wit, H. Samtleben and M. Trigiante, *On Lagrangians and gaugings of maximal supergravities*, *Nucl. Phys.* **B 655** (2003) 93 [hep-th/0212239] [INSPIRE].

[107] B. de Wit, *Supergravity*, in the proceedings of *Unity from duality. Gravity, gauge theory and strings. NATO Advanced Study Institute, Euro Summer School, 76th session*, July 30–August 31, Les Houches, France (2001), hep-th/0212245 [INSPIRE].

[108] T. Fischbacher, *Fourteen new stationary points in the scalar potential of* SO(8)*-gauged $N = 8$, $D = 4$ supergravity*, *JHEP* **09** (2010) 068 [arXiv:0912.1636] [INSPIRE].

[109] G. Y. Rainich, *Electrodynamics in the general relativity theory*, *Proc. Natl. Acad. Sci.* **10** (1924) 124.

[110] N.P. Warner, *Some properties of the scalar potential in gauged supergravity theories*, *Nucl. Phys.* **B 231** (1984) 250 [INSPIRE].

[111] P. Breitenlohner and D.Z. Freedman, *Stability in gauged extended supergravity*, *Annals Phys.* **144** (1982) 249 [INSPIRE].

[112] T. Fischbacher, K. Pilch and N.P. Warner, *New supersymmetric and stable, non-supersymmetric phases in supergravity and holographic field theory*, arXiv:1010.4910 [INSPIRE].

[113] H. Godazga, *An* SO(3) × SO(3) *invariant solution of $D = 11$ supergravity*, *JHEP* **01** (2015) 056 [arXiv:1410.5090] [INSPIRE].

[114] A. Borghese, R. Linares and D. Roest, *Minimal Stability in Maximal Supergravity*, *JHEP* **07** (2012) 034 [arXiv:1112.3939] [INSPIRE].

[115] T. Fischbacher, *The many vacua of gauged extended supergravities*, *Gen. Rel. Grav.* **41** (2009) 315 [arXiv:0811.1915] [INSPIRE].

[116] N.P. Warner, *Some new extrema of the scalar potential of gauged $N = 8$ supergravity*, *Phys. Lett.* **128B** (1983) 169 [INSPIRE].

[117] B. de Wit and H. Nicolai, *A new* SO(7) *invariant solution of $d = 11$ supergravity*, *Phys. Lett.* **148B** (1984) 60 [INSPIRE].

[118] N. Bobev, N. Halmagyi, K. Pilch and N.P. Warner, *Supergravity instabilities of non-supersymmetric quantum critical points*, *Class. Quant. Grav.* **27** (2010) 235013.

[119] B. de Wit and H. Nicolai, *The parallelizing S_7 torsion in gauged $N = 8$ supergravity*, *Nucl. Phys.* **B 231** (1984) 506 [INSPIRE].

[120] *Google colaboratory*, https://colab.sandbox.google.com.

[121] N.P. Jouppi et al., *In-datacenter performance analysis of a tensor processing unit*, in the proceedings of the 44th *Annual International Symposium on Computer Architecture (ISCA'17)*, June 24–28, Toronto, Canada (2017), arXiv:1704.04760.

[122] J. Nocedal and S. Wright, *Numerical optimization*, 2nd edition, Springer Series in Operations Research and Financial Engineering, Springer, Germany (2006).

JHEP08(2019)057

[123] M. Morse, *The calculus of variations in the large*, *Monat. Math. Phys.* **47** (1939) A10.

[124] N. Bobev, T. Fischbacher and K. Pilch, *A new $\mathcal{N} = 1$ AdS_4 vacuum of maximal supergravity*, work in progress.

[125] T. Fischbacher, *The encyclopedic reference of critical points for* SO(8)*-gauged $N = 8$ supergravity. Part 1: cosmological constants in the range $-\Lambda/g^2 \in [6 : 14.7)$*, arXiv:1109.1424 [INSPIRE].

[126] S. Dittmaier, *Precision standard model physics*, talv given at *LoopFest V*, June 19–21, SLAC, Stanford, U.S.A. (2006).

[127] G.P. Collins, *The Large Hadron Collider: the discovery machine*, *Sci. Amer.* (2008) 39.

[128] F. Johansson et al., *mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 0.18)* (2013).

[129] D. Maclaurin, *Modeling, inference and optimization with composable differentiable procedures*, Ph.D. thesis, Harvard University, Cambridge, U.S.A. (2016).

[130] P.B. Davenport, *Rotations about nonorthogonal axes*, *AIAA J.* **11** (1973) 853.

[131] J. Wittenburg and L. Lilov, *Decomposition of a finite rotation into three rotations about given axes*, *Mult. Syst. Dyn.* **9** (2003) 353.

[132] D.H. Bailey and J.M. Borwein, *PSLQ: an algorithm to discover integer relations*, (2009).

[133] B. de Wit and H. Nicolai, *Properties of $\mathcal{N} = 8$ supergravity*, in the proceedings of the 19$^{\text{th}}$ *Winter School and Workshop on Theoretical Physics: Supersymmetry and Supergravity*, February 14–26, Karpacz, Poland (1983).

[134] A. Borghese, A. Guarino and D. Roest, *Triality, periodicity and stability of* SO(8) *gauged supergravity*, *JHEP* **05** (2013) 107 [arXiv:1302.6057] [INSPIRE].

[135] M.B. Green, J.H. Schwarz and E. Witten, *Superstring theory*, Cambridge Monographs on Mathematical Physics volume 150, Camrbidge University PRess, Cambridge U.K. (2012).

[136] J.C. Baez, *The octonions*, *Bull. Amer. Math. Soc.* **39** (2001) 145 [math/0105155].

[137] T. Fischbacher, *Numerical tools to validate stationary points of* SO(8)*-gauged $N = 8$ $D = 4$ supergravity*, *Comput. Phys. Commun.* **183** (2012) 780 [arXiv:1007.0600] [INSPIRE].

[138] F. Englert, *Spontaneous compactification of eleven-dimensional supergravity*, *Phys. Lett.* **B 119** (1982) 339 [INSPIRE].

JHEP08(2019)057