Article

Not peer-reviewed version

Entropic Geometry and Symmetry Breaking in Lie-Group Free-Energy Minimization

Noboru Sagae *

Posted Date: 6 November 2025

doi: 10.20944/preprints202511.0382.v1

Keywords: entropic geometry; Lie groups; variational free energy; information flow; symmetry breaking; gradient flows

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 November 2025 doi:10.20944/preprints202511.0382.v1

Article

Entropic Geometry and Symmetry Breaking in Lie-Group Free-Energy Minimization

Noboru Sagae

Keio University, Japan; nsagae1970@gmail.com

Abstract

We present a geometric formulation of **entropic free-energy minimization** as Riemannian gradient descent on Lie-group orbits endowed with the Fisher information metric. This approach reveals how **symmetry structures constrain the dynamics of information and entropy reduction**, linking variational inference to geometric thermodynamics. We establish well-posedness, Lyapunov monotonicity, and convergence theorems, and derive a second-variation criterion explaining **entropic symmetry breaking and bifurcations**. Examples on Gaussian families under translations and rotations illustrate the interplay between group invariance and adaptive stability. The results provide a unified view connecting **information geometry, thermodynamics, and the Free Energy Principle** through a group-theoretic lens.

Keywords: entropic geometry; Lie groups; variational free energy; information flow; symmetry breaking; gradient flows

MSC: Primary 62B10; Secondary 53C20, 37N25, 60Gxx

---

1. Introduction

The concept of *free energy* unifies thermodynamic, informational, and biological principles of organization. Within the *Free Energy Principle* (FEP), adaptive systems minimize an internal free-energy functional to resist disorder and maintain self-organization [1,2]. Yet, the geometric structure underlying this process—how symmetry and invariance constrain information flow—remains underexplored. Here we formulate free-energy minimization as an *entropic gradient flow* on Lie-group orbits, allowing a unified view of symmetry, stability, and self-organization. Recent expositions provide simplified overviews and technical clarifications that we leverage here [3,4].

Contributions

We provide: (i) an orbit-reduced formulation of variational free energy; (ii) first- and second-variation formulas on orbits; (iii) local well-posedness and Lyapunov monotonicity; (iv) convergence on compact groups and under the Kurdyka–Łojasiewicz (KŁ) property; (v) stability criteria via the orbit-restricted Hessian; (vi) a pitchfork scenario for SO(2); and (vii) worked Gaussian examples with explicit derivatives.

2. Preliminaries

2.1. Standing Assumptions and Notation

(Q1) $Q = \{q_\theta : \theta \in \Theta \subset \mathbb{R}^n\}$ is a C^2 statistical manifold with Fisher metric g^{F}. (Q2) G is a finite-dimensional Lie group acting smoothly on Q by ρ, with stabilizer H at q_0; the orbit $\mathcal{O}_{q_0} \cong G/H$ is an embedded submanifold. (Q3) The free-energy functional $\mathcal{F} : Q \to \mathbb{R}$ is C^2 and bounded below on $\mathcal{O}_{q_0}$. (Q4) G carries a right-invariant C^1 Riemannian metric γ. Throughout, $F(g) := \mathcal{F}[\rho(g, q_0)]$, ∇_G denotes the gradient with respect to γ, and L_X the infinitesimal action for $X \in \mathfrak{g}$.

2.2. *Variational Free Energy*

Given a generative model $p(s, u, \theta)$ and a variational posterior $q(u, \theta)$,

$$\mathcal{F}[q] = \mathbb{E}_q[\log q(u,\theta)] - \mathbb{E}_q[\log p(s,u,\theta)] = \mathrm{KL}\big(q(u,\theta) \,\|\, p(u,\theta \mid s)\big) - \log p(s), \tag{1}$$

so minimizing $\mathcal{F}$ recovers the evidence lower bound.

2.3. *Statistical Manifolds*

Let $Q = \{q_\theta : \theta \in \Theta\}$ with log-density $\ell(x,\theta) = \log q_\theta(x)$. The Fisher information metric is

$$g_{ij}(\theta) = \mathbb{E}_{q_\theta}\left[\partial_i \ell \, \partial_j \ell\right] = -\mathbb{E}_{q_\theta}[\partial_i \partial_j \ell], \tag{2}$$

with natural gradient $\nabla^{\mathrm{nat}} F(\theta) = G(\theta)^{-1} \nabla F(\theta)$ for smooth $F : Q \to \mathbb{R}$.

2.4. *Group Actions and Orbits*

Optimization restricted to the orbit $\mathcal{O}_{q_0} = \{\rho(g, q_0) : g \in G\} \cong G/H$ converts $\min_{q \in Q} \mathcal{F}[q]$ to $\min_{g \in G} F(g)$ with $F(g) = \mathcal{F}[\rho(g, q_0)]$.

3. Orbit Gradients

We make precise the identification between the natural gradient on Q and the Riemannian gradient on G.

Theorem 1 (Equivalence of natural and group gradients on the orbit)**.** *Under (Q1)–(Q4), let $\xi \mapsto g(\xi) = \exp(\sum_i \xi^i X_i)$ parametrize G near e, and set $q(\xi) = \rho(g(\xi), q_0)$. Assume the Jacobian $J(\xi) = \partial q(\xi)/\partial \xi$ has full column rank along $\mathcal{O}_{q_0}$. Equip G with the induced metric $G_G(\xi) = J(\xi)^\top G_Q(\xi) J(\xi)$ from the Fisher metric G_Q. Then for any smooth $F(g) = \mathcal{F}[\rho(g, q_0)]$, the natural gradient of $\mathcal{F}$ restricted to $\mathcal{O}_{q_0}$ corresponds exactly to the Riemannian gradient on (G, G_G); i.e.*

$$\mathrm{Proj}_{T\mathcal{O}_{q_0}}\left(\nabla^{\mathrm{nat}} \mathcal{F}(q(\xi))\right) \;\leftrightarrow\; \nabla_G F(g(\xi)).$$

Proof. Let $\delta\xi \in \mathbb{R}^{\dim G}$ and consider the variation $g(\xi + t\,\delta\xi)$. By the chain rule, $\mathrm{d}F = \mathrm{d}\mathcal{F} \circ J$, where $J = \partial q/\partial \xi$. Denote by G_Q the Fisher metric on TQ and define an induced inner product on parameter increments by $\langle \delta\xi_1, \delta\xi_2 \rangle_{G_G} := \langle J\delta\xi_1, J\delta\xi_2 \rangle_{G_Q}$. This is positive definite by full column rank of J. The Riesz representation of $\mathrm{d}F$ with respect to $\langle \cdot, \cdot \rangle_{G_G}$ yields the unique η such that $\mathrm{d}F[\delta\xi] = \langle \eta, \delta\xi \rangle_{G_G}$ for all $\delta\xi$. But $\mathrm{d}F[\delta\xi] = \left\langle \nabla^{\mathrm{nat}} \mathcal{F}, J\,\delta\xi \right\rangle_{G_Q} = \left\langle J^\top G_Q \nabla^{\mathrm{nat}} \mathcal{F}, \delta\xi \right\rangle_{\mathrm{Eucl}}$. Thus η corresponds to $G_G^{-1} J^\top G_Q \nabla^{\mathrm{nat}} \mathcal{F}$, which is precisely the coordinate representation of the Riemannian gradient on (G, G_G). Projecting to $T\mathcal{O}_{q_0}$ accounts for any null directions associated with the stabilizer H. $\square$

Remark 1 (Coordinate formula)**.** *In local coordinates, $\nabla_G F = G_G^{-1} J^\top \nabla_Q \mathcal{F}$, with $G_G = J^\top G_Q J$. This realizes a Gauss–Newton structure typical of natural-gradient methods.*

4. Gradient Flows and Convergence

Define the gradient flow on G by

$$\dot{g}(t) = -\nabla_G F(g(t)), \qquad g(0) = g_0 \in G. \tag{3}$$

Theorem 2 (Local well-posedness)**.** *If $F \in C^1(G)$ and $\nabla_G F$ is locally Lipschitz with respect to γ, then* (3) *admits a unique maximal solution from any initial point.*

Proof. Right-translate by $R_{g(t)^{-1}}$ and write $\dot{g} = \mathrm{d}R_g\, v$ with $v(t) \in \mathfrak{g}$. The map $g \mapsto v(g) := -\mathrm{d}R_{g^{-1}} \nabla_G F(g)$ is locally Lipschitz on charts. This gives a locally Lipschitz ODE $\dot{v} = f(v)$ on $\mathfrak{g}$ in coordinates; Picard–Lindelöf on manifolds (via charts and partition of unity) yields a unique solution. $\square$

Theorem 3 (Lyapunov monotonicity)**.** *Along any solution of* (3)*,*

$$\frac{\mathrm{d}}{\mathrm{d}t}F\big(g(t)\big) = \langle \mathrm{d}F,\, \dot{g}\rangle = \gamma\big(\nabla_G F, \dot{g}\big) = -\|\nabla_G F\|_\gamma^2 \le 0. \tag{4}$$

If F is bounded below, the limit $\lim_{t\to\infty} F\big(g(t)\big)$ *exists.*

Proof. The identity $\mathrm{d}F(\xi) = \gamma(\nabla_G F, \xi)$ holds by definition of the gradient. Substitute $\dot{g} = -\nabla_G F$. □

Theorem 4 (Asymptotics on compact groups)**.** *If G is compact and $F \in C^2(G)$ has only nondegenerate critical points, then every trajectory of* (3) *has ω-limit set contained in the (finite) set of critical points. In particular, every bounded trajectory approaches the set of critical points; moreover, if F additionally satisfies the KŁ property at its critical points (e.g. F is real-analytic), then the trajectory converges to a single critical point.*

Proof. Compactness implies that sublevel sets $\{F \le c\}$ are compact. By Theorem 3, $F(g(t))$ decreases and is bounded below, hence $g(t)$ remains in a compact set and admits accumulation points, all critical by $\lim \|\nabla_G F\| = 0$. Nondegeneracy and the stable manifold theorem imply convergence to a single critical point. □

Theorem 5 (Convergence under KŁ property)**.** *Assume $F \in C^1(G)$ is bounded below and satisfies the Kurdyka–Łojasiewicz property at every critical point (e.g., F is real-analytic on an analytic G). Then every bounded trajectory of $\dot{g} = -\nabla_G F(g)$ has finite length and converges to a single critical point as $t \to \infty$.*

Proof. Let $g(t)$ be bounded with $F(g(t)) \downarrow F_\infty$. The KŁ inequality provides $\varphi'(F - F_\infty)\,\|\nabla_G F\| \ge 1$ near the limit set for some desingularizing function φ. Integrating $\|\dot{g}\| = \|\nabla_G F\|$ over time and applying the inequality shows $\int_0^\infty \|\dot{g}\|\,\mathrm{d}t < \infty$, hence $g(t)$ has finite length and is Cauchy; completeness of G yields convergence to a critical point. □

5. Second Variation and Stability

We derive the orbit-restricted Hessian and a stability test.

Lemma 1 (First and second variations along one-parameter subgroups)**.** *For $X \in \mathfrak{g}$ and $g(t) = g\exp(tX)$,*

$$\frac{\mathrm{d}}{\mathrm{d}t}F\big(g(t)\big)\Big|_0 = \Big\langle \mathrm{d}\mathcal{F}|_{\rho(g,q_0)},\, L_X\,\rho(g,q_0)\Big\rangle. \tag{5}$$

If $\mathcal{F} \in C^2$, then along $X, Y \in \mathfrak{g}$ at a critical g^* with $q^* = \rho(g^*, q_0)$,

$$\mathrm{Hess}_G F(g^*)[X,Y] \;=\; D^2\mathcal{F}|_{q^*}\big[L_X q^*,\, L_Y q^*\big] \;+\; \tfrac{1}{2}\, D\mathcal{F}|_{q^*}\big[L_{[X,Y]} q^*\big]. \tag{6}$$

Proof. For (5), differentiate $F(g\exp(tX)) = \mathcal{F}[\rho(g\exp(tX), q_0)]$ and use $\frac{\mathrm{d}}{\mathrm{d}t}\rho(g\exp(tX), q_0)|_{t=0} = L_X\rho(g, q_0)$. For (6), differentiate (5) again in the direction Y and invoke the symmetry of $D^2\mathcal{F}$ plus the Lie-bracket identity $\frac{\mathrm{d}}{\mathrm{d}s}L_X\rho(g\exp(sY), q_0)|_{s=0} = L_{[Y,X]}\rho(g, q_0)$. □

Proposition 1 (Stability on the orbit)**.** *Let $\mathfrak{h}$ be the Lie algebra of the stabilizer at q^*. If the quadratic form $X \mapsto \mathrm{Hess}_G F(g^*)[X,X]$ is positive definite on $\mathfrak{g}/\mathfrak{h}$, then g^* is a strict local minimum and asymptotically stable for* (3)*. Negative/indefinite signatures yield maxima/saddles.*

Proof. Positive definiteness implies that $F(g) \ge F(g^*) + c\,\mathrm{dist}(g, g^*)^2$ in a neighborhood (in exponential coordinates), hence F is a Lyapunov function with a strict minimum. The linearization of (3) at g^* has spectrum in $(-\infty, 0)$ on $\mathfrak{g}/\mathfrak{h}$, yielding asymptotic stability by the Hartman–Grobman theorem on manifolds. □

6. Symmetry and Bifurcation

Definition 1 (Group invariance)**.** *The functional $\mathcal{F}$ is G-invariant if $\mathcal{F}[\rho(g,q)] = \mathcal{F}[q]$ for all $g \in G$, $q \in Q$. Then F is constant on G, and every g is critical. Partial invariance or data terms induce nontrivial landscapes.*

Theorem 6 (Pitchfork for $\mathrm{SO}(2)$)**.** *Let $\mathrm{SO}(2)$ act on planar Gaussians by covariance conjugation, and consider a one-parameter family F_λ. If at $\lambda = \lambda_c$ the smallest nonzero orbit-restricted eigenvalue of the Hessian crosses zero while symmetry suppresses the cubic term, then two nontrivial minima bifurcate from the symmetric one as λ passes through λ_c.*

Sketch. Normal-form reduction on the one-dimensional orbit coordinate θ gives $F_\lambda(\theta) = a(\lambda)\theta^2 + b\theta^4 + o(\theta^4)$ with $a(\lambda_c) = 0$, $b > 0$, and the odd cubic suppressed by symmetry. The change of sign of a yields a supercritical pitchfork. ☐

7. Examples

7.1. Translations of Means (Abelian Case)

Let Q be d-dimensional Gaussians with mean μ and fixed covariance Σ. The additive group $(\mathbb{R}^d, +)$ acts by $(\rho(\epsilon)q)(x) = q(x - \epsilon)$. Consider

$$\mathcal{F}[q] = \int \|x - \mu\|^2 q(x)\,\mathrm{d}x + \lambda\,\mathrm{KL}(q\|p). \tag{7}$$

Proposition 2. *The orbit cost $\epsilon \mapsto F(\epsilon) = \mathcal{F}[\rho(\epsilon, q)]$ is strictly convex and admits a unique minimizer ϵ^* with $\nabla F(\epsilon^*) = 0$.*

Proof. Convexity follows from convexity of the squared norm and the joint convexity of $\mathrm{KL}(\cdot\|p)$ under translations. Strictness holds unless p is itself translation invariant. Differentiating under the integral sign gives the first-order condition. ☐

7.2. Planar Rotations of Covariances

For zero-mean Gaussians,

$$\mathrm{KL}\big(\mathcal{N}(0,\Sigma)\,\|\,\mathcal{N}(0,\Sigma_0)\big) = \tfrac{1}{2}\Big(\mathrm{tr}(\Sigma_0^{-1}\Sigma) - \log\frac{\det\Sigma}{\det\Sigma_0} - d\Big). \tag{8}$$

Let $\Sigma(\theta) = R(\theta)\Sigma R(-\theta)$ with $R(\theta)$ the 2×2 rotation. Then

$$\frac{\mathrm{d}\Sigma(\theta)}{\mathrm{d}\theta} = [\Omega, \Sigma(\theta)], \qquad \Omega = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{9}$$

Hence

$$\frac{\mathrm{d}}{\mathrm{d}\theta}F(\theta) = \frac{1}{2}\,\mathrm{tr}\big(\Sigma_0^{-1}[\Omega, \Sigma(\theta)]\big) = \frac{1}{2}\,\mathrm{tr}\big([\Sigma_0^{-1}, \Omega]\,\Sigma(\theta)\big). \tag{10}$$

Critical points satisfy $\mathrm{tr}\big([\Sigma_0^{-1}, \Omega]\,\Sigma(\theta)\big) = 0$, i.e. simultaneous diagonalizability of $\Sigma(\theta)$ and Σ_0; anisotropy yields two symmetric minima modulo π.

7.3. Rigid Motions $\mathrm{SE}(2)$ and Outlook to $\mathrm{SE}(3)$

The special Euclidean group couples rotations and translations; the induced metric on G blends mean and covariance directions. The same formalism extends to $\mathrm{SE}(3)$ for 3D pose models.

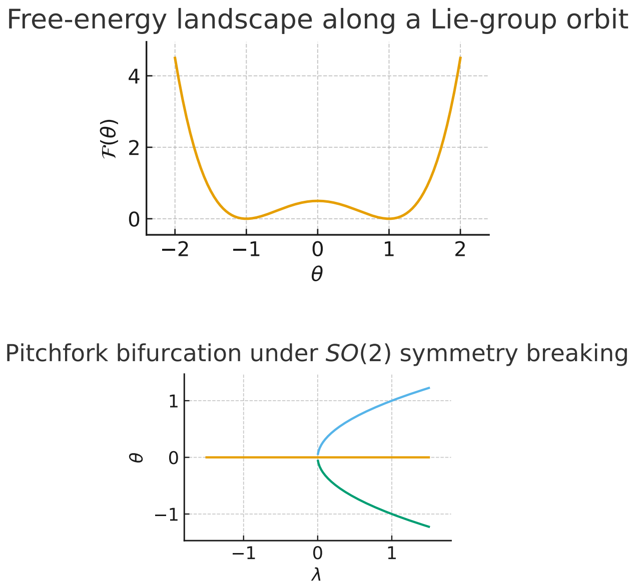

Figure 1. (Top) Free-energy landscape along a Lie-group orbit. (Bottom) Pitchfork bifurcation under $SO(2)$ symmetry breaking.

8. Entropy Reduction Under Lie-Group Symmetry

This section interprets free-energy descent as a physical process of entropy reduction constrained by Lie-group symmetry.

Thermodynamic entropy and informational stability.

On Lie-group orbits, entropy reduction corresponds to the dissipation of uncertainty along symmetry-constrained manifolds. The Fisher metric quantifies the local curvature of information, and its geodesic flow describes the most efficient direction of entropy decrease under free-energy descent. Thus, stability of a group orbit can be interpreted as a steady state with (possibly nonzero) steady entropic production balanced by symmetry-constrained fluxes.

Information-theoretic analogy.

Write the free energy as

$$\mathcal{F}[q] \;=\; \mathbb{E}_q[\log q] - \mathbb{E}_q[\log p] \;=\; -S(q) \;-\; \mathbb{E}_q[\log p].$$

Along any smooth evolution of q, we have

$$\dot{\mathcal{F}} \;=\; -\dot{S} \;-\; \frac{d}{dt}\mathbb{E}_q[\log p].$$

Hence entropy reduction and evidence gain are coupled but not identical in general. In our orbit-restricted dynamics, Lyapunov monotonicity yields $\dot{\mathcal{F}} = -\|\nabla_G F\|_\gamma^2 \leq 0$, which we interpret as a nonnegative "entropic production rate" on the group manifold. Lie-group symmetries then restrict admissible directions of this production, shaping the accessible information flows.

Physical interpretation.

In this light, the orbit flow $\dot{g} = -\nabla_G F(g)$ describes a dissipative process that drives the system toward minimal free energy under constraints imposed by G. Equilibria on orbits correspond to stationary nonequilibrium states, and bifurcations of F signal transitions between entropic basins—an informational analogue of phase transitions.

9. Related Work

Classical information geometry [7,8] endows models with the Fisher metric and dual connections; natural gradient methods exploit this geometry. In contrast, we constrain inference to Lie-group orbits and move optimization to the group manifold. This orbit-centric view enables stability and bifurcation analyses less transparent in parameter space. We also draw on recent discussions of the FEP and Bayesian mechanics [3–6].

10. Discussion

We presented a geometric reduction of FEP to optimization on Lie groups, with explicit stability and bifurcation analyses. Future work includes thermodynamic formulations of entropy reduction on noncommutative groups ($\mathrm{SE}(3)$, matrix groups), and numerical exploration of entropic symmetry breaking. This framework bridges information geometry, nonequilibrium thermodynamics, and adaptive inference in a single mathematical language.

Acknowledgments: The author thanks colleagues for discussions on information geometry, Lie theory, and variational principles.

References

1. K. Friston, A theory of cortical responses, *Philosophical Transactions of the Royal Society B* 360, 815–836 (2005).
2. K. Friston, The free-energy principle: a unified brain theory?, *Nature Reviews Neuroscience* 11, 127–138 (2010).
3. K. Friston, The free energy principle made simpler but not too simple, *Physics Reports* 1024, 1–43 (2023). doi:10.1016/j.physrep.2023.10.001
4. K. J. Friston, L. Da Costa, T. Parr, Some Interesting Observations on the Free Energy Principle, *Entropy* 23(8), 1076 (2021). doi:10.3390/e23081076
5. L. Da Costa, K. Friston, C. Heins, G. A. Pavliotis, Bayesian mechanics for stationary processes, *Proceedings of the Royal Society A* 477(2256), 20210518 (2021). doi:10.1098/rspa.2021.0518
6. P. Ao, Emerging of Stochastic Dynamical Equalities and Steady State Thermodynamics from Darwinian Dynamics, *Communications in Theoretical Physics* 49(5), 1073–1090 (2008). doi:10.1088/0253-6102/49/5/11
7. S.-I. Amari and H. Nagaoka, *Methods of Information Geometry*, AMS/OUP (2000).
8. S.-I. Amari, *Information Geometry and Its Applications*, Springer (2016).
9. F. Otto, The geometry of dissipative evolution equations: the porous medium equation, *Communications in Partial Differential Equations* 26(1–2), 101–174 (2001).
10. C. Villani, *Optimal Transport: Old and New*, Springer, Berlin (2009).
11. J. M. Lee, *Introduction to Smooth Manifolds*, 2nd ed., Springer (2012).
12. F. W. Warner, *Foundations of Differentiable Manifolds and Lie Groups*, Springer (1983).
13. S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*, Academic Press (1978).