Energy Minimizing Flows on Equivariant Space of Connections

L. J. Pereira

1 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 bounding space, a top-down energy-based attention mechanism 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 two 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, producing a Ricci Yang-Mills flow with solutions akin to Ricci solitons. Stable solutions represent islands of agreement formed through consensus within a dynamic part-whole hierarchy. The theoretical objects representing solutions of these flows are studied as quasiparticles and can be made physical in quantum materials or gasses. Our geometric formulation of general intelligence provides a natural and biologically plausible model for quantum AI.

2 Geometry of Latent Information

2.1 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 fitting 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 comes 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 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 connections between fibers exists, known as a principal Ehresmann connection, 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 being itself a differentiable manifold. An interesting phenomena 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.

2.2 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 (or synapses) and is increasingly positive when the pair show similar behavior and is negative when dissimilar. Moving from discrete random variables to fields and continuous manifolds (like EM fields), the covariant derivative between fibers of a bundle naturally arises and is known as a principal Ehressmann connection. To allow holonomy of dynamics on a bundle, it is necessary to impose a generalized a priori covariance principle 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 a Lie Group, but can also be achieved by imposing restrictions on the projection map of Riemannian manifolds without explicitly defining the structure group beforehand (Gao, 2021).

2.3 Moduli Spaces

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 Yang-Mills moduli space defined to be a finite dimensional manifold. This reduced space has local and global minima being connections with minimized energy known as Yang-Mills connections or instantons. Yang-Mills connections 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 between sampled trajectories of two manifolds which happen to be gauge equivariant. The gauge field strength is the curvature F_A of the connection, and the energy of the gauge field is given by the Yang–Mills action functional:

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

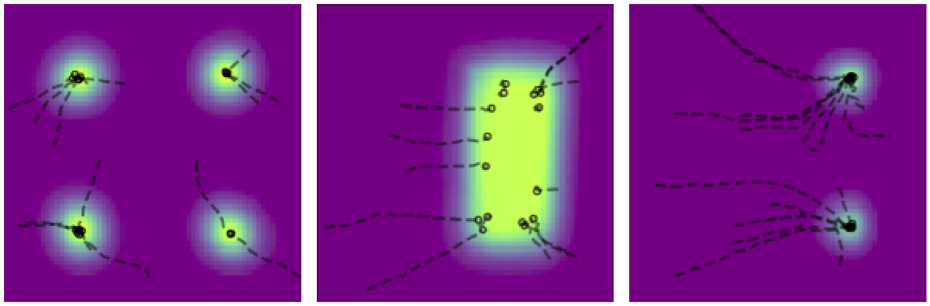
With the aim of having zero or vanishing curvature, we vary paramaters in search of a connection with curvature as small as possible. The Yang–Mills action functional simply corresponds to the L^2-norm of the curvature, and its Euler–Lagrange equations describe the critical points of this functional, either the absolute minima or local minima.

3 Variational Geometric Flows

3.1 Variational Inference and Energy Composition

As underlying neural networks perform inference, their latent vector trajectories pass through hierarchies of weighted hidden layers 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 that is learned on the top-down manifold as well as creating a generative effect on underlying networks through variational inference. The idea behind variational methods is to pick a family of distributions over the latent variables with its own variational parameters, $q(z_{1:m}|v)$, and then attempt to find the setting of the parameters that makes q close to the posterior of interest (the instanton). Closeness of the two distributions in variational inference is measured with the Kullback-Leibler (KL) divergence,

$$KL(q||p) = E_q\left[\log \frac{q(z)}{p(z|x)}\right]. \tag{2}$$

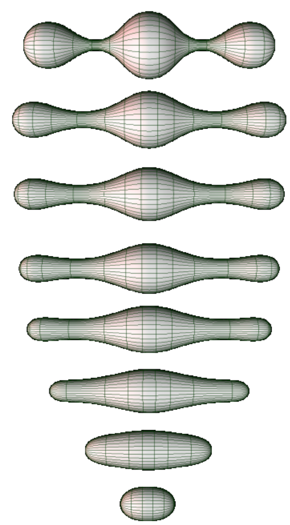


ENERGY A ENERGY B ENERGY A + B

A 2D example of combining energy functions through their summation and the resulting sampling trajectories. (Yilun, 2019)

3.2 Geometric Flows

We can interpret energy minimization through variational methods as a geometric flow on the moduli space (for intrinsic flows) or on the parameter space (for extrinsic flows). A Ricci flow is an example of an intrinsic flow on the metric, 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 variational method for minimizing energy described previously can be understood as a mix of the two; changing both the embedded manifolds constructed with hierarchical neural networks and the moduli space of connections (or metric) itself. We can further classify it as both a variational and curvature flow since it evolves to minimize the Yang-mills action functional which is the L^2 norm of curvatures. The Ricci flow, Calabi flow, and Yamabe flow all arise in similar ways. 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. To avoid this, it's possible to normalize the flow by fixing the volume.



Ricci flow of a metric manifold (Rubinstein, 2005)

3.3 Islands in Part-Whole Hierarchies

The GLOM model (Hinton 2021) aims to represent part-whole hierarchies, using islands of matching vectors, known as islands of agreement, to construct nodes in parse tree representations within a neural network. A part-whole hierachy is simply a system consisting of subsystems or components that can themselves be broken down into further subcomponents. By finding clusters of coherence, the model attempts learn intuition and analogical reasoning by encoding parts that are not well defined, unconscious, or possibly obscured. Moreover, Hinton describes the importance of viewpoint and coordinate-invariance allowing for exchangibility within the network, which can be used to adapt to and generate novelty in a way similar to humans. We can draw comparisons between the GLOM model and learning energy minimizations on a moduli space, where vector encodings correspond to trajectories on vector bundles, frame invariance corresponds to the gauge invariance established by quotienting the space of principal connections of the vector bundles by Lie structure groups, and islands of agreement correspond to using variational, curvature Geometric flows that manifest stable instanton or soliton solutions.

Iterative consensus generally doesn't converge in coherent ways because of the difficulty of encoding a prior understanding of desirable representations to be generated through the agglomerative clustering. However, using a geometric model we find that we are better posed to converge to coherent clusters because of the use of an a priori structure groups which imposes meaningful symmetry to the clusters and an optmized organization of information throughout the space. Moreover, using Ricci flows to interpret the process, we gain access to a trove of mathematical research on the solubility, stability, and other convergence properties and behaviour.

3.4 Variational Methods in EBMs

An energy function E in an EBM can be thought of as an unnormalized negative log probability. 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 $\check{y} = \text{argmin}_y E(x, y)$. This inverse temperature limit appears similar to the critical temperature in BEC. 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.

4 Topological Quantum Computing (Exploratory)

4.1 Quasiparticles

Quasiparticles and collective excitations are emergent phenomena that encapsulate sections of a microscopically complicated system with their behaviour imitating different behaviour 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. Note, a Wick rotation is a transformation that substitutes an imaginary-number variable for a real-number variable in order to solve a problem of (complex) Minkowski space in Euclidean space. Therefore, 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 are believed to tunnel between the topologically different color vacua.

Given a smooth manifold M and an open real interval (a, b), a Ricci flow assigns to each $t \in (a, b)$ a Riemannian metric g_t on M such that

$$\frac{\partial}{\partial t} g_t = -2 \operatorname{Ric}^{g_t} . \tag{4}$$

Ricci solitons are exactly the solutions to the Ricci flow that are of the form

$$g(t) = \sigma(t) \phi_t^* g_0 \tag{5}$$

where g_0 is some fixed metric, ϕ_t is a one-parameter family of diffeomorphisms with $\phi_0 = id$, and σ is a real-valued function satisfying $\sigma(0) = 1$ (so that $g(0) = g_0$). One can picture the Ricci flow in this situation as moving the manifold around by internal symmetries (the family of diffeomorphisms) and a uniform-in-space scaling at each time. In this sense, the intuition is that the manifold maintains the same shape but expands or contracts. 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

Aside from providing a quantum computational schema, it's also worth noting the philosophical implications of associating mathematics of physical unified field theories with neuroscience and machine learning theory. Namely, that it may justify further introspection into the Anthropic Principle and the nature of cosmological fine-tuning which appears to be reflective in our own consciousness.

4.2 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. This transition to BEC occurs below a critical temperature that is given by:

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

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;

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

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] Yann LeCun, DS-GA 1008, DEEP LEARNING, NYU CENTER FOR DATA SCIENCE

[9] 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.

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

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

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