Control Functionals in the $O(3)$ Sigma Model (Lie-derivative formulation)

Kostas Orginos

January 31, 2026

Abstract

We formulate control functionals for Monte-Carlo simulations of the lattice $O(3)$ non-linear sigma model using Lie derivatives on the target manifold. The construction parallels the scalar-field formulation based on the defining PDE $(\partial^2 - \partial S \cdot \partial)\, U = O - \mu$, but replaces ordinary field derivatives by left-invariant (Lie) derivatives at each lattice site. We derive the associated control variate, prove it has zero mean, and show that the PDE operator is self-adjoint and positive with respect to the Gibbs inner product. We work out the leading terms of the weak-coupling expansion in β for the two-point function at separation $|y - z| > 1$, and suggest symmetry-respecting parametrizations of U suitable for machine learning.

1 Notation and model

Lattice sites vs. Lie algebra indices. Throughout, $y, z, w, \ldots$ denote lattice sites. Lie algebra directions are denoted by $a, b, c, \ldots \in \{1, 2, 3\}$.

1.1 $O(3)$ sigma model

Let $n_y \in S^2 \subset \mathbb{R}^3$ be a unit vector at each lattice site y. We take the nearest-neighbor action

$$S[n] = \beta S_0[n], \qquad S_0[n] = -\sum_{\langle yw \rangle} n_y \cdot n_w. \tag{1}$$

The partition function and expectation values are

$$Z = \int \mathcal{D}n\, e^{-S[n]}, \qquad \langle A \rangle \equiv \frac{1}{Z} \int \mathcal{D}n\, A[n]\, e^{-S[n]}. \tag{2}$$

Here $\mathcal{D}n$ is the product of normalized invariant measures on S^2 at each site.

We will focus on the (translation-summed) two-point observable

$$C(r) \equiv \sum_y n_y \cdot n_{y+r}, \qquad \mu_r \equiv \langle C(r) \rangle\,. \tag{3}$$

Many derivations are clearer for fixed sites y, z; in that case set $O[n] = n_y \cdot n_z$ and $\mu = \langle O \rangle$, and at the end one may sum over translations.

2 Lie derivatives and ∂^2

2.1 Lie derivatives

Let $\{\lambda^a\}$ be anti-hermitian generators of $\mathfrak{so}(3)$ in the defining representation: λ^a are real antisymmetric 3×3 matrices obeying

$$[\lambda^a, \lambda^b] = \epsilon^{abc}\lambda^c. \tag{4}$$

Define the Lie derivative at site y by its action on spins:

$$\partial^a_y n_w = (\lambda^a n_w)\, \delta_{y,w}. \tag{5}$$

For any functional $U[n]$, $\partial^a_y U[n]$ is defined by the induced variation of U under an infinitesimal rotation at site y.

2.2 The Casimir / Laplacian ∂^2

Define the quadratic Casimir at site y and globally by

$$\partial^2_y \equiv \sum_{a=1}^{3} (\partial^a_y)^2, \qquad \partial^2 \equiv \sum_y \partial^2_y. \tag{6}$$

On functions depending only on one spin n_y, ∂^2_y coincides with the Laplace–Beltrami operator on S^2.

Useful identities (with fixed $u, v \in \mathbb{R}^3$) are

$$\partial^2_y (n_y \cdot u) = -2(n_y \cdot u), \tag{7}$$

$$\partial^2_y [(n_y \cdot u)(n_y \cdot v)] = -6(n_y \cdot u)(n_y \cdot v) + 2(u \cdot v). \tag{8}$$

Finally, from the $\mathfrak{so}(3)$ algebra in the defining representation one obtains the projector identity

$$\sum_{a=1}^{3} \big[(\lambda^a n) \cdot u\big] \big[(\lambda^a n) \cdot v\big] = u \cdot v - (n \cdot u)(n \cdot v). \tag{9}$$

3 Control functionals: definition and why they work

3.1 The defining PDE and the control variate

Let $O[n]$ be any observable and $\mu \equiv \langle O \rangle$. Given a functional $U[n]$, define

$$F_U[n] \equiv \partial^2 U[n] - \left(\partial^a_y U[n]\right) \left(\partial^a_y S[n]\right), \tag{10}$$

with the usual implied sums over y and a.

The "control-functional PDE" is

$$\partial^2 U[n] - \left(\partial^a_y U[n]\right) \left(\partial^a_y S[n]\right) = O[n] - \mu. \tag{11}$$

If U satisfies (11) exactly, then $F_U = O - \mu$ and the improved estimator $O - F_U$ is the constant μ (i.e. zero variance). In practice, one uses approximations to U.

3.2 Why F_U has zero mean (control variate property)

We show $\langle F_U \rangle = 0$ for any sufficiently regular U. The key tool is integration by parts with respect to Lie derivatives (invariance of the Haar measure on each S^2 factor).

Lie integration by parts. For any functionals $A[n], B[n]$ with appropriate regularity,

$$\int \mathcal{D}n \, (\partial_y^a A[n]) \, B[n] = -\int \mathcal{D}n \, A[n] \, (\partial_y^a B[n]). \tag{12}$$

This follows because ∂_y^a generates an infinitesimal symmetry of the measure at site y.

Now compute

$$\begin{aligned}
\langle \partial^2 U \rangle &= \frac{1}{Z} \int \mathcal{D}n \, (\partial_y^a \partial_y^a U) \, e^{-S} \\
&= -\frac{1}{Z} \int \mathcal{D}n \, (\partial_y^a U) \, (\partial_y^a e^{-S}) \qquad \text{by (12)} \\
&= \frac{1}{Z} \int \mathcal{D}n \, (\partial_y^a U) \, (\partial_y^a S) \, e^{-S} \\
&= \langle (\partial_y^a U)(\partial_y^a S) \rangle .
\end{aligned} \tag{13}$$

Therefore

$$\langle F_U \rangle = \langle \partial^2 U \rangle - \langle (\partial_y^a U)(\partial_y^a S) \rangle = 0. \tag{14}$$

Hence F_U is a valid *control variate* for any U.

3.3 Variance reduction

Given any U, define the improved estimator

$$\widehat{O}_U[n] \equiv O[n] - F_U[n]. \tag{15}$$

Since $\langle F_U \rangle = 0$, it is unbiased: $\left\langle \widehat{O}_U \right\rangle = \langle O \rangle = \mu$. Moreover, choosing U to make F_U strongly correlated with O reduces the variance of $\widehat{O}_U$. The formal optimal choice is the exact solution of (11), yielding $\widehat{O}_U = \mu$.

4 Operator formulation: self-adjointness and positivity

4.1 Gibbs inner product

Define the Gibbs inner product

$$(A, B) \equiv \langle A \, B \rangle = \frac{1}{Z} \int \mathcal{D}n \, A[n] \, B[n] \, e^{-S[n]}. \tag{16}$$

(For complex functionals one may insert complex conjugation on the first slot; in most applications U is real.)

4.2 The PDE operator

Define the differential operator L by

$$(LU)[n] \equiv -\, e^{S[n]}\, \partial_y^a\Big(e^{-S[n]}\, \partial_y^a U[n]\Big). \tag{17}$$

Expanding (17) using $\partial_y^a(e^{-S}) = -(\partial_y^a S)e^{-S}$ gives

$$(LU)[n] = -\partial^2 U[n] + (\partial_y^a S[n])(\partial_y^a U[n]). \tag{18}$$

Thus the defining PDE (11) is equivalently

$$LU = \mu - O. \tag{19}$$

4.3 Self-adjointness

Using (16) and (17),

$$\begin{aligned}
(A, LB) &= \frac{1}{Z}\int \mathcal{D}n\ A\Big(-e^{S}\partial_y^a(e^{-S}\partial_y^a B)\Big)e^{-S} \\
&= -\frac{1}{Z}\int \mathcal{D}n\ A\,\partial_y^a(e^{-S}\partial_y^a B) \\
&= \frac{1}{Z}\int \mathcal{D}n\ (\partial_y^a A)\,(e^{-S}\partial_y^a B) \qquad \text{by (12)} \\
&= \big\langle(\partial_y^a A)(\partial_y^a B)\big\rangle\,.
\end{aligned} \tag{20}$$

By symmetry of the final expression,

$$(A, LB) = (LA, B), \tag{21}$$

so L is self-adjoint with respect to $(\cdot,\cdot)$.

4.4 Positivity and the zero mode

Setting $A = B$ in (20) yields

$$(A, LA) = \big\langle(\partial_y^a A)(\partial_y^a A)\big\rangle \geq 0, \tag{22}$$

so L is positive semidefinite.

Moreover, constants are a zero mode:

$$L\mathbb{1} = 0. \tag{23}$$

Hence L is invertible on the subspace orthogonal to constants, i.e. those A with $\langle A\rangle = 0$. Equation (19) is solvable because $\langle \mu - O\rangle = 0$ by definition of μ.

ML remark. The identity (20) is useful for ML: it gives a natural quadratic form

$$(A, LA) = \big\langle \|\partial A\|^2 \big\rangle\,, \tag{24}$$

and suggests loss functions based on the PDE residual measured in the Gibbs inner product.

5 Strong-coupling expansion in β for $O(3)$ two-point functions

We now specialize to $O[n] = n_y \cdot n_z$ with $|y - z| > 1$.

Write $S = \beta S_0$ and expand

$$U = U_0 + \beta U_1 + \mathcal{O}(\beta^2), \qquad \mu = \mu_0 + \beta\mu_1 + \mathcal{O}(\beta^2). \tag{25}$$

At $\beta = 0$, spins are i.i.d. uniform on S^2, so for $y \neq z$,

$$\mu_0 = \langle n_y \cdot n_z \rangle_{\beta=0} = 0. \tag{26}$$

For $|y - z| > 1$ one also has $\mu_1 = 0$ (the first nonzero contribution to the correlator appears at higher order).

Define

$$BU \equiv (\partial_y^a S_0)(\partial_y^a U), \tag{27}$$

so the PDE (11) reads

$$(\partial^2 - \beta B)\, U = O - \mu. \tag{28}$$

5.1 Zeroth order: U_0

At $\mathcal{O}(\beta^0)$:

$$\partial^2 U_0 = O. \tag{29}$$

Using (7),

$$\partial^2(n_y \cdot n_z) = (\partial_y^2 + \partial_z^2)(n_y \cdot n_z) = (-2-2)(n_y \cdot n_z) = -4(n_y \cdot n_z), \tag{30}$$

hence a particular solution is

$$U_0 = -\frac{1}{4}(n_y \cdot n_z). \tag{31}$$

5.2 First order: closure under ∂^2 and explicit U_1

At $\mathcal{O}(\beta^1)$:

$$\partial^2 U_1 = BU_0 - \mu_1 = BU_0, \qquad (|y - z| > 1). \tag{32}$$

Derivatives of S_0 and U_0. From (1) and (5),

$$\partial_y^a S_0 = -\sum_{w \in \mathrm{nn}(y)} \partial_y^a (n_y \cdot n_w) = -\sum_{w \in \mathrm{nn}(y)} (\lambda^a n_y) \cdot n_w. \tag{33}$$

From (31),

$$\partial_y^a U_0 = -\frac{1}{4}\,(\lambda^a n_y) \cdot n_z, \qquad \partial_z^a U_0 = -\frac{1}{4}\,(\lambda^a n_z) \cdot n_y. \tag{34}$$

Therefore only sites y and z contribute to BU_0:

$$BU_0 = (\partial_y^a S_0)(\partial_y^a U_0) + (\partial_z^a S_0)(\partial_z^a U_0). \tag{35}$$

Using (33), (34), and (9) yields

$$\begin{aligned}(\partial_y^a S_0)(\partial_y^a U_0) &= \frac{1}{4} \sum_{w \in \mathrm{nn}(y)} \sum_{a=1}^{3} \left[(\lambda^a n_y) \cdot n_w\right] \left[(\lambda^a n_y) \cdot n_z\right] \\ &= \frac{1}{4} \sum_{w \in \mathrm{nn}(y)} \Big[(n_w \cdot n_z) - (n_y \cdot n_w)(n_y \cdot n_z)\Big],\end{aligned} \tag{36}$$

and similarly with $y \leftrightarrow z$.

Closure under ∂^2. Fix a neighbor $w \in \mathrm{nn}(y)$ (with $|y-z|>1$ implying $w \neq z$). Define two monomials

$$f_1 \equiv (n_w \cdot n_z), \qquad f_2 \equiv (n_y \cdot n_w)(n_y \cdot n_z). \tag{37}$$

Using (7) and (8) one obtains

$$\partial^2 f_1 = -4 f_1, \qquad \partial^2 f_2 = 2 f_1 - 10 f_2. \tag{38}$$

Thus $\mathrm{span}\{f_1, f_2\}$ is closed. The eigen-combination

$$g_2 \equiv f_2 - \frac{1}{3} f_1 \tag{39}$$

satisfies $\partial^2 g_2 = -10 g_2$, while f_1 satisfies $\partial^2 f_1 = -4 f_1$.

Solving (32) in the closed subspace. The y-side RHS contribution per neighbor $w \in \mathrm{nn}(y)$ is

$$R_{y,w} = \frac{1}{4}(f_1 - f_2) = \frac{1}{4}\left(\frac{2}{3} f_1 - g_2\right). \tag{40}$$

Inverting ∂^2 mode-by-mode gives

$$U_{1,(y,w)} = -\frac{1}{20} f_1 + \frac{1}{40} f_2 = -\frac{1}{20}(n_w \cdot n_z) + \frac{1}{40}(n_y \cdot n_w)(n_y \cdot n_z). \tag{41}$$

Adding the symmetric z-side contributions yields, for $|y-z|>1$,

$$\begin{aligned} U_1 = &\sum_{w \in \mathrm{nn}(y)} \left[-\frac{1}{20}(n_w \cdot n_z) + \frac{1}{40}(n_y \cdot n_w)(n_y \cdot n_z)\right] \\ &+ \sum_{w \in \mathrm{nn}(z)} \left[-\frac{1}{20}(n_w \cdot n_y) + \frac{1}{40}(n_z \cdot n_w)(n_y \cdot n_z)\right]. \end{aligned} \tag{42}$$

Control variate through $\mathcal{O}(\beta)$. With $U \approx U_0 + \beta U_1$, define

$$F_U = \partial^2 U - (\partial_y^a U)(\partial_y^a S), \tag{43}$$

and the improved estimator $\widehat{O}_U = O - F_U$. By (14), $\langle F_U \rangle = 0$ exactly (no approximation), so $\widehat{O}_U$ remains unbiased for any approximate U.

6 Symmetry-respecting parametrizations of U for ML

We want a trainable approximation $U_\theta[n]$ that: (i) is translation invariant (for $C(r)$), (ii) is $O(3)$ invariant, (iii) is local/quasi-local for efficiency, (iv) generalizes to G/H.

6.1 (A) Translation-invariant local density + invariant features

Use a shared-weights local density (convolutional ansatz)

$$U_\theta[n] = \sum_y u_\theta(\mathcal{N}_R(y); r)\,, \tag{44}$$

where $\mathcal{N}_R(y)$ is a radius-R neighborhood of y and r is the displacement in $C(r)$.

Impose $O(3)$ invariance by feeding u_θ only invariants, e.g. dot products

$$X_y^{(\delta)} = n_y \cdot n_{y+\delta}, \qquad Y_y^{(\delta)}(r) = n_y \cdot n_{y+r+\delta}, \tag{45}$$

for a chosen stencil δ (nearest neighbors, next-to-nearest, etc). Then u_θ can be an ordinary MLP/CNN acting on scalar channels.

6.2 (B) CNN on invariant channels

Construct multi-channel scalar fields $\{X_y^{(\delta)}, Y_y^{(\delta)}(r)\}$ and feed them into a standard CNN (with shared kernels), producing an output field $\rho_\theta(y)$ and set

$$U_\theta[n] = \sum_y \rho_\theta(y). \tag{46}$$

This enforces translation invariance and $O(3)$ invariance by construction.

6.3 (C) Residual learning around the perturbative solution

Use the explicit baseline

$$U^{(0+1)} \equiv U_0 + \beta U_1, \tag{47}$$

and train only a correction:

$$U_\theta = U^{(0+1)} + \Delta U_\theta, \tag{48}$$

with ΔU_θ constrained as in (A)–(B). This typically stabilizes training and keeps the model in the right symmetry class.

6.4 (D) Losses based on the self-adjoint inner product

Since L is self-adjoint in (16), natural losses are:

(i) PDE residual in the Gibbs inner product. Let the residual be

$$R_\theta[n] \equiv \partial^2 U_\theta - (\partial_y^a U_\theta)(\partial_y^a S) - (O - \widehat{\mu}), \tag{49}$$

with $\widehat{\mu}$ an empirical estimate of μ. Minimize

$$\mathcal{L}_{\text{PDE}}(\theta) = (R_\theta, R_\theta) = \left\langle R_\theta^2 \right\rangle. \tag{50}$$

(ii) Variance minimization of the improved estimator. Define F_θ from (10) and minimize the empirical variance of

$$\widehat{O}_\theta[n] = O[n] - F_\theta[n]. \tag{51}$$

(iii) Dirichlet form regularization. Using (22), one may add

$$\mathcal{R}(\theta) = (U_\theta, LU_\theta) = \left\langle (\partial_y^a U_\theta)^2 \right\rangle \tag{52}$$

as a smoothness/complexity control.

7 Outlook: extension to G/H and principal chiral models

The Lie-derivative formulation generalizes directly: replace $n_y \in S^2$ by fields valued in G/H (or G), replace ∂_y^a by the corresponding left-invariant derivatives (and for cosets, the projected derivatives), and replace dot-product invariants by group invariants such as $\mathrm{Re}\,\mathrm{Tr}(g_y^\dagger g_w)$ (principal chiral) or appropriate coset invariants. The low-order closure/eigenspace strategy is then governed by the representation theory of the site-wise quadratic Casimir.