Energy Minimizing Flows on Equivariant Space of Connections

L. J. Pereira

Overview

In gauge theory, the moduli space of connections results from quotienting the space of principal connections of a fiber or vector bundle by the structure group, generating a gauge equivariant parameter space. A moduli space of connections can be used to cover the activity of a collection of hierarchical or deep neural networks, represented as trajectories on vector bundles. On this upper covering space, a top-down energy-based attention mechanism, referred to as the moduli attention, can be trained from sampled activity of the bottom-up trajectories of underlying networks through a composition of energies. Dynamics of attention are computed by minimizing the energy functional of the connection space. In particular, the Yang-Mills moduli space, a subset of the total connection space, can be constructed to be a smooth, compact, and oriented manifold in 4 dimensions with critical points known as Yang-Mills connections (or instantons). These connections minimize curvature between bundles and, from an information geometric perspective, they minimize relative entropy or KL divergence between gauge equivariant manifolds. Allowing these connections to function as sources of variational noise, we train a neural network to minimize total energy of the moduli space. This process can be understood as finding stable solutions of an intrinsic flow on the curvature metric, namely a Ricci Yang-Mills flow (RYM flow) that produces solutions which are real non-volume preserving objects around invariant points, akin to Ricci solitons. Stable solutions represent islands of agreement formed through consensus within a dynamic part-whole hierarchy.

1 Geometry of Latent Information

1.1 Manifolds, Vector Bundles, Connections

A fiber bundle serves as a useful mathematical object to analyse both the recursive construction of artificial and biological neurons and neural networks, as well as providing appropriate descriptions of the geometry of latent information, which uses manifold representations for information processing and statistical learning. A fiber bundle formalizes the notion of one topological space (called a fiber) being parameterized by another topological space (called a base). The bundle is equipped with a group action on the fiber that represents different ways the fibers can be viewed as equivalent. Bundles also have a property, known as local trivialization, allowing neighborhoods of the bundle to be computed as simple, oriented product spaces, despite the global space being unoriented or twisted.

A family of fibers associated to a base can be described by defining a standard (or template) fiber from which all other fibers are isomporphic to. This is formalized by defining a diffeomorphic or homotopic projection mapping, that connects positional data from the entirety of the space of fibers to a base, and implicitly from one fiber to another. When the template fiber is a vector space, the bundle is called a vector bundle. Similarly, a standard connection between fibers, known as a principal Ehresmann connection, will always exist and can be understood as a covariant directional derivative on the tangent spaces of the manifolds. Intuitively understood, Lie groups have a special recursive nature as a result of the group itself being a differentiable manifold of a continuous symmetry. A useful phenomenon occurs when equipping a bundle with a Lie group action; the bundle structure can be used to represent both the original vector bundle as well as a higher-level collection of mappings of their tangent spaces, in what's known as a bundle of connections.

1.2 A Priori Structure Groups

A biological first principle of covariance arises naturally from analysis of neuronal activity, which favours functional localization and Hebbian learning. Moreover, cognitive networks in the brain flow in connectome-specific diffusive waves along gyrification paths, which are theorized to be caused by differential tangential growth. Recall, covariance is a measure of the joint variability of a pair random variables and is increasingly positive when the pair show similar behavior and is negative when dissimilar. Moving from discrete random variables (i.e. synapses) to continuous fields and manifolds (i.e. EM fields), the covariant derivative between fibers of a bundle naturally arises in the principal Ehressmann connection. It proves necessary to impose a generalized a priori covariance principle to achieve holonomy of dynamics on a bundle to maintain integrity of information during parallel transportion in bilateral and hierarchical directions. Yet for a standard learning model, covariance of functionality is an a posteriori feature since the joint variability is unknown until individual modules are fully trained. Covariance of fibers can be achieved by imposing the structure group to be Lie groups, but can also be achieved by imposing restrictions on the projection map of Riemannian manifolds without explicitly defining the structure group beforehand (Gao 2021).

1.3 Moduli Spaces and Instantons

With covariance established on connections, it becomes possible to perform inference using higher levels of abstraction on gauge fields. This is done by constructing a gauge equivariant bundle of connections known as a moduli space of connections, which can be further reduced to a finite dimensional manifold known as a Yang-Mills moduli space. This reduced space has local and global minima being connections with minimized energy known as Yang-Mills connections or instantons which serve as a natural choice of connection on principal and vector bundles since they minimize their curvature. From an information geometric perspective, this can be thought of as minimizing relative entropy or KL divergence between sampled trajectories of gauge equivariant manifolds. The infinitesimal form of the the KL divergence is comparable to the Fisher information metric. The gauge field strength is the curvature F_A of the connection A, and the energy of the gauge field is given by the Yang–Mills action functional YM. With the aim of having zero or vanishing curvature, we vary parameters in search of a connection with curvature as small as possible. The Yang–Mills action functional corresponds to the L^2-norm of the curvature, and its Euler–Lagrange equations describe the critical points of this functional, either the absolute local minima.

$$\mathrm{YM}(A) = \int_X \|F_A\|^2 \, d\mathrm{vol}_g. \tag{1}$$

2 Dynamics of Latent Information

2.1 Variational Methods in Energy-Based Models

As underlying neural networks perform inference, their latent trajectories pass through layers of a neural network in a bottom-up manner. At the same time, a top-down variational noise is produced around the instanton solutions that best minimize energy of the total activity on the moduli space of connections. Using an energy-based model (EBM) we sample the joint distribution as a sum of each latent trajectory, corresponding to a product of experts model. This forms an attention mechanism, which we refer to as the moduli attention, and is learned on the top-down manifold while having a generative effect on underlying networks through approximate inference. This has recently been implemented using Stein Variational Gradient Descent algorithm (Jaini 2021). Recall, variational methods pick a family of distributions over the latent variables with their own variational parameters, $q(z_{1:m}|v)$, then attempt to find settings of the parameters that makes q close to the posterior of interest. Closeness of the two distributions is measured with the KL divergence, $KL(q||p) = E_q\left[\log \dfrac{q(z)}{p(z|x)}\right]$.

2.2 Instantons and Self-Organized Criticality

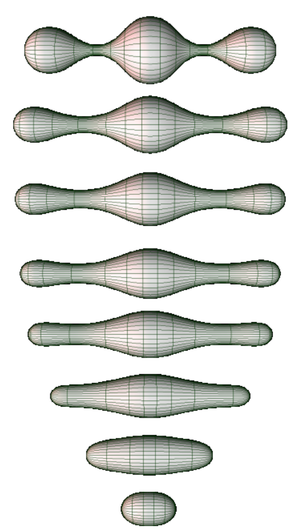
It is known that biological neural networks are organized based on self-organized criticality (SOC). Neuronal avalanches are scale-invariant neuronal population activity patterns in the cortex that are proposed to be a mechanism of cortical information processing and storage. Theory and experiments suggest neuronal avalanches allow for the transient and selective formation of local and system-wide spanning neuronal groups. The condensation of instantons can describe the noise-induced chaotic phase of SOC. A generic SOC system can be formulated as a Witten-type topological field theory (W-TFT) with spontaneously broken Becchi-Rouet-Stora-Tyutin (BRST) symmetry. In the parameter space, there must exist regions where the BRST-symmetry is spontaneously broken by instantons, which in the context of SOC are avalanches (Ovchinnikov 2011). Stochastic neural networks models have demonstrated how spontaneously broken BRST symmetry can describe SOC (Jian 2021).

2.3 Instantons as Islands in Part-Whole Hierarchies

The GLOM model (Hinton 2021) represents part-whole hierarchies using clusters of matching vectors, known as islands of agreement, as nodes in parse tree representations within a neural network. A part-whole hierachy system consists of components that can themselves be further broken down into subcomponents. The model learns analogical reasoning by encoding parts that are not well defined or obscured. An emphasis is given on viewpoint and coordinate invariance, which allows for exchangeability within the network and is used to adapt to and generate novelty in a manner similar to creativity in humans. Comparisons can be drawn between the GLOM model and moduli attention, with vector encodings corresponding to dynamic trajectories on vector bundles, frame invariance corresponding to the gauge equivariant connection space, and islands of agreement corresponding to stable instanton or soliton solutions. Iterative consensus often does not converge in meaningful ways due to the difficulty of encoding prior understandings of desirable representations to be generated through agglomerative clustering. Using a geometric model we find that we are better able to converge to coherent clusters because of the use of an a priori structure group that imposes meaningful symmetry on the clusters and optimizes organization of information throughout the space.

2.4 Intrinsic Geometric Flows

We can interpret energy minimization as a geometric flow, either on the parameter space as an extrinsic flow through variational gradient descent, or on the moduli space as an intrinsic flow. A Ricci flow is an example of an intrinsic flow on the metric by which one can take an arbitrary manifold and smooth out the geometry to make it more symmetric, whereas a mean curvature flow (as found in soap films, with critical points as minimal surfaces) is an example of an extrinsic flow on an embedded manifold. The moduli attention mechanism can be understood as a mix of the two categories; manipulating both the embedded manifolds constructed from hierarchical neural networks and the moduli space of connections (or metric) itself. We can further classify it as both a variational and a curvature flow since it evolves to minimize the Yang-mills action functional which is the L^2 norm of curvatures. Curvature flows may not necessarily preserve volume (the Calabi flow does, while the Ricci flow does not), meaning the flow may simply shrink or grow the manifold, rather than regularizing the metric. Instead of normalizing the flow by fixing the volume, we allow dissipative solutions to exist, forming a memory and compression mechanism.



Ricci flow of a metric manifold (Rubinstein, 2005)

2.5 Ricci Yang-Mills Solitons

A Ricci flow is a differential equation on the space of Riemannian metrics on $M, \mathfrak{Met}$. We can picture the Ricci flow as moving a manifold around by internal symmetries (the family of diffeomorphisms) and a uniform-in-space scaling at each time. If one works in the moduli space of $\mathfrak{Met}/\mathfrak{Diff}$, where $\mathfrak{Diff}$ is the group of diffeomorphisms on M, then one allows for a family of fixed points that are metrics that flow by scaling and diffeomorphism. i.e. $g(t) = \sigma(t)\phi(t)^* g_0$, where $\phi(t) : M \to M$ is a one parameter family of diffeomorphisms. These are the Ricci soliton metrics. In the standard quantum field theoretic interpretation of the Ricci flow in terms of the renormalization group, the parameter t corresponds to length or energy rather than time. One can show that Ricci soliton metrics satisfy the following equation: $Rc + \mathcal{L}_X g + \frac{\epsilon}{2} g = 0$, where X is the vector field generating the diffeomorphisms, and $\epsilon = -1, 0, 1$ corresponds to shrinking, steady, and expanding solitons respectively. If X is the gradient of some function, i.e. $X = \nabla f$, then a solution is said to be a gradient Ricci soliton. Similarly, the Ricci Yang-Mills flow is a natural coupling of the Ricci flow and the Yang-Mills heat flow. It was discovered that the Ricci Yang-Mills flow is an ideal candidate for studying magnetic flows. Given a choice h of metric on the Lie algebra g of G, a one-parameter family of metrics g_t on Σ and principal connections μ_t satisfies the RYM flow if,

$$\frac{\partial}{\partial t} g = -2Rc\, g + F_\mu^2, \qquad \frac{\partial}{\partial t} \mu = -d_g^* F_\mu. \tag{2}$$

2.6 Solitons as Invariant Points

Ricci solitons and Ricci Yang-Mills solitons can be naturally associated to invariant points in metric spaces using Geometric Invariant Theory (Jablonski 2013). Similarly, the Banach fixed-point theorem guarantees existence of a fixed point in certain metric spaces given a contractive mapping, and can be interpreted as solition solutions of a Ricci de Turck flow.

A comparison can be made with the Invariant Point Attention (IPA) mechanism introduced in Alphafold 2 (Jumper 2021). The invariant point attention augments each of the standard attention queries, keys, and values with 3-D points produced in the local frame of each protein residue gas such that the final value is invariant to global rotations and translations. After each attention operation and element-wise transition block, the module computes an update to the rotation and translation of each backbone frame. The application of these updates within the local frame of each residue makes the overall attention and update block an equivariant operation on the residue gas.

3 Topological Quantum Computing (Exploratory)

3.1 Quasiparticles

The objects representing solutions of variational flows can be studied as quasiparticles and have been made physical in quantum materials or gasses. This geometric formulation of general intelligence is a naturally arising and biologically plausible model for quantum AI. Aside from providing a computational scheme, it's worth acknowledging the philosophical implications of associating physical unified field theories with neuroscience and machine learning theory. Namely, that it may justify further introspection into the Anthropic Principle and cosmological fine-tuning.

Quasiparticles or collective excitations are emergent phenomena that encapsulate macroscopic portions of a complicated microscopic system such that the behaviour of these encapsulated parts imitate behaviours of weakly interacting particles in a vacuum. A soliton is a localized, non-dispersive solution of a nonlinear theory in Euclidean space and is a real object. Conversely, instantons are not real and only exist as solutions to the equations of motion of a quantum field theory after a Wick rotation, in which time is made imaginary. Thus, instantons are not observable, but are used to calculate and explain quantum mechanical effects that can be observed, such as tunneling. In quantum chromodynamics (QCD) instantons tunnel between the topologically different color vacua.

3.2 Bose-Einstein Condesate

A Bose–Einstein condensate (BEC) is a state of matter which is typically formed when a gas of bosons at low densities is cooled to temperatures very close to absolute zero causing a large fraction of the bosons to occupy the lowest quantum state, at which point microscopic quantum mechanical phenomena, particularly wavefunction interference, become macroscopic. The transition to BEC occurs below a critical temperature that is given by:

$$T_{\mathrm{c}} = \left(\frac{n}{\zeta(3/2)} \right)^{2/3} \frac{2\pi\hbar^2}{mk_{\mathrm{B}}}$$

T_{c} is the critical temperature, n the particle density, m the mass per boson, $\hbar$ the reduced Planck constant, k_{B} the Boltzmann constant and ζ the Riemann zeta function;

It was demonstrated to be possible to create magnetic solitons in a dipolar BEC made from quantum gases of atoms with different spins (Farolfi, 2020). These quantum solitons are density waves, meaning they are local waves of particles. Similarly, spatial phase distributions can be optically imprinted onto a BEC of atoms and have been shown to create solitons (Denschlag, 2000).

3.3 Energy-Based Models in Particle Systems

An energy function E in an EBM can be thought of as an unnormalized negative log probability (LeCun 2020). To convert an energy function to its equivalent probabilistic representation after normalization, $P(y \mid x)$, we apply the Gibbs-Boltzmann formula with latent variables z being marginalized implicitly through integration, i.e. $P(y \mid x) = \int_z P(y, z|x)$. Then,

$$P(y \mid x) = \frac{\int_z \exp(-\beta E(x, y, z))}{\int_y \int_z \exp(-\beta E(x, y, z))}$$

The derivation introduces a β term which is the inverse of temperature T, so as $\beta \to \infty$ the temperature goes to zero, and we see that $\check{y} = \text{argmin}_y E(x, y)$. We can redefine our energy function as an equivalent function with free energy F_β,

$$F_\infty(x, y) = \text{argmin}_z E(x, y, z)$$
$$F_\beta(x, y) = -\frac{1}{\beta} \log \int_z \exp(-\beta E(x, y, z)).$$

If we have a latent variable model and want to eliminate the latent variable z in a probabilistically correct way, we just need to redefine the energy function in terms of F_β,

$$P(y \mid x) = \frac{\exp(-\beta F_\beta(x, y, z))}{\int_y \exp(-\beta F_\beta(x, y, z))}. \tag{3}$$

With variational methods, instead of only minimizing the energy function with respect to z we must prevent the energy function from being 0 everywhere by constraining the flexibility of the latent variable z. The energy function is defined as sampling z randomly according to a distribution whose logarithm is the cost that links it to z. This distribution is commonly chosen to be a Gaussian with mean $\bar{z}$ which results in Gaussian noise being added to $\bar{z}$. The reparameterization trick is often used to allow for backpropagation during training despite the random sampling.

3.4 Abelian Sandpile Model

The toppling of several vertices during one iteration is referred to as an avalanche. After a finite number of topplings some stable configuration is reached such that the automaton is well defined. Although there will often be many possible choices for the order in which to topple vertices, the final stable configuration does not depend on the chosen order; this is one sense in which the sandpile is abelian.

Work in progress

References

[1] Gao, Tingran. "The diffusion geometry of fibre bundles: Horizontal diffusion maps." Applied and Computational Harmonic Analysis 50 (2021): 147-215.

[2] Eckhard Meinrenken, "Principal bundles and connections", Lecture Notes.

[3] Tao, T. "What is a gauge?" (2008) https://terrytao.wordpress.com/2008/09/27/what-is-a-gauge/

[4] Wetzel, Sebastian J., and Manuel Scherzer. "Machine learning of explicit order parameters: From the Ising model to SU (2) lattice gauge theory." Physical Review B 96.18 (2017): 184410.

[5] Du, Yilun, and Igor Mordatch. "Implicit generation and generalization in energy-based models." arXiv preprint arXiv:1903.08689 (2019).

[6] Blei, David M., Alp Kucukelbir, and Jon D. McAuliffe. "Variational inference: A review for statisticians." Journal of the American statistical Association 112.518 (2017): 859-877.

[7] Hinton, Geoffrey. "How to represent part-whole hierarchies in a neural network." arXiv preprint arXiv:2102.12627 (2021).

[8] Hinton, Geoffrey E. "Mapping part-whole hierarchies into connectionist networks." Artificial Intelligence 46.1-2 (1990): 47-75.

[9] Yann LeCun and Alfredo Canziani. "DEEP LEARNING". DS-GA 1008, SPRING 2020. NYU CENTER FOR DATA SCIENCE

[10] Russo, Abigail A., et al. "Neural trajectories in the supplementary motor area and motor cortex exhibit distinct geometries, compatible with different classes of computation." Neuron 107.4 (2020): 745-758.

[11] Farolfi, A., et al. "Observation of magnetic solitons in two-component Bose-Einstein condensates." Physical Review Letters 125.3 (2020): 030401.

[12] Denschlag, J., et al. "Generating solitons by phase engineering of a Bose-Einstein condensate." Science 287.5450 (2000): 97-101.

[13] J. Hyam Rubinstein and Robert Sinclair. "Visualizing Ricci Flow of Manifolds of Revolution", Experimental Mathematics v. 14 n. 3, pp. 257–384

[14] Jaini, Priyank, Lars Holdijk, and Max Welling. "Learning Equivariant Energy Based Models with Equivariant Stein Variational Gradient Descent." arXiv preprint arXiv:2106.07832 (2021).

[15] I.V. Ovchinnikov. Self-organized criticality as Witten-type topological field theory with spontaneously broken Becchi-Rouet-Stora-Tyutin symmetry. Physical Review E, 83,051129 (2011).

[16] Zhai, Jian, Chaojun Yu, and You Zhai. "Witten-type topological field theory of self-organized criticality for stochastic neural networks." arXiv preprint arXiv:2106.10851 (2021).

[17] Michael Jablonski and Andrea Young,Ricci Yang-Mills solitons on nilpotent Lie groups, J. Lie Theory23 (2013), no. 1, 177–202. MR 3060772

[18] Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature (2021). https://doi.org/10.1038/s41586-021-03819-2