



End-to-end learning and auto-differentiation: forces, uncertainties, observables, trajectories and scales.

IPAM Learning and Emergence in Molecular Systems LA Jan 24 2023 Rafael Gómez-Bombarelli rafagb@mit.edu http://gomezbombarelli.mit.edu/

Virtuous cycle for design



Quick thoughts on Emergence

Weak emergence:

- ~ Generalization
- ~ Inductive bias

train on a subset of the "phase space" and confirm that known larger scale / more collective effects show up

Actual emergence:

~ Discovery

observe unknown larger scale / more collective effects show up – trust them to be physical and not artifacts !

Autodiff, uncertainty, and ML potentials

Atomistic simulations

Molecular Dynamics

a simulation technique that uses classical (Newtonian) dynamics of many particles and statistical mechanics to understand the collective behavior of atoms and molecules

Essentially, propagates the motion of

- 100s to millions of atoms
- 1 to ~5 femtoseconds at at time
- Many processes we care about take milliseconds or longer.
 The *ergodic principle* suggests we can trade simulation length with ensemble size

Matter typically involves Avogadro's number of particle (10²³)

- Given their interaction potential

$$\dot{\mathbf{q}} = rac{\partial H}{\partial \mathbf{p}} \qquad \dot{\mathbf{p}} = -rac{\partial H}{\partial \mathbf{q}}$$

Interactions between atoms are governed by Quantum Mechanics.
 Solving them has huge cost–accuracy tradeoffs. Surrogates are needed

We aspire to use ML to take on as many of these better as we can: replace QM, reduce particle count, enhance sampling



Using NN potentials for molecules

Excited State potentials

Multi-target neural networks that predict the energies (and forces) of multiple states

- Excited-state dynamics and reactivity
- Non-adiabatic dynamics and internal conversion in photo switches
- Diabatic Neural Networks learn smooth diabatic surfaces that diagonalize into the training-data adiabatic energies and forces
- Diazobenzene compounds have rich, fast photophysics with CI close to cis/trans transition state. Photopharmacological applications.





trans-azobenzene

cis-azobenzene





Energies, forces, and gaps

$$(E_0, E_1) = \begin{cases} (d_{00}, d_{11}), & \text{if } \vec{R} \in trans \\ (d_{22}, d_{00}), & \text{if } \vec{R} \in cis, \end{cases}$$

$$\left(\mathbf{U}^{\dagger}\left[\nabla_{R}\mathbf{H}_{d}\right]\mathbf{U}\right)_{nm} = \begin{cases} -\vec{f}_{n}, \text{ if } n = m\\ \vec{g}_{nm}, \text{ if } n \neq m. \end{cases}$$

640,000 geometries SF-TDDFT BHHLYP/6-31G* 8,000 unique azobenzenes

		E_0	E_1	ΔE_{01}	$(\Delta E_{01})_{\rm small}^{\rm a}$	$ec{F_0}$	$ec{F_1}$	$ec{g}_{01}$	
Seen species	MAE (\downarrow)	0.86	1.01	0.75	0.47	1.00	1.17	0.87	
	$R^2 (\uparrow)$	1.00	1.00	1.00	0.97	0.99	0.99	0.84	
Ungoon species	MAE (\downarrow)	3.06	3.77	1.89	0.97	1.72	2.31	1.36	
Unseen species	R^2 (\uparrow)	0.99	0.98	0.98	0.95	0.97	0.86	0.50	

LGPS Nat Commun 10, 5260 (2019)

Ion transport in solids

Crystalline inorganic solids are also potential alternative to liquid electrolytes

NN Potentials work just the same for crystals (just need periodic boundaries)

LGPS (Li₁₀GeP₂S₁₂) has very high ionic conductivity.

Experiments and simulations did not agree, and neither could bridge gap in observations

NN-MD demonstrates a switchover in the behavior of the material where one of the modes of Litransport shuts down at intermediate temperatures



Differentiable uncertainty

Uncertainty and active learning

Quantifying uncertainty in NNs is possible (kernel methos are naturals at this)

Mean-variance estimation

Monte-Carlo dropout

Ensembling

Evidential uncertainty

Every prediction (E, F) now comes with a standard deviation that – in principle – is correlated with error enabling Active Learning

How to find uncertain points to improve the model – we can call the oracle!

- •Run the simulation and pause when the model is uncertain, retrain (unpractical for NNs)
- Run the simulation and gather many frames where model is uncertain
- Have the full converged simulation already (not that much need for AL then?)



Adversarial Attacks on Interatomic potentials

MD to get training data is impractical

NN-MD is very unstable. Ground-truth MD is very expensive.

Gradient ascent attack on probability-weighted uncertainty for Active Learning



Adversarial Attacks on Interatomic potentials

Ensemble uncertainty on the forces

Likelihood based on visited points

[energy uncertainty is less correlated with error that force uncertainty with force error]

$$\sigma_F^2(X) = \frac{1}{M-1} \sum_{m=1}^M \left[\frac{1}{3n} \sum_{i,j} \|\hat{F}_{ij}^{(m)}(X) - \bar{F}_{ij}(X)\|^2 \right]$$

$$Q = \sum_{(X,E,\mathbf{F})\in\mathcal{D}} \exp\left(-\frac{E}{kT}\right)$$
$$p(X_{\delta}) = \frac{1}{Q} \exp\left(-\frac{\bar{E}(X_{\delta})}{kT}\right)$$

Displace atoms from starting frames towards region that maximize the likelihoodweighted uncertainty through gradient ascent

$$\max_{\delta} \mathcal{L}_{adv}(X,\delta;\theta) = \max_{\delta} p(X_{\delta}) \sigma_F^2(X_{\delta}) \qquad \delta^{(i+1)} = \delta^{(i)} + \alpha_{\delta} \frac{\partial \mathcal{L}_{adv}}{\partial \delta}$$

 $\gamma C \rho$

2D Well



Better data efficiency and more predictive potential

101

10²

Diffusion of organics in silica nanopores

On molecule-zeolite interactions

- 10% of the data produces more improvement than random



Rev-MD17 uncertainty



Uncertainty Attribution for condensed phase

Going from molecular to bulk, it is not enough to do adversarial sampling, we need to carve out uncertain environments: attribution

Attribution allows extracting only the uncertain local environments





Testing the potential – is this emergence?



fragments, confirmed generalization to large crystalline solids

Beyond Forces

Turns out we wanted good forces and energies to produce good MD. That was the end goal

Reporting how good an architecture does at reproducing (heldout) data is a necessary condition, but it is not sufficient.

The real question is whether I can run MD and produce accurate, rigorous and valuable statistics.

Fu et al, put together this amazing benchmark

Forces are not Enough: Benchmark and Critical Evaluation for Machine Learning Force Fields with Molecular Simulations





Differentiable Simulations

Differentiable MD simulations

Use Neural ODE setting to learn through time evolution or ensemble (ergodic principle) of MD simulation

Hamiltonian dynamics

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \qquad \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i},$$

Adjoint

1/24/2023

$$\begin{split} a(t) &= dL/d(p(t),q(t)) \\ \frac{dL}{d\theta} &= \int_{t_{i+1}}^{t_i} a(t) \frac{df(q(t),p(t),\theta)}{d\theta} dt \end{split}$$

The reverse-mode automatic differentiation computes the gradient through the adjoint states without backpropagating through the forward computations



Differentiable MD simulations



Reproduce Radial Distribution Functional

Learn interatomic potential that results in the atomistic distribution with the given RDF water 298K 1atm

O-O O-H and H-H



6

Å

More on Pair potentials

Can fit to multiple state points and composition at the same time, improving generalization.



Fitting CG RDF of water at with temperature dependence



Henderson theorem states there is a one-toone mapping between potential and g(r) but the fitting can be very sensitive



arXiv:2209.07679v1 Wang et al. 2022

DiffSim for barrier crossing

If the initial and final states are known. Set up two DiffSim trajectories from both ends

$$p_{\alpha} = P\left(\int_{t_0}^{t_e} \mathbb{1}_{\alpha}(\boldsymbol{x}(t)) \, \mathrm{d}t > 0 \ \middle| \ \boldsymbol{x}(t_0) \in W_{-\alpha}\right)$$
$$L = L_{\boldsymbol{\xi}_{\alpha}} = \begin{cases} 0 & \text{if } \exists_{t_0 < t < t_e} \ \boldsymbol{x}(t) \in W_{-\alpha} \\ \lim_{t_0 < t < t_e} d(\boldsymbol{\xi}(\boldsymbol{x}(t)), \boldsymbol{\xi}_{\alpha}) - \epsilon & \text{otherwise} \end{cases}$$











Sipka et al. **2023** arXiv:2301.03480

Issues and tools needed

Exploding / vanishing gradients

Slow and fast dynamics

Chaotic behavior (the longer the trajectory)

One update per trajectory

Partial backprop Detach gradients so DiffSim is Markovian: the bias potential B only depends on **x** and **n**

Mini batching

Langevin dynamics finite memory dynamics and therefore decaying adjoints



More examples

5D Muller-Brown

Ala2



Un- or semisupervised coarse-graining

Coarse Graining Auto-Encoding Framework



- AutoEncoder automatically coarse-grains atomistic coordinates to CG coordinates in a data-driven way
- Force matching also helps to shape the learning of CG and obtain $V_{CG}(z = E(x))$ for CG simulations

Automatic CG for small molecules









CG of molecular liquids



Ionic liquid

Temperature-transferable, long+short range NN potential recovers structure and kinetics



Self-diffusion activation energy

		E_a	$(kJ mol^{-1})$
Molecule	GT	T-NFF	NFF
Cation	20.7 ± 1.5	14.3 ± 1.6	10.4 ± 0.6
Anion	$21.6{\pm}2.9$	$16.7 {\pm} 1.8$	12.2 ± 0.5

Equivariant generative decoder

Information is always lost in CG – needs to be recovered statistically.

All atom to CG is surjective, a generative (nondeterministic) model is needed Create latent variable to hold info for decoding (depends on x and X at train and only on X during inference.

Equivariant decoding through inter-bead vectors.

FG Pooling &



Avoid FF refinement.

Recovers structure even in ultra-CG

Evaluate on alanine dipeptide and chignolin

Metrics are RMSD and fraction of chemical bonds missing + diversity of samples.

Baselines are equivariant linear decoder and non-equivariant NN decoder.



GenZProt



Internal coordinates



Generalizes and improves over CG-VAE

	D) (CD				and a second sec			
	RMSD				GED			
	PED00055	PED00090	PED00151	PED00218	PED00055	PED00090	PED00151	PED00218
(m1) Equiv. enc/prior +	0.433	0.537	0.540	0.499	0.004	0.008	0.004	0.001
Z-matrix dec (= GenZProt)	(0.001)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
(m2) Inv. enc/prior +	0.601	0.783	0.671	0.592	0.007	0.013	0.008	0.003
Z-matrix dec	(0.003)	(0.004)	(0.004)	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)
(m3) Equiv. enc/prior +	2.517	2.497	2.785	2.459	0.463	0.549	0.488	0.466
xyz dec	(0.000)	(0.001)	(0.007)	(0.003)	(0.001)	(0.001)	(0.002)	(0.004)
(m4) Inv. enc/prior +	2.091	2.190	2.490	2.113	0.280	0.339	0.329	0.247
xyz dec (= CGVAE)	(0.028)	(0.026)	(0.221)	(0.032)	(0.004)	(0.020)	(0.020)	(0.009)
(m5) GenZProt		-	0.723	-	-	-	0.027	
(PED00151)	-		(0.001)				(0.000)	-
(m6) CGVAE	-	-	2.062	-	-	-	0.155	
(PED00151)			(0.000)				(0.000)	-



Reconstruction vs Generation





Virtual discovery

Interplaying ML and simulations allows accelerated design of materials.

If the methods are there it's engineering

Sometimes one needs to make their own tools

Thanks!





IBM | Hii