Equivariant Flows: sampling configurations for multi-body systems with symmetric energies

Jonas Köhler [*†]   **Leon Klein** [*†]   **Frank Noé** [†‡§]

[†] Freie Universität Berlin, Department of Mathematics and Computer Science.
[‡] Freie Universität Berlin, Department of Physics.
[§] Rice University, Department of Chemistry.
{jonas.koehler, leon.klein, frank.noe}@fu-berlin.de
[*] Authors contributed equally to this work.

Abstract

Flows are exact-likelihood generative neural networks that transform samples from a simple prior distribution to the samples of the probability distribution of interest. Boltzmann Generators (BG) combine flows and statistical mechanics to sample equilibrium states of strongly interacting many-body systems such as proteins with 1000 atoms. In order to scale and generalize these results, it is essential that the natural symmetries of the probability density – in physics defined by the invariances of the energy function – are built into the flow. Here we develop theoretical tools for constructing such equivariant flows and demonstrate that a BG that is equivariant with respect to rotations and particle permutations can generalize to sampling nontrivially new configurations where a nonequivariant BG cannot.

1   Introduction

Generative learning using exact-likelihood methods based on invertible transformations has had remarkable success in domains like accurately representing images [5, 8, 9], audio [17], 3D point cloud data [11, 12] as well as applications in physics [1, 12, 16].

Recently, *Boltzmann Generators* (BG) [12] have been introduced for sampling Boltzmann type distributions $p(x) \propto \exp(-u(x))$ of high-dimensional multi-body problems, such as valid conformations of proteins. In contrast to typical generative learning problems, the target density $p(x)$ is specified by definition of the multi-body energy function $u(x)$ and the difficulty lies in learning to sample it efficiently. BGs do that in two steps: (1) An exact-likelihood method that generates samples from a density $q_{X,\theta}(x)$ that approximates the Boltzmann density $p(x)$. (2) An algorithm to reweigh the generated density to the target density $p(x)$.

In [12], it has been demonstrated that such BGs can be trained to efficiently sample prototypical many-body systems such as proteins in implicit solvent and condensed matter systems. In order to make further progress, it is essential to develop exact-likelihood generative models that respect the symmetries of $u(x)$, for example invariance of the energy with respect to global rotation or permutations of identical particles.

In this work, we derive a sufficient criterion to design symmetry invariant BGs and design a simple instance that is symmetric with respect to translations, rotations and permutations. By comparing this model with the non-symmetric approach based on *RealNVP* layers [5] as used in [12], we show that encoding symmetries in the model is critical for generalization.

A Hamiltonian flow framework for incorporating equivariances that can be seen as a special case of our equivariant flow construction was developed in parallel in [15]. In contrast to the present paper,

[15] does not use equivariant flows as part of an exact likelihood model but in a variational inference framework.

2   Boltzmann-generating flows

In order to reweigh the generated density $q_{X,\theta}(x)$ to the target density $p_X(x) \propto \exp(-u(x))$ (step 2 above), BGs require an exact likelihood generative model. This can be achieved by transforming a simple prior density $q_Z(z)$, e.g. a multivariate normal distribution, via an invertible function f_θ [2, 4–7, 10, 13, 14]. Sampling from $q_{X,\theta}(x)$ is achieved by sampling $z \sim q_Z(z)$ and transforming to $f_\theta(z) \sim q_{X,\theta}(x)$. Due to the invertibility of f_θ, the probability density of any generated point can be computed with the *change of variables* equation

$$q_{X,\theta}(x) = q_Z \left(f_\theta^{-1}(x) \right) \left| \det \frac{\partial f_\theta^{-1}(x)}{\partial x} \right|.$$

Two obvious options are available in order to train the generator to match $q_{X,\theta} \approx p_X$ [1, 12]:

1. *ML-training*: If some data $\{x_n\}_{n=1\ldots N}$ is given that at least represents one or a few high-probability modes of $p(x)$, we can maximize the likelihood under the model, as is typically done when training flow-based models for images.

2. *KL-training* We minimize the reverse Kullback-Leibler divergence $KL(q_{X,\theta}\|p_X)$. This approach is also known as energy-based training where the energy corresponding to the generated density is trained to match $u(x)$.

3   Equivariant Flows and Boltzmann Generators

3.1   Equivariant normalizing flows yield symmetric densities

Symmetries may be discussed in terms of a *group G* acting on a vector space $V \cong \mathbb{R}^n$. A *representation ρ* of G is a map $\rho\colon G \to GL(V)$ satisfying $\rho(g)\rho(h) = \rho(gh)$. Examples are permutations and rotations in 3D which can be represented by permutation and rotation matrices in $GL(V)$.

We call any map $f\colon V \to V'$ *G-invariant*, iff $f(\rho(g)x) = f(x)$ for all g and x. We further call a map $f\colon V \to V$ *G-equivariant*, iff $f(\rho(g)x) = \rho(g)f(x)$ for all g and x. In that sense an energy $u\colon \mathbb{R}^n \to \mathbb{R}$ is invariant with respect to a symmetry, if there is a symmetry group G, a representation ρ of the group in the conformation space, and the energy satisfies $u(\rho(g)x) = u(x)$.

Our first result is the following theorem, which gives a sufficient criterion for when a transformed density resulting from a normalizing flow is symmetric (proof in Appendix):

Theorem 1. *Let $q_Z(z)$ be a G-invariant prior density and f be an G-equivariant bijection. Let $q_{X,\theta}(x)$ be the density of $x = f(z)$ for $z \sim q_Z(z)$. Then $q_{X,\theta}$ is G-invariant.*

3.2   Equivariant dynamics yield equivariant continuous normalizing flows

Continuous normalizing flows (CNFs) [3] have recently been proposed as a way to describe the bijection implicitly via solving the initial value problem

$$\frac{\partial y(t)}{\partial t} = f(y(t), t), \qquad y(t_0) = z. \tag{1}$$

for a fixed time horizon $[t_0, t_1]$, where f can be a freely chosen / learned dynamics function. The solution of this initial value problem is $x := y(t_1) = y(t_0) + \int_{t_0}^{t_1} f(y(t), t)dt$, which (under mild constraints) gives a bijection $z \to x$.

As shown in [3], the change of density can be computed at any integration time, using the *continuous change of variable* rule, given by the differential equation

$$\frac{\partial \log p(y(t))}{\partial t} = -\mathrm{tr}\left[\frac{\partial f(y(t), t)}{\partial y(t)} \right] = -\operatorname{div} f(y(t), t), \tag{2}$$

$$\log p(y(t_0)) = \log p(z).$$

Integrating eqs. (1) & (2), e.g. using a black-box solver, yields an exact-likelihood generative model (within discretization error of the solver).

Our second result is that using such CNFs, 1 can be realized (proof in Appendix):

Theorem 2. *Let f be a G-equivariant dynamics function. If $q_Z(z)$ is a G-invariant prior density, z is sampled from $q_Z(z)$ and x is obtained from z by solving* (1)*. Then x is distributed according to a G-invariant density.*

3.3 Modeling symmetric multi-body systems with normalizing flows

In this work we concentrate on modeling the equilibrium distribution of multi-body systems of K particles $x = (x_1, \ldots, x_K) \in \mathbb{R}^{K \times D}$ ($D = 2, 3$), interacting with each other via a potential energy function $u(x_1, \ldots, x_K)$. The equilibrium distribution of such a system is given by a Boltzmann-type distribution $p(x) \propto \exp(-u(x))$ with an a priori defined energy $u(x)$. $u(x)$ is typically invariant with respect to certain transformations, here we assume the invariances of a molecule described by quantum mechanics without external field:

1. Permutation invariance: Swapping the labels of any two interchangeable particles $(x_1, \ldots, x_K) \rightarrow (x_{\sigma(1)}, \ldots, x_{\sigma(K)})$.

2. Rotation invariance: Any 2D/3D rotation of the system $Rx = (Rx_1, \ldots, Rx_K)$.

3. Translation invariance: Any 2D/3D translation $x + v = (x_1 + v, \ldots, x_K + v)$.

A natural prior distribution $q_Z(z)$ being invariant to all these symmetries and thus satisfying the conditions of thm. 1 can be obtained by sampling $z' = (z'_1, \ldots, z'_N) \sim \mathcal{N}(0, I_{D \times D})$ i.i.d. and subtract its center of mass $\mu(x) := \sum_{i=1}^{K} x_i$ yielding $z = z' - \mu(z')$. For a mean-free configuration z, we can evaluate its likelihood by $\log q_Z(z) = -1/2 \sum_i^K \|z_i\|^2 + const.$, which is invariant with respect to the symmetries 1.-3. above.

3.4 Equivariant flows for multi-body systems

We give a simple example of an equivariant flow which obeys the symmetries 1.-3. above. For this, we employ thms. 1 & 2 to design equivariant dynamics functions with tractable exact traces.

Each particle is updated using a vector field in radial direction

$$\frac{\partial x_i(t)}{\partial t} = f_\theta(x(t), t)_i := \sum_{j=1}^{K} \psi_\theta(\|x_i - x_j\|_2)(x_i - x_j) \tag{3}$$

where $\psi_\theta \colon \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ can be an arbitrary scalar function parameterized by θ. These dynamics correspond to particle updates due to a conservative force generated by a potential that depends on particle distances. The divergence is then given by

$$\text{div } f_\theta(x(t), t) = \sum_{i=1}^{K} \sum_{j=1}^{K} \frac{\partial \psi_\theta(\|x_i - x_j\|_2)}{\partial \|x_i - x_j\|_2} \|x_i - x_j\|_2 + D \sum_{i=1}^{K} \sum_{j=1}^{K} \psi_\theta(\|x_i - x_j\|_2), \tag{4}$$

which can be computed exactly in an efficient way using one backpropagation pass through the graph.

It is simple to see that this vector field is equivariant with respect to symmetries 1.-3. Further this field does not alter the center of mass of a particle system, thus integrating mean-free states along it will always result in mean-free states.

4 Experiments

Using a simple 2D toy system, we show empirically, that a Boltzmann Generator using an equivariant flow encoding the symmetries of the system's energy function explicitly as given in the last section (*eqBG*) can generalize beyond the training data whereas a non-equivariant approach (furthermore abbreviated as *nBG*) based on RealNVP transformations cannot.
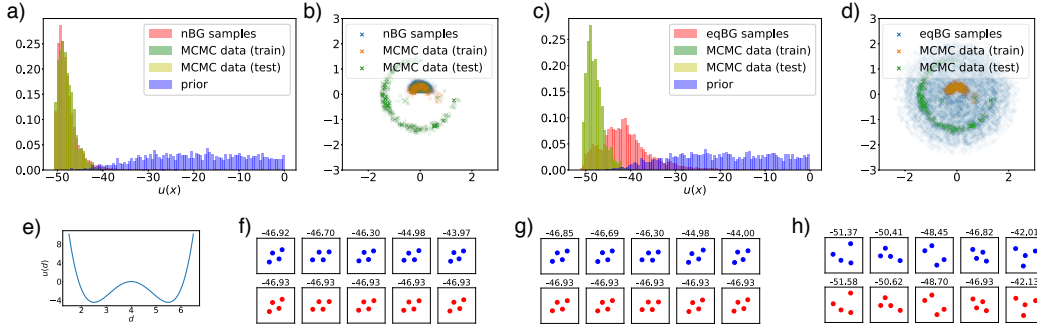
Figure 1: **Comparison: sampling equivariant and non-equivariant Boltzmann Generators**. **a, c)** Energies for samples taken from trained nBG and eqBG respectively. **b, d)** Marginal distribution of the first particle position x_1 for states sampled from trained nBG and eqBG respectively. **e)** Potential energy of the system illustrated for $N = 2$. **f-h)** *Top row:* sampled states from training data, trained nBG and trained eqBG respectively together with corresponding energies. *Bottom row:* corresponding configurations after energy minimization.

Our system ($N = 4, D = 2$) is given by the pairwise potential

$$u(x) = \sum_{i=1,j=1,j\neq i}^{N} -4 \left(\|x_i - x_j\|_2 - 4 \right)^2 + 0.9 \left(\|x_i - x_j\|_2 - 4 \right)^4,$$

which produces two distinct low energy modes separated by a high energy barrier (see Figure 1 e). By coupling multiple particles with such double-well interactions we can create a frustrated system with multiple metastable states.

4.1 First experiment: generalization on unseen trajectories

In a first experiment, we compare how both models generalize to unseen trajectories. For this we use standard MCMC sampling using the classic Metropolis-Hasting algorithm to produce short out-of-equilibrium trajectories of the system. We split a trajectory using the first 50% of samples for initializing both, the nBG and the eqBG, with ML training. After that we fine-tune both models on the energy using a weighted sum of ML and KL training. Both models are able to sample states with acceptable energies (Figure 1 a, c). However, evaluating both models on the second 50% of the trajectory, we clearly see that the nBG is not generalizing at all, whereas the eqBG is able to achieve similar likelihoods on train and test set simultaneously (Figure 1 b, d, Table 1).

Table 1: Negative log-likelihood of trained equivariant and non-equivariant Boltzmann Generators.

nBG (train)	nBG (test)	eqBG (train)	eqBG (test)
-12.64	372.69	5.91	6.72

4.2 Second experiment: discovery of unseen metastable states

In a second experiment, we evaluate to which extend both models are able to discover new metastable states, which have not been observed in the data set. Here we (1) create a training set, by perturbing a single minimum state with a tiny amount of Gaussian noise (Figure 1 f), (2) initialize both, the nBG and eqBG, with ML training on the training set, (3) fine-tune both models using a weighted sum ML and KL training on the energy, (4) generate samples from the converged models (Figure 1 g,h, top row) and (5) locally minimize the energies $u(x)$ of the generated structures x using a non-momentum second-order optimizer (Figure 1 g, h, bottom row). As can be seen, the nBG will only reproduce the unique state from which the training set was constructed. However, the eqBG will produce a variety of local minima, which did not appear in the training set.

4

5 Discussion

We gave a sufficient criterion, by which equivariant flows can be constructed and implemented an instance for symmetries frequently appearing in multi-particle systems. Using a simple example we have demonstrated that encoding symmetries of the target density/energy into the invertible function is critical, if we expect generative flows such as BGs to generalize beyond seen trajectories. Scaling such equivariant flows and BGs to large real-world systems remains a major challenge for future research.

References

[1] MS Albergo, G Kanwar, and PE Shanahan. Flow-based generative models for markov chain monte carlo in lattice field theory. *arXiv preprint arXiv:1904.12072*, 2019.

[2] Jens Behrmann, David Duvenaud, and Jörn-Henrik Jacobsen. Invertible residual networks. *CoRR*, abs/1811.00995, 2018.

[3] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, pages 6571–6583, 2018.

[4] L. Dinh, D. Krueger, and Y. Bengio. Nice: Nonlinear independent components estimation. *arXiv:1410.8516*, 2015.

[5] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

[6] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. *arXiv preprint arXiv:1906.04032*, 2019.

[7] Will Grathwohl, Ricky TQ Chen, Jesse Betterncourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*, 2018.

[8] Jonathan Ho, Xi Chen, Aravind Srinivas, Yan Duan, and Pieter Abbeel. Flow++: Improving flow-based generative models with variational dequantization and architecture design. *arXiv preprint arXiv:1902.00275*, 2019.

[9] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 10215–10224. Curran Associates, Inc., 2018.

[10] Durk P Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improved variational inference with inverse autoregressive flow. In *Advances in neural information processing systems*, pages 4743–4751, 2016.

[11] Xingyu Liu, Charles R. Qi, and Leonidas J. Guibas. Flownet3d: Learning scene flow in 3d point clouds. *arXiv:1806.01411*, 2019.

[12] Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators-sampling equilibrium states of many-body systems with deep learning. *Science*, 365:eaaw1147, 2019.

[13] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked autoregressive flow for density estimation. In *Advances in Neural Information Processing Systems*, pages 2338–2347, 2017.

[14] D. J. Rezende and S. Mohamed. Variational inference with normalizing flows. *arXiv:1505.05770*, 2015.

[15] Danilo Jimenez Rezende, Sébastien Racanière, Irina Higgins, and Peter Toth. Equivariant hamiltonian flows. *arXiv preprint arXiv:1909.13739*, 2019.

[16] Or Sharir, Yoav Levine, Noam Wies, Giuseppe Carleo, and Amnon Shashua. Deep autoregressive models for the efficient variational simulation of many-body quantum systems. *arXiv preprint arXiv:1902.04057*, 2019.

[17] A. van den Oord, Y. Li, I. Babuschkin, K. Simonyan, O. Vinyals, K. Kavukcuoglu, G. van den Driessche, E. Lockhart, L. C. Cobo, F. Stimberg, N. Casagrande, D. Grewe, S. Noury, S. Dieleman, E. Elsen, N. Kalchbrenner, H. Zen, A. Graves, H. King, T. Walters, D. Belov, and D. Hassabis. Parallel WaveNet: Fast High-Fidelity Speech Synthesis. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 3918–3926, Stockholmsmässan, Stockholm Sweden, 2018. PMLR.

A Proof of thm. 1

Let $V = \mathbb{R}^n$ and $q\colon V \to \mathbb{R}_{\geq 0}$ be a probability density on V. Let G be a group acting on V and let $\rho\colon G \to GL(V)$ be a representation of G over V. As V is finite-dimensional every $\rho(g)$ is represented by a matrix and thus $\det \rho(g)$ is well-defined. Furthermore, for a function $f \in C^1(\mathbb{R}^n, \mathbb{R}^m)$ let $J_f(x) \in \mathbb{R}^{n \times m}$ denote its Jacobian evaluated at x and define the push-forward density of q along a diffeomorphism $f \in C^1(V, V)$ by $q_f(x) := q(f^{-1}(x)) \left| \det J_{f^{-1}}(x) \right|$.

Lemma 1. *Let* $A \in GL(V)$, *if* $q(Ax) = q(x)$ *for all* $x \in V$, *then* $\det A \in \{-1, 1\}$

Proof. Set $h(x) = Ax$. By substituting $z = A^{-1}x$ we get

$$1 = \int_V q(x)dx \tag{5}$$

$$= \int_{h^{-1}(V)} q(h(z)) \left| \det A \right| dz \tag{6}$$

$$= \int_V q(z) \left| \det A \right| dz \tag{7}$$

$$= \left| \det A \right| \underbrace{\int_V q(z)dz}_{=1} \tag{8}$$

$$= \left| \det A \right| \tag{9}$$

$\square$

From Lemma 1 we get $\det \rho(g) \in \{-1, 1\}$ for each $g \in G$. Let $g \in G$. Define the transformation $T_g\colon V \to V$, $x \mapsto \rho(g)x$. If $f \in C^1(V, V)$ is G-equivariant, it means $f \circ T_g = T_g \circ f$ for each $g \in G$. If q is an G-invariant density it means $q \circ T_g = q$. Together with the lemma we obtain

$$q_f(\rho(g)x) = q_f(T_g(x)) \tag{10}$$

$$= q_f(T_g(x)) \underbrace{\left| \det J_{T_g}(x) \right|}_{=|\det \rho(g)|=1} \tag{11}$$

$$= q_{T_{g^{-1}} \circ f}(x) \tag{12}$$

$$= q_{f \circ T_{g^{-1}}}(x) \tag{13}$$

$$= q((T_g \circ f^{-1})(x)) \left| \det J_{T_g \circ f^{-1}}(x) \right| \tag{14}$$

$$= (q \circ T_g \circ f^{-1})(x) \left| \det J_{T_g}(f^{-1}(x)) J_{f^{-1}}(x) \right| \tag{15}$$

$$= (q \circ f^{-1})(x) \left| \det J_{T_g}(f^{-1}(x)) \right| \left| \det J_{f^{-1}}(x) \right| \tag{16}$$

$$= q(f^{-1}(x)) \left| \det \rho(g) \right| \left| \det J_{f^{-1}}(x) \right| \tag{17}$$

$$= q(f^{-1}(x)) \left| \det J_{f^{-1}}(x) \right| \tag{18}$$

$$= q_f(x) \tag{19}$$

B Proof of thm. 2

Let $g \in G$ and $\rho(g)$ be its representation. Let f be an G-equivariant dynamics function. Then

$$\rho(g)y(t_0) + \int_{t_0}^{t_1} f(\rho(g)y(t), t) \, dt = \rho(g)y(t_0) + \int_{t_0}^{t_1} \rho(g)f(y(t), t) \, dt \tag{20}$$

$$= \rho(g) \left(y(t_0) + \int_{t_0}^{t_1} f(y(t), t) \, dt \right). \tag{21}$$

7

This implies that the bijection given by solving

$$y(t_0) = z \tag{22}$$

$$\frac{\partial y(t)}{\partial t} = f(y(t), t) \tag{23}$$

for the time interval $[t_0, t_1]$ is an equivariant map. Now we can use thm. 1.