Learning Coarse-Grained Models from Molecular Dynamics

Data-Driven and Embedded-Physics ML.

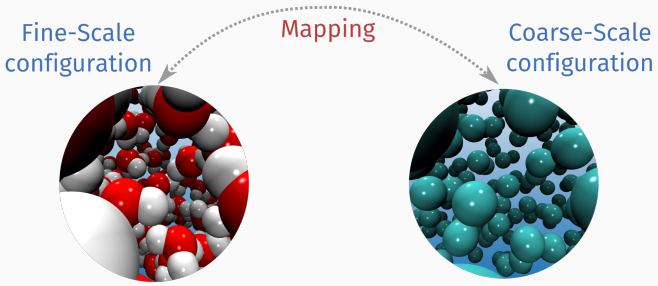
M. Schöberl[1,2], N. Zabaras[1], P.-S. Koutsourelakis[2]

SIAM CSE 2019, Spokane WA
February 27, 2019.

[1]Center for Informatics and Computational Science, University of Notre Dame, IN, USA.
[2]Continuum Mechanics Group, Technical University of Munich, Germany.

Coarse-graining and collective variable discovery

mschoeberl@gmail.com

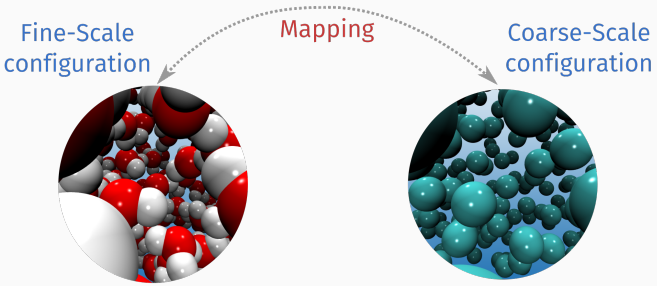


Why coarse-graining?

- Overcome spatiotemporal limitations.
- Reveal physical insight from reduced representation.

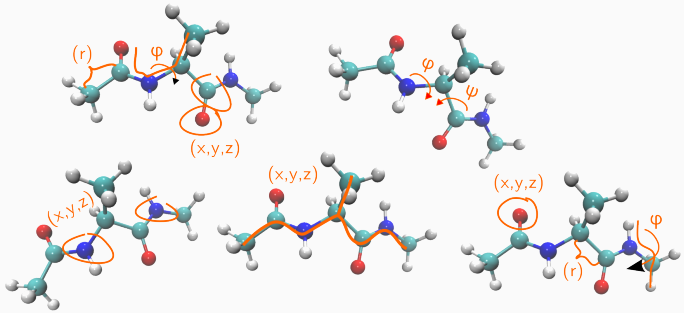
Coarse-graining and collective variable discovery

mschoeberl@gmail.com



Why coarse-graining?

- Overcome spatiotemporal limitations.
- Reveal physical insight from reduced representation.



Which set of collective variables (CVs) captures resilient and parsimonious features of atomistic systems?

- Relevant coordinates are highly clustered around a set of *lower dimensional collective variables.* [ban (2017), Chen and Ferguson (2017)]
- Vast combinatoric possibilities for *choosing* CVs. [Chakraborty et al. (2018)]

mschoeberl@gmail.com

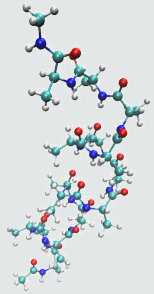
Questions we address

- How to learn a *predictive* coarse-grained representation
  - in the *small data regime* and
  - *in absence of* any data?
- How to identify good CVs?
- What are *good* CVs?
- Are identified CVs physically interpretable?

Problem definition - Equilibrium statistical mechanics

mschoeberl@gmail.com

Atomistic model



$$p_{\text{target}}(\mathbf{x}) \propto e^{-\beta U(\mathbf{x})}$$

- $\mathbf{x} \in \mathcal{M}$: atomistic coordinates
- $U(\mathbf{x})$: atomistic potential
- Observables: $\mathbb{E}_{p(\mathbf{x})}[a] = \int a(\mathbf{x})\, p(\mathbf{x})\, d\mathbf{x}$

Coarse-grained representation

$$\mathbf{z} = \mathcal{R}(\mathbf{x}), \quad dim(\mathbf{z}) << dim(\mathbf{x})$$

- $\mathbf{z}$: reduced CG / collective variables.
- $\mathcal{R}$: mapping operator (mapping to CG variables).

Overview - Coarse-graining

mschoeberl@gmail.com

Data-driven

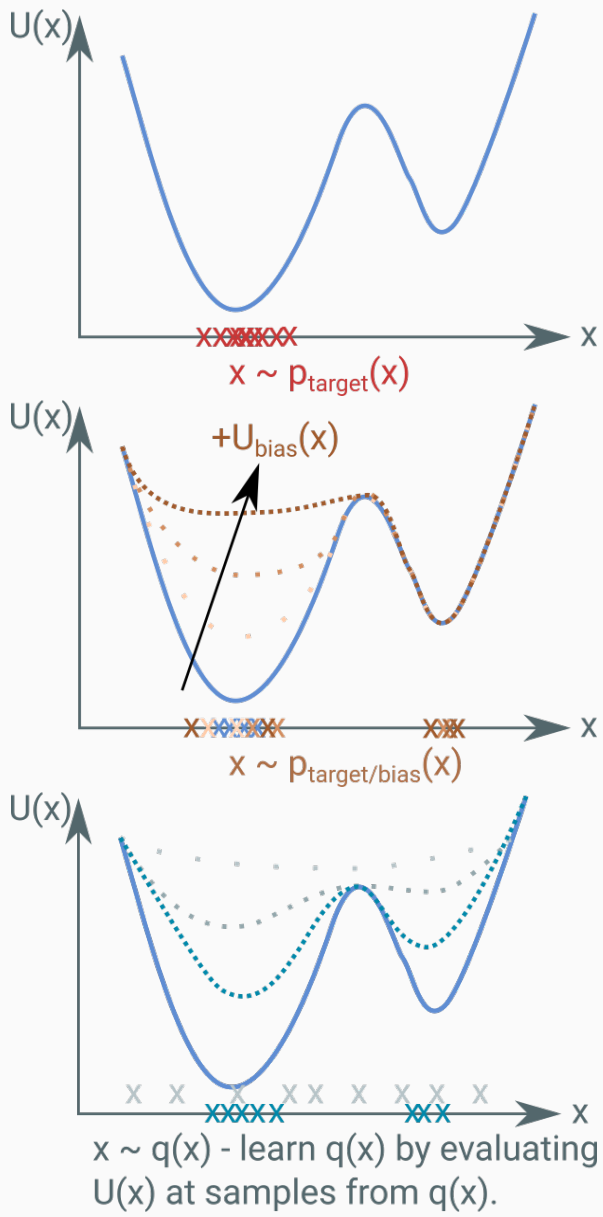
Simulate sufficiently long reference governing equation and obtain (limited) data approximating $p_{\text{target}}(x)$. [Shell (2008), Katsoulakis and Trashorras (2006), Trashorras and Tsagkarogiannis (2010), Noid (2013), Brunton et al. (2016), Wehmeyer and Noé (2018)]

Data-augmenting approach

E.g. enhanced sampling, learning based on consecutively gathered insight. [Laio and Parrinello (2002), Darve et al. (2008), Bilionis and Koutsourelakis (2012), Ferguson et al. (2011), Ferguson (2017), Chen and Tuckerman (2018)]

Embedded-Physics approach

Instead of simulating the target (based on known $U(x)$), directly incorporate physical constraints at our disposal, e.g. potential/force field. [Noé and Wu (2018)]



Data-Driven Coarse-Graining

Methodology - Probabilistic generative model

mschoeberl@gmail.com

Introduce $q(\mathbf{z})$ and $q(\mathbf{x}|\mathbf{z})$: $\underbrace{q(\mathbf{z})}_{\text{CV}} \xrightarrow{q(\mathbf{x}|\mathbf{z})} q(\mathbf{x}) = \int q(\mathbf{x}|\mathbf{z})\, q(\mathbf{z})\, d\mathbf{z}$ [Schöberl, Zabaras, and Koutsourelakis (2017)]

Which latent CVs z give rise to observations $x^{(i)}$?

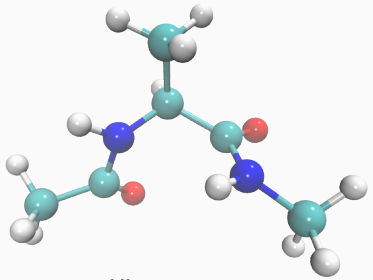
CV

Latent - - - - - - - - - - - - - - - - $\mathbf{z} \sim q(\mathbf{z})$

Generate / Decode

$\mathbf{x} \sim q(\mathbf{x}|\mathbf{z})$

Observed - - - - - - - - - - - -



Data $x^{(i)} \sim p_{\text{target}}(\mathbf{x})$

Methodology - Probabilistic generative model

mschoeberl@gmail.com

Introduce $q(\mathbf{z})$ and $q(\mathbf{x}|\mathbf{z})$: $\underbrace{q(\mathbf{z})}_{\text{CV}} \xrightarrow{q(\mathbf{x}|\mathbf{z})} q(\mathbf{x}) = \int q(\mathbf{x}|\mathbf{z})\, q(\mathbf{z})\, d\mathbf{z}$ [Schöberl, Zabaras, and Koutsourelakis (2017)]

Which latent CVs z give rise to observations $x^{(i)}$?

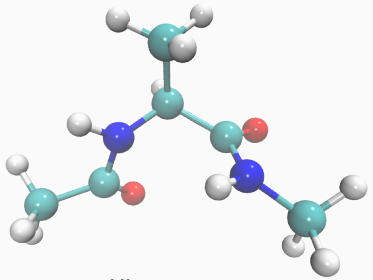
CV

Latent - - - - - - - - - - - - - - - - $\mathbf{z} \sim q(\mathbf{z})$

Generate / Decode

$\mathbf{x} \sim q(\mathbf{x}|\mathbf{z})$

Observed - - - - - - - - - - - -



Data $x^{(i)} \sim p_{\text{target}}(\mathbf{x})$

Methodology - Probabilistic generative model

mschoeberl@gmail.com

Introduce $q(\mathbf{z})$ and $q(\mathbf{x}|\mathbf{z})$: $\underbrace{q(\mathbf{z})}_{\text{CV}} \xrightarrow{q(\mathbf{x}|\mathbf{z})} q(\mathbf{x}) = \int q(\mathbf{x}|\mathbf{z})\, q(\mathbf{z})\, d\mathbf{z}$ [Schöberl, Zabaras, and Koutsourelakis (2017)]

Which latent CVs z give rise to observations $x^{(i)}$?

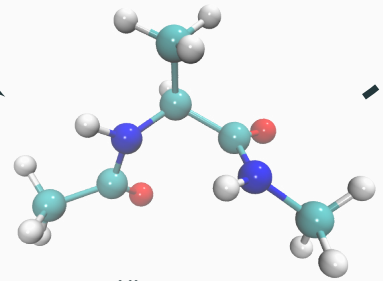
CV

Latent - - - - - - - - - - - - - - - - $\mathbf{z} \sim q(\mathbf{z})$

Generate / Decode

$\mathbf{x} \sim q(\mathbf{x}|\mathbf{z})$

Observed - - - - - - - - - - -

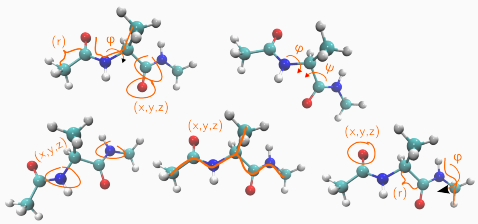


Data $x^{(i)} \sim p_{\text{target}}(\mathbf{x})$

Encode $\mathbf{x}$ with

$q_p(\mathbf{z}|\mathbf{x}) = \frac{q(\mathbf{x}|\mathbf{z})q(\mathbf{z})}{q(\mathbf{x})}$

Code / Dictionary



Methodology - Learning from data

mschoeberl@gmail.com

Parametrize $\underbrace{q(\mathsf{z}|\boldsymbol{\theta})}$ $\underbrace{q(\mathsf{x}|\mathsf{z};\ \boldsymbol{\theta})}$

Model for CVs. Probabilistic mapping given CVs to atomistic coordinates.

mschoeberl@gmail.com

Parametrize $\underbrace{q(\mathsf{z}|\boldsymbol{\theta})}$ $\underbrace{q(\mathsf{x}|\mathsf{z};\ \boldsymbol{\theta})}$

Model for CVs. Probabilistic mapping given CVs to atomistic coordinates.

Data-driven approach: Minimize

$$D_{\mathrm{KL}}\left(p_{\mathrm{target}}(\mathsf{x})||q(\mathsf{x})\right) = -\int p_{\mathrm{target}}(\mathsf{x}) \log \frac{q(\mathsf{x})}{p_{\mathrm{target}}(\mathsf{x})}\, d\mathsf{x}$$

equals maximizing marginal log-likelihood of the data set $\mathsf{X} = \{x^{(i)}\}_{i=1}^{N}$:

$$\log q(\mathsf{x}^{(i)}, \cdots, \mathsf{x}^{(N)}) = \sum_{i=1}^{N} \log q(x^{(i)}).$$

mschoeberl@gmail.com

Stochastic Variational Bayesian approximation

The log-likelihood is decomposed into: [Beal and Ghahramani (2006); Kingma and Welling (2013); Rezende et al. (2014)]

$$\log q_{\boldsymbol{\theta}}(x^{(i)}) = \underbrace{\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; x^{(i)})}_{\text{variational lower bound, since } D_{KL} \geq 0} + \underbrace{D_{KL}(r_{\boldsymbol{\phi}}(z^{(i)}|x^{(i)})||q_P(z^{(i)}|x^{(i)}))}_{\geq 0}$$

$$\log q_{\boldsymbol{\theta}}(x^{(i)}) \geq \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; x^{(i)}) = - \underbrace{D_{KL}(r_{\boldsymbol{\phi}}(z^{(i)}|x^{(i)})||q_{\boldsymbol{\theta}}(z))}_{\substack{\text{Regularize } \boldsymbol{\phi}, \text{ such that } r_{\boldsymbol{\phi}}(z^{(i)}|x^{(i)}) \\ \text{in aggregation is close to } q_{\boldsymbol{\theta}}(z).}} + \underbrace{\mathbb{E}_{r_{\boldsymbol{\phi}}(z^{(i)}|x^{(i)})}[\log q_{\boldsymbol{\theta}}(x^{(i)}|z^{(i)})]}_{\text{Expected neg. reconstruction error.}}$$

with the approximate posterior $r_{\boldsymbol{\phi}}(z^{(i)}|x^{(i)})$, e.g. a distribution of the exponential family and parameters $\boldsymbol{\phi}$.

Discovery of CVs as approximate Bayesian inference, i.e. identify the code/dictionary $r_{\boldsymbol{\phi}}(z|x)$. [Schöberl, Zabaras, and Koutsourelakis (2019)]

Methodology - CV discovery as VI

mschoeberl@gmail.com

- MLE estimate

$$\max_{\phi,\theta} \mathcal{L}(\theta, \phi; \mathbf{X})$$

- MAP estimate

$$\max_{\phi,\theta} \; \mathcal{L}(\theta, \phi; \mathbf{X}) + \underbrace{\log p(\theta)}_{log-prior}$$

Methodology - CV discovery as VI

mschoeberl@gmail.com

- MLE estimate

$$\max_{\phi,\theta} \mathcal{L}(\theta, \phi; X)$$

- MAP estimate

$$\max_{\phi,\theta} \mathcal{L}(\theta, \phi; X) + \underbrace{\log p(\theta)}_{log-prior}$$

- Approximate posterior of decoding parameters θ, $p(\theta|X)$, with Laplace approximation.

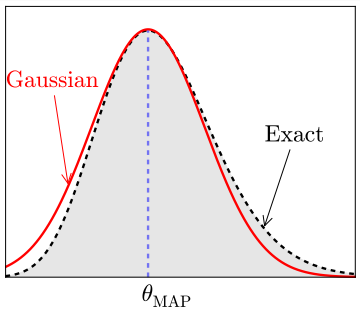


Figure 1: Laplace approximation: $q(\theta|X) \approx \mathcal{N}(\mu, S)$

- $q(\theta|X) \approx \mathcal{N}(\mu, S)$
- $\mu = \theta_{MAP}$
- $S^{-1} = -\frac{\partial^2 \mathcal{L}(\theta,\phi;X)}{\partial\theta_k \partial\theta_l} - \frac{\partial^2 \log p(\theta)}{\partial\theta_k \partial\theta_l}$

mschoeberl@gmail.com

Simulation details

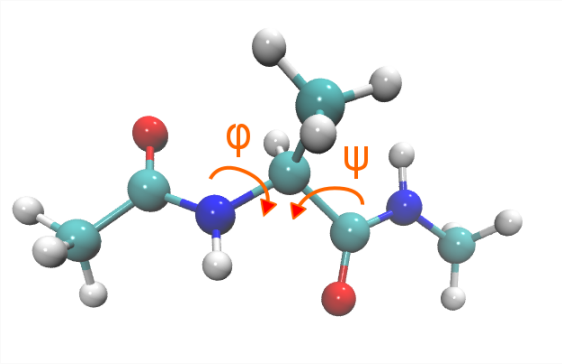


Figure 2: ALA-2 peptide with indicated dihedral angels.

- 22 atoms $\rightarrow$ $\mathbf{dim(x)} = 66$ in implicit solvent [Still et al. (1990)] .
- AMBER-FF96 force field [Salomon-Ferrer et al.] with Andersen thermostat at $T = 330\text{K}$ and $\Delta t = 1$ fs.
- Equilibration for 50 ns. Snapshots taken every 10 ps.
- *No* pre-processing of data the data: $\boldsymbol{x}^{(i)}$ are the Cartesian coordinates of *all* atoms in the system.

mschoeberl@gmail.com

Characteristic conformations of the ALA-2 peptide

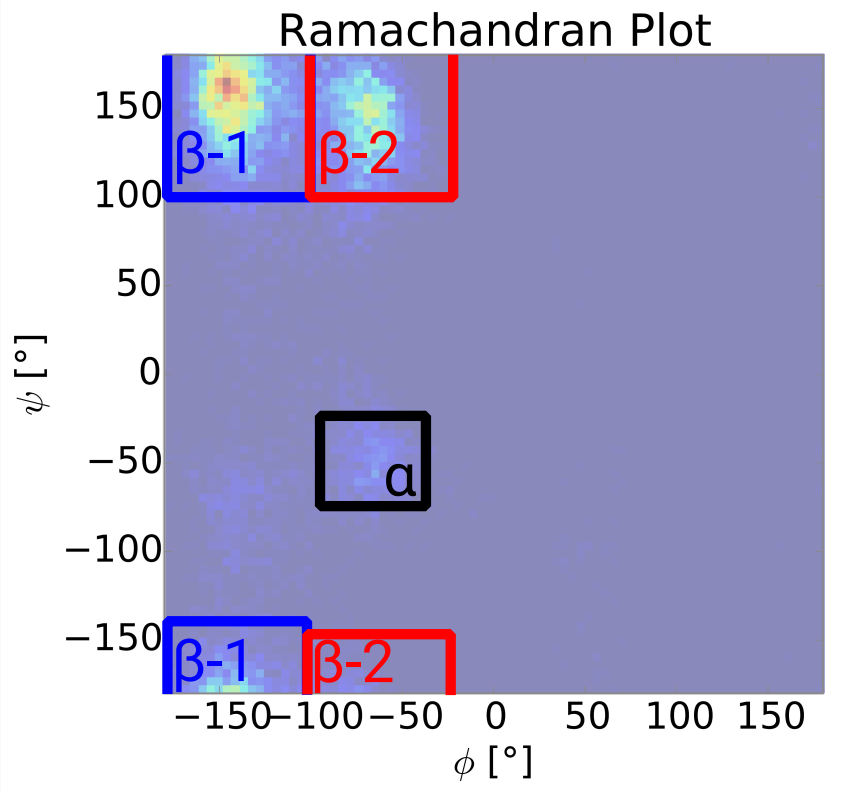
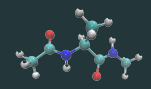


Figure 3: Characteristic conformations (α, β-1, β-2) and their labelling as used in the sequel.



mschoeberl@gmail.com

Auto-Encoding Variational Bayes

According [Kingma and Welling (2013)] .

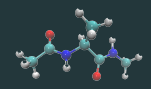
$$q(\mathsf{x}|\boldsymbol{\theta}) = \int q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta})\, q(\mathsf{z})\, d\mathsf{z}$$

Mapping $\mathsf{z} \to \mathsf{x}$ (decoder):

$$\boxed{q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}(\mathsf{z};\boldsymbol{\theta}),\ S_{\boldsymbol{\theta}} = \mathrm{diag}(\boldsymbol{\sigma}^2_{\boldsymbol{\theta}})).}$$

With,

- $\boldsymbol{\mu}(\mathsf{z};\boldsymbol{\theta})$ output of fully connected decoding neural network.
- $\boldsymbol{\sigma}^2_{\boldsymbol{\theta}}$ independent of z due to [Mattei and Frellsen (2018)] .



mschoeberl@gmail.com

Auto-Encoding Variational Bayes

According [Kingma and Welling (2013)] .

$$q(\mathsf{x}|\boldsymbol{\theta}) = \int q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta})\, q(\mathsf{z})\, d\mathsf{z}$$

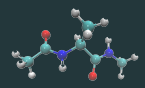
Mapping $\mathsf{z} \to \mathsf{x}$ (decoder):

$$\boxed{q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}(\mathsf{z};\boldsymbol{\theta}),\ S_{\boldsymbol{\theta}} = \mathrm{diag}(\boldsymbol{\sigma}^2_{\boldsymbol{\theta}})).}$$

With,

- $\boldsymbol{\mu}(\mathsf{z};\boldsymbol{\theta})$ output of fully connected decoding neural network.
- $\boldsymbol{\sigma}^2_{\boldsymbol{\theta}}$ independent of z due to [Mattei and Frellsen (2018)] .

- Proposed model does not pre-assume any physical insight.
- No a priori assumption or data pre-processing needed.



mschoeberl@gmail.com

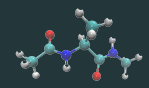
Auto-Encoding Variational Bayes

$$q(\mathbf{x}|\boldsymbol{\theta}) = \int q(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta})\, q(\mathbf{z}|\boldsymbol{\theta})\, d\mathbf{z}$$

We assume CVs are normal distributed according,

$$\boxed{q(\mathbf{z}) = \mathcal{N}(\mathbf{0}, I).}$$

- Here: $\dim(\mathbf{z}) = 2$.
- Explore assigned meaning of CVs given $\dim(\mathbf{z}) = 2$.



mschoeberl@gmail.com

Auto-Encoding Variational Bayes

The approximate posterior is of the following form,

$$\boxed{r(\mathsf{z}|\mathsf{x};\phi) = \mathcal{N}\left(\boldsymbol{\mu}(\mathsf{x};\phi), S_\phi = \mathsf{diag}(\boldsymbol{\sigma}(\mathsf{x};\phi)^2)\right).}$$

With,

- $\boldsymbol{\mu}(\mathsf{x};\phi)$ and $\boldsymbol{\sigma}^2(\mathsf{x};\phi)$ described by a fully connected encoding network.
- Last layers separate into $\boldsymbol{\mu}(\mathsf{x};\phi)$ and $\boldsymbol{\sigma}^2(\mathsf{x};\phi)$.

mschoeberl@gmail.com

Prediction of atomistic configurations for $\{\mathsf{z}|z_1 = [-4, 4], z_2 = 0\}$

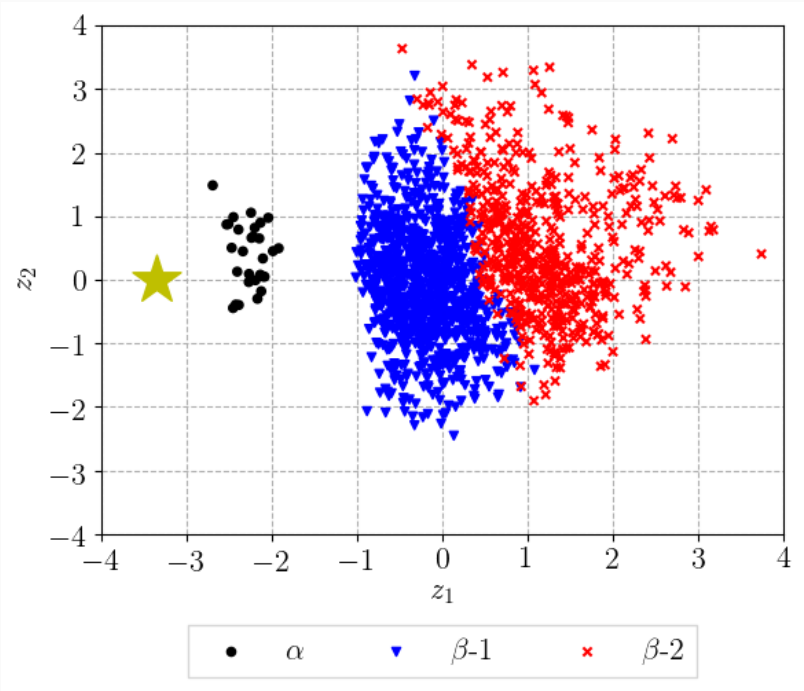


Figure 4: CVs are highly correlated with the dihedral angles (ϕ, ψ).

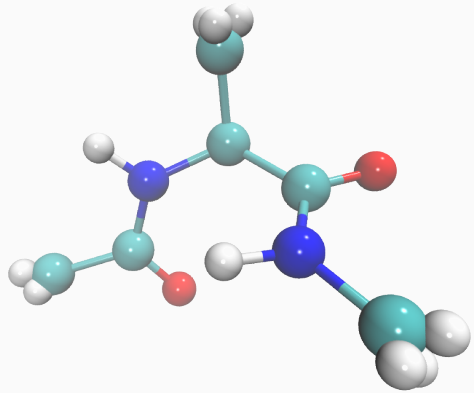


Figure 5: Mean prediction $\mathsf{x} = \boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathsf{z})$ of decoder $q_{\boldsymbol{\theta}}(\mathsf{x}|\mathsf{z})$, given $\{\mathsf{z}|z_1 = [-4, 4], z_2 = 0\}$.

mschoeberl@gmail.com

Prediction of atomistic configurations for $\{\mathsf{z}|z_1 = [-4, 4], z_2 = 0\}$

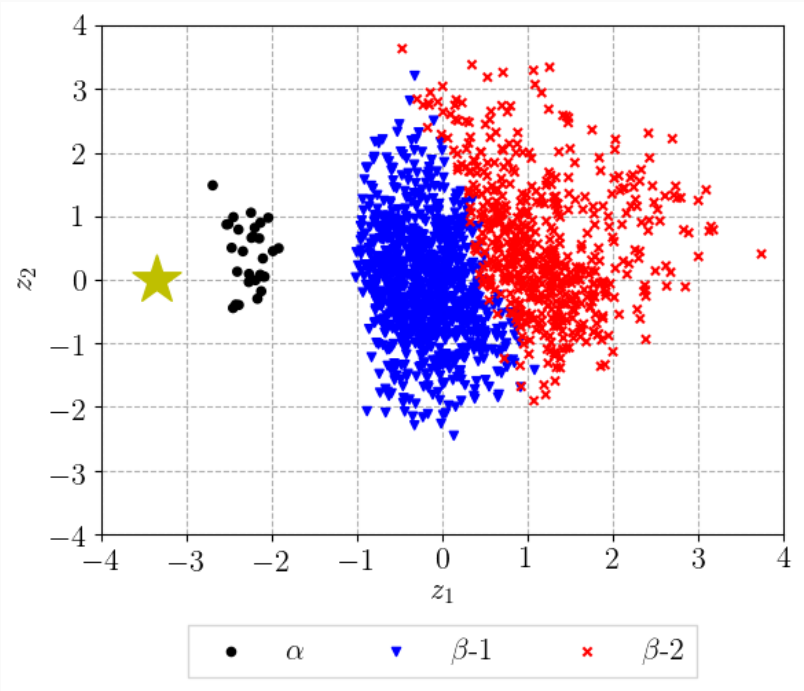


Figure 4: CVs are highly correlated with the dihedral angles (ϕ, ψ).

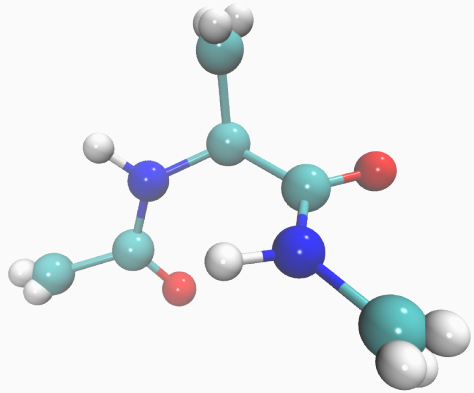
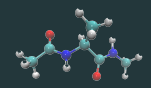
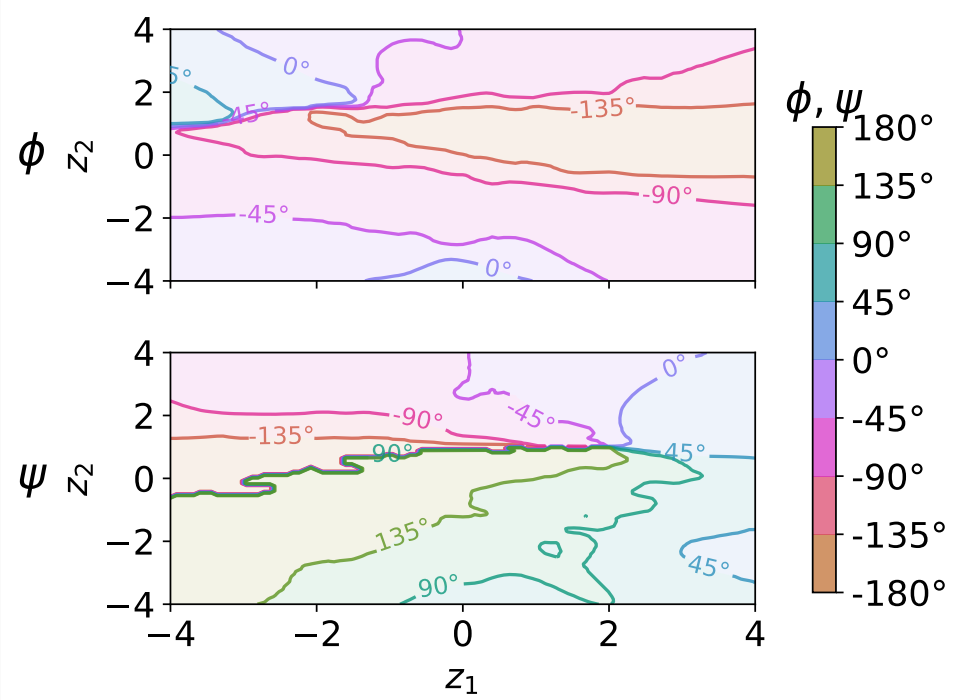


Figure 5: Mean prediction $\mathsf{x} = \boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathsf{z})$ of decoder $q_{\boldsymbol{\theta}}(\mathsf{x}|\mathsf{z})$, given $\{\mathsf{z}|z_1 = [-4, 4], z_2 = 0\}$.



mschoeberl@gmail.com

Identified CVs show high correlation to known optimal description, the dihedral angles (ϕ, ψ)



- Discovered CVs $\mathbf{z}$: Map to the (ϕ, ψ) angles.
- Note, instead of having a distinct (ϕ, ψ) value, we obtain a distribution of CVs implied by $r_{\phi}(\mathbf{z}|\mathbf{x})$.

Figure 6: Predicted dihedral angles (ϕ, ψ) given the latent variables $\mathbf{z} \in [-4, 4]^2$.

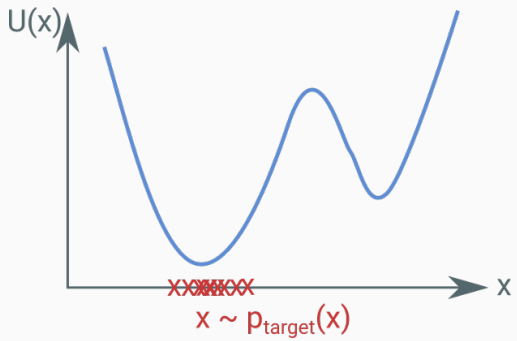
mschoeberl@gmail.com

Objective utilizing data $x^{(i)} \sim p_{\text{target}}(\mathbf{x})$

$$\min_{q(\mathbf{x})} D_{\text{KL}}\left(p_{\text{target}}(\mathbf{x})||q(\mathbf{x})\right) \leq -\mathbb{E}_{r(\mathbf{z}|\mathbf{x})}\left[\log q(\mathbf{x},\mathbf{z}|\boldsymbol{\theta}) - \log r(\mathbf{z}|\mathbf{x};\boldsymbol{\phi})\right]$$

Trade-off

- ✓ Revealing latent lower-dimensional embedding.
- ✓ Generative model, throughout Bayesian formulation feasible.
- × Explorative capabilities are limited.
- × Biased data-based approximation of $p_{\text{target}}(\mathbf{x})$ (e.g. unseen modes) yields biased models.
- × Biased predictions without being aware of it.



Embedded-Physics Approach

mschoeberl@gmail.com

Incorporate available physics of $p_{\text{target}}(x)$

Instead of the forward KL-divergence,

$$\min_{q(x)} D_{KL}(p_{\text{target}}(x)||q(x)) = \min_{q(x)} \left[- \int p_{\text{target}}(x) \log \frac{q(x)}{p_{\text{target}}(x)} \, dx \right],$$

mschoeberl@gmail.com

Incorporate available physics of $p_{\text{target}}(x)$

Instead of the forward KL-divergence,

$$\min_{q(x)} D_{KL}(p_{\text{target}}(x)||q(x)) = \min_{q(x)} \left[- \int p_{\text{target}}(x) \log \frac{q(x)}{p_{\text{target}}(x)} \, dx \right],$$

"flip" the distance measure:

$$\min_{q(x)} D_{KL}(q(x)||p_{\text{target}}(x)) = \min_{q(x)} \left[- \int q(x) \log \frac{p_{\text{target}}(x)}{q(x)} \, dx \right].$$

mschoeberl@gmail.com

Incorporate available physics of $p_{\text{target}}(x)$

Instead of the forward KL-divergence,

$$\min_{q(x)} D_{KL}(p_{\text{target}}(x)||q(x)) = \min_{q(x)} \left[- \int p_{\text{target}}(x) \log \frac{q(x)}{p_{\text{target}}(x)} \, dx \right],$$

"flip" the distance measure:

$$\min_{q(x)} D_{\text{KL}}(q(x)||p_{\text{target}}(x)) = \min_{q(x)} \left[- \int q(x) \log \frac{p_{\text{target}}(x)}{q(x)} \, dx \right].$$

- Forward KL-divergence involves expectations with respect to $p_{\text{target}}(x) \rightarrow$ Requires data.
- Reverse KL-divergence: Evaluate $\log p_{\text{target}}(x)$ (i.e. evaluate $U(x)$) at samples $x^{(i)} \sim q(x)$.
- *Problem independent* and *explorative* CG approach.

The objective

$$\min_{q(\mathbf{x})} D_{KL}(q(\mathbf{x})||p_{\text{target}}(\mathbf{x})) = \min_{q(\mathbf{x})} \left[- \int q(\mathbf{x}) \log \frac{p_{\text{target}}(\mathbf{x})}{q(\mathbf{x})} \, d\mathbf{x} \right]$$

How to model $q(\mathbf{x})$?

- Hierarchical Variational Models [Ranganath et al. (2016)]

$$q(\mathbf{x}) = \int q(\mathbf{x}|\mathbf{z}) q(\mathbf{z}) d\mathbf{z}.$$

  with an expressive probabilistic mapping $q(\mathbf{x}|\mathbf{z})$.

✓ Hierarchical extension facilitates the construction of an expressive $q(\mathbf{x})$.

✓ Inference is feasible by bounding the objective.

mschoeberl@gmail.com

Remarks on the objective

$$\min_{q(\mathsf{x})} D_{KL}(q(\mathsf{x})||p_{\text{target}}(\mathsf{x})) = \max_{q(\mathsf{x})} \underbrace{\left[\int q(\mathsf{x}) \log \frac{p_{\text{target}}(\mathsf{x})}{q(\mathsf{x})} \, d\mathsf{x}\right]}_{\mathcal{L}}$$

Consider a parametrization $\boldsymbol{\theta}$:

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})} \left[\log p_{\text{target}}(\mathsf{x}) - \log q(\mathsf{x};\boldsymbol{\theta})\right]$$

mschoeberl@gmail.com

Remarks on the objective

$$\min_{q(\mathbf{x})} D_{KL}(q(\mathbf{x})||p_{\text{target}}(\mathbf{x})) = \max_{q(\mathbf{x})} \underbrace{\left[\int q(\mathbf{x}) \log \frac{p_{\text{target}}(\mathbf{x})}{q(\mathbf{x})} \, d\mathbf{x}\right]}_{\mathcal{L}}$$

Consider a parametrization $\boldsymbol{\theta}$:

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{q(\mathbf{x};\boldsymbol{\theta})} \left[\log p_{\text{target}}(\mathbf{x}) - \log q(\mathbf{x};\boldsymbol{\theta}) \right]$$

- $\mathbb{E}_{q(\mathbf{x};\boldsymbol{\theta})} [\log p_{\text{target}}(\mathbf{x})]$ tractable if we can sample from $q(\mathbf{x};\boldsymbol{\theta})$.

mschoeberl@gmail.com

Remarks on the objective

$$\min_{q(\mathsf{x})} D_{KL}(q(\mathsf{x})||p_{\text{target}}(\mathsf{x})) = \max_{q(\mathsf{x})} \underbrace{\left[\int q(\mathsf{x}) \log \frac{p_{\text{target}}(\mathsf{x})}{q(\mathsf{x})} \, d\mathsf{x}\right]}_{\mathcal{L}}$$

Consider a parametrization $\boldsymbol{\theta}$:

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})} \left[\log p_{\text{target}}(\mathsf{x}) - \log q(\mathsf{x};\boldsymbol{\theta})\right]$$

- $\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})} [\log p_{\text{target}}(\mathsf{x})]$ tractable if we can sample from $q(\mathsf{x};\boldsymbol{\theta})$.
- $-\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})} [\log q(\mathsf{x};\boldsymbol{\theta})]$ is the entropy of $q(\mathsf{x};\boldsymbol{\theta})$, $\mathbb{H}(q(\mathsf{x};\boldsymbol{\theta}))$.
  - It comprises an integration with respect to z: $q(\mathsf{x}) = \int q(\mathsf{x}|\mathsf{z})q(\mathsf{z}) \, d\mathsf{z}$
  - In general not analytically tractable.

Construct a tractable (lower-)bound on $-\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})} [\log q(\mathsf{x};\boldsymbol{\theta})]$.

mschoeberl@gmail.com

Lower bound the entropy $\mathbb{H}(q(\mathsf{x};\boldsymbol{\theta}))$

$$-\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta})\right] = -\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta}) + \underbrace{D_{KL}\left(q_P(\mathsf{z}|\mathsf{x})||q_P(\mathsf{z}|\mathsf{x})\right)}_{=0}\right]$$

$$-\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta})\right] \geq -\mathbb{E}_{q(\mathsf{x},\mathsf{z};\boldsymbol{\theta})}\left[\log q(\mathsf{z}) + \log q(\mathsf{x}|\mathsf{z}) - \log r(\mathsf{z}|\mathsf{x})\right]$$

mschoeberl@gmail.com

Tractable objective

By employing the entropy bound derived before in

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log p_{\text{target}}(\mathsf{x}) - \log q(\mathsf{x};\boldsymbol{\theta})\right],$$

we obtain a tractable lower-bound for the variational approach with,

$$\tilde{\mathcal{L}}(\boldsymbol{\theta},\boldsymbol{\phi}) = \mathbb{E}_{q(\mathsf{x},\mathsf{z};\boldsymbol{\theta})}\Big[\log p_{\text{target}}(\mathsf{x}) + \log r(\mathsf{z}|\mathsf{x};\boldsymbol{\phi}) - \log q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta}) - \log q(\mathsf{z};\boldsymbol{\theta})\Big].$$

Proposed approach does not require any data and any physical insight - but learns from the potential energy and the force field

mschoeberl@gmail.com

How to connect the objective with the physics, e.g. the atomistic potential $U(\mathsf{x})$?

With $p_{\text{target}} \propto -\beta U(\mathsf{x})$:

$$\tilde{\mathcal{L}}(\boldsymbol{\theta}, \boldsymbol{\phi}) = -\beta \mathbb{E}_{q(\mathsf{x},\mathsf{z};\boldsymbol{\theta})}\big[\, U(\mathsf{x}) \,\big] + \mathbb{E}_{q(\mathsf{x},\mathsf{z};\boldsymbol{\theta})}\big[\, \log r(\mathsf{z}|\mathsf{x};\boldsymbol{\phi}) - \log q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta}) - \log q(\mathsf{z};\boldsymbol{\theta}) \big]$$

$$= -\beta \mathbb{E}_{q(\mathsf{x},\mathsf{z};\boldsymbol{\theta})}\big[\, \underbrace{U(\mathsf{x}) - \frac{1}{\beta} \log r(\mathsf{z}|\mathsf{x};\boldsymbol{\phi})}_{\tilde{U}(\mathsf{x},\mathsf{z})} \,\big] + \mathbb{H}(q(\mathsf{x},\mathsf{z}|\boldsymbol{\theta}))$$

mschoeberl@gmail.com

How to connect the objective with the physics, e.g. the atomistic potential $U(\mathsf{x})$?

With $p_{\text{target}} \propto -\beta U(\mathsf{x})$:

$$\tilde{\mathcal{L}}(\boldsymbol{\theta}, \boldsymbol{\phi}) = -\beta \mathbb{E}_{q(\mathsf{x},\mathsf{z};\boldsymbol{\theta})}\big[\, U(\mathsf{x}) \,\big] + \mathbb{E}_{q(\mathsf{x},\mathsf{z};\boldsymbol{\theta})}\big[\, \log r(\mathsf{z}|\mathsf{x};\boldsymbol{\phi}) - \log q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta}) - \log q(\mathsf{z};\boldsymbol{\theta}) \big]$$

$$= -\beta \mathbb{E}_{q(\mathsf{x},\mathsf{z};\boldsymbol{\theta})}\big[\, \underbrace{U(\mathsf{x}) - \frac{1}{\beta} \log r(\mathsf{z}|\mathsf{x};\boldsymbol{\phi})}_{\tilde{U}(\mathsf{x},\mathsf{z})} \,\big] + \mathbb{H}(q(\mathsf{x},\mathsf{z}|\boldsymbol{\theta}))$$

mschoeberl@gmail.com

How to connect the objective with the physics, e.g. the atomistic potential $U(\mathsf{x})$?

With $p_{\text{target}} \propto -\beta U(\mathsf{x})$:

$$\begin{aligned}\tilde{\mathcal{L}}(\boldsymbol{\theta}, \boldsymbol{\phi}) &= -\beta \mathbb{E}_{q(\mathsf{x},\mathsf{z};\boldsymbol{\theta})}\big[\, U(\mathsf{x}) \,\big] + \mathbb{E}_{q(\mathsf{x},\mathsf{z};\boldsymbol{\theta})}\big[\, \log r(\mathsf{z}|\mathsf{x};\boldsymbol{\phi}) - \log q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta}) - \log q(\mathsf{z};\boldsymbol{\theta}) \big] \\ &= -\beta \mathbb{E}_{q(\mathsf{x},\mathsf{z};\boldsymbol{\theta})}\big[\, \underbrace{U(\mathsf{x}) - \frac{1}{\beta} \log r(\mathsf{z}|\mathsf{x};\boldsymbol{\phi})}_{\tilde{U}(\mathsf{x},\mathsf{z})} \,\big] + \mathbb{H}(q(\mathsf{x},\mathsf{z}|\boldsymbol{\theta}))\end{aligned}$$

Maximization of $\tilde{\mathcal{L}}(\boldsymbol{\theta}, \boldsymbol{\phi})$ as balance between:

- Minimization of the averaged (joint) potential energy $\tilde{U}(\mathsf{x}, \mathsf{z})$.
- Maximization of the entropy of the generative model $\mathbb{H}(q(\mathsf{x}, \mathsf{z}|\boldsymbol{\theta}))$.

mschoeberl@gmail.com

Gradient estimation of the term involving $U(\mathsf{x})$

$$\tilde{\mathcal{L}}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \mathbb{E}_{q(\mathsf{x},\mathsf{z};\boldsymbol{\theta})}\big[-\beta U(\mathsf{x}) + \log r(\mathsf{z}|\mathsf{x};\boldsymbol{\phi}) - \log q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta}) - \log q(\mathsf{z};\boldsymbol{\theta})\big].$$

Consider $q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathsf{z}), S_{\boldsymbol{\theta}}(\mathsf{z}))$ and $q(\mathsf{x}) = \mathcal{N}(0, I)$

$$\frac{\partial}{\partial \boldsymbol{\theta}} \Big\langle \langle -\beta U(\mathsf{x}) \rangle_{q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta})} \Big\rangle_{q(\mathsf{z})}$$

mschoeberl@gmail.com

Gradient estimation of the term involving $U(\mathsf{x})$

$$\tilde{\mathcal{L}}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \mathbb{E}_{q(\mathsf{x},\mathsf{z};\boldsymbol{\theta})}\big[-\beta U(\mathsf{x}) + \log r(\mathsf{z}|\mathsf{x};\boldsymbol{\phi}) - \log q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta}) - \log q(\mathsf{z};\boldsymbol{\theta})\big].$$

Consider $q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathsf{z}), S_{\boldsymbol{\theta}}(\mathsf{z}))$ and $q(\mathsf{x}) = \mathcal{N}(0, I)$

$$\frac{\partial}{\partial \boldsymbol{\theta}} \Big\langle \langle -\beta U(\mathsf{x}) \rangle_{q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta})} \Big\rangle_{q(\mathsf{z})}$$

- Estimator highly affected by *noise*.
- Reparametrize x by auxiliary random variable $\boldsymbol{\epsilon}$ and differentiable transformation $g(\boldsymbol{\epsilon}; \mathsf{z})$:

$$\begin{aligned} \mathsf{x}(\boldsymbol{\epsilon}; \mathsf{z}) &= g_{\boldsymbol{\theta}}(\boldsymbol{\epsilon}; \mathsf{z}) \\ &= \boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathsf{z}) + \boldsymbol{\sigma}_{\boldsymbol{\theta}}(\mathsf{z}) \odot \boldsymbol{\epsilon} \quad \text{with } p(\boldsymbol{\epsilon}) = \mathcal{N}(0, I). \end{aligned}$$

mschoeberl@gmail.com

Gradient estimation of the term involving $U(\mathsf{x})$

$$\tilde{\mathcal{L}}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \mathbb{E}_{q(\mathsf{x},\mathsf{z};\boldsymbol{\theta})}\Big[-\beta U(\mathsf{x}) + \log r(\mathsf{z}|\mathsf{x};\boldsymbol{\phi}) - \log q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta}) - \log q(\mathsf{z};\boldsymbol{\theta})\Big].$$

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\theta}} \Big\langle \langle -\beta U(\mathsf{x}) \rangle_{q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta})} \Big\rangle_{q(\mathsf{z})} &= \left\langle \left\langle -\beta \frac{\partial U(\mathsf{x}(\boldsymbol{\epsilon};\mathsf{z}))}{\partial \boldsymbol{\theta}} \right\rangle_{p(\boldsymbol{\epsilon})} \right\rangle_{q(\mathsf{z})} \\
&= \left\langle \left\langle -\beta \frac{\partial U(\mathsf{x}(\boldsymbol{\epsilon};\mathsf{z}))}{\partial \mathsf{x}} \frac{\partial \mathsf{x}(\boldsymbol{\epsilon};\mathsf{z})}{\partial \boldsymbol{\theta}} \right\rangle_{p(\boldsymbol{\epsilon})} \right\rangle_{q(\mathsf{z})} \\
&= \left\langle \left\langle \beta \underbrace{F(\mathsf{x}(\boldsymbol{\epsilon};\mathsf{z}))}_{\text{Atomistic force-field}} \frac{\partial \mathsf{x}(\boldsymbol{\epsilon};\mathsf{z})}{\partial \boldsymbol{\theta}} \right\rangle_{p(\boldsymbol{\epsilon})} \right\rangle_{q(\mathsf{z})}
\end{aligned}$$

mschoeberl@gmail.com

Gradient estimation of the term involving $U(\mathsf{x})$

$$\tilde{\mathcal{L}}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \mathbb{E}_{q(\mathsf{x},\mathsf{z};\boldsymbol{\theta})}\Big[-\beta U(\mathsf{x}) + \log r(\mathsf{z}|\mathsf{x};\boldsymbol{\phi}) - \log q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta}) - \log q(\mathsf{z};\boldsymbol{\theta})\Big].$$

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\theta}} \Big\langle \langle -\beta U(\mathsf{x}) \rangle_{q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta})} \Big\rangle_{q(\mathsf{z})} &= \left\langle \left\langle -\beta \frac{\partial U(\mathsf{x}(\boldsymbol{\epsilon};\mathsf{z}))}{\partial \boldsymbol{\theta}} \right\rangle_{p(\boldsymbol{\epsilon})} \right\rangle_{q(\mathsf{z})} \\
&= \left\langle \left\langle -\beta \frac{\partial U(\mathsf{x}(\boldsymbol{\epsilon};\mathsf{z}))}{\partial \mathsf{x}} \frac{\partial \mathsf{x}(\boldsymbol{\epsilon};\mathsf{z})}{\partial \boldsymbol{\theta}} \right\rangle_{p(\boldsymbol{\epsilon})} \right\rangle_{q(\mathsf{z})} \\
&= \left\langle \left\langle \beta \underbrace{F(\mathsf{x}(\boldsymbol{\epsilon};\mathsf{z}))}_{\text{Atomistic force-field}} \frac{\partial \mathsf{x}(\boldsymbol{\epsilon};\mathsf{z})}{\partial \boldsymbol{\theta}} \right\rangle_{p(\boldsymbol{\epsilon})} \right\rangle_{q(\mathsf{z})}
\end{aligned}$$

Learning by evaluating $U(\mathsf{x})$ and $F(\mathsf{x})$

Gradient estimation involves the *evaluation* of the force field $F(\mathsf{x})$ at configurations $\mathsf{x}^{(i)} \sim q(\mathsf{x})$, *not* the *simulation* of $p_{\text{target}}(\mathsf{x})$.

mschoeberl@gmail.com

Target distribution

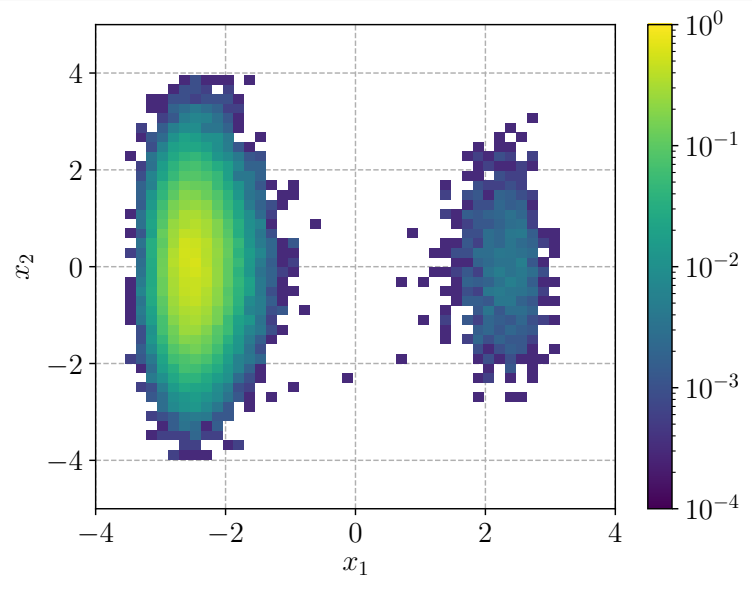


Figure 7: Reference histogram obtained by importance sampling.

$$p_{\text{target}}(\mathbf{x}) \propto e^{-\beta U(\mathbf{x})}$$

with $\beta = 1$ and,

$$\boxed{U(x_1, x_2) = \frac{1}{4} a x_1^4 - \frac{1}{2} b x_1^2 + c x_1 + \frac{1}{2} d x_2^2.}$$

- Bistable in x_1 and harmonic in x_2.
- Two distinct modes, one less pronounced.
- Random-walk MCMC is not able to capture both modes properly.

mschoeberl@gmail.com

Model details

Three components to specify:

1. Latent representation:

$$q(\mathsf{z}) = \mathcal{N}(\mathbf{0}, I)$$

2. Probabilistic **de**coder:

$$q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathsf{z}), S_{\boldsymbol{\theta}} = \mathrm{diag}(\boldsymbol{\sigma}^2_{\boldsymbol{\theta}}(\mathsf{z})))$$

3. Probabilistic **en**coder:

$$r(\mathsf{z}|\mathsf{x};\boldsymbol{\phi}) = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\phi}}(\mathsf{x}), S_{\boldsymbol{\phi}} = \mathrm{diag}(\boldsymbol{\sigma}^2_{\boldsymbol{\phi}}(\mathsf{x})))$$

mschoeberl@gmail.com

Model details

Three components to specify:

1. Latent representation:

$$q(\mathsf{z}) = \mathcal{N}(\mathbf{0}, I)$$

2. Probabilistic **de**coder:

$$q(\mathsf{x}|\mathsf{z};\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathsf{z}), S_{\boldsymbol{\theta}} = \mathrm{diag}(\boldsymbol{\sigma}^2_{\boldsymbol{\theta}}(\mathsf{z})))$$

3. Probabilistic **en**coder:

$$r(\mathsf{z}|\mathsf{x};\boldsymbol{\phi}) = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\phi}}(\mathsf{x}), S_{\boldsymbol{\phi}} = \mathrm{diag}(\boldsymbol{\sigma}^2_{\boldsymbol{\phi}}(\mathsf{x})))$$

Training details

- Tempering during training is employed, initially start with $\beta_s = 1\mathrm{e} - 10$ and $\bar{p}_{\mathrm{target}}(\mathsf{x}) \propto e^{-\beta_s U(\mathsf{x})}$.
- Convergence with β_s.
- Increase β_s such that $D_{\mathrm{KL}}(q(\mathsf{x},\mathsf{z})||\bar{p}(\mathsf{x},\mathsf{z}))$ does not exceed a threshold.
- Repeat until $\beta_s = \beta$.

Numerical illustration - KL-divergence bounds

mschoeberl@gmail.com

Bounds of the KL-divergence and predictions during training

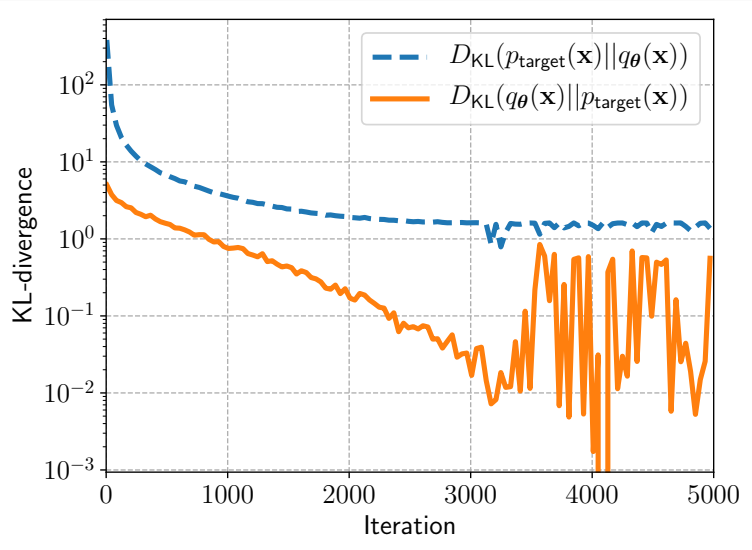


Figure 8: Upper- and lower bound of the training objective.

Observable estimation

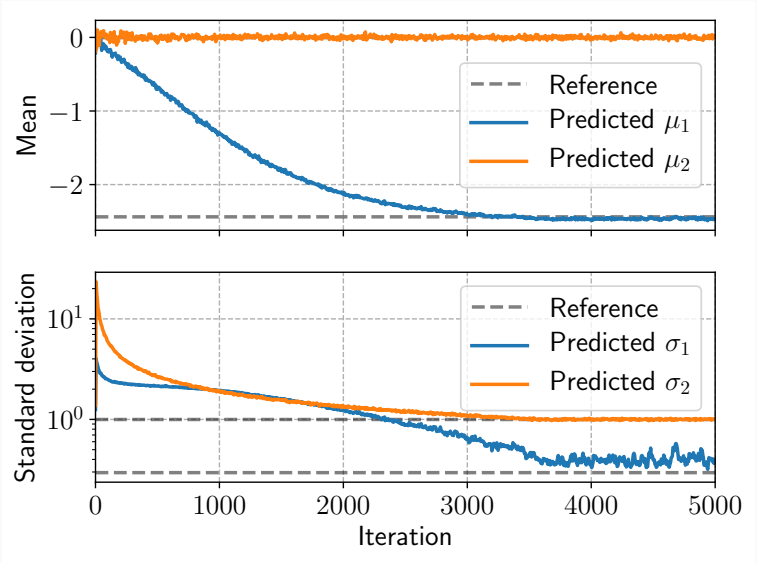


Figure 9: Mean and standard deviation compared to reference.

mschoeberl@gmail.com

Predictive distribution $q(\mathsf{x}) = \int q(\mathsf{x}|\mathsf{z})q(\mathsf{z})d\mathsf{z}$

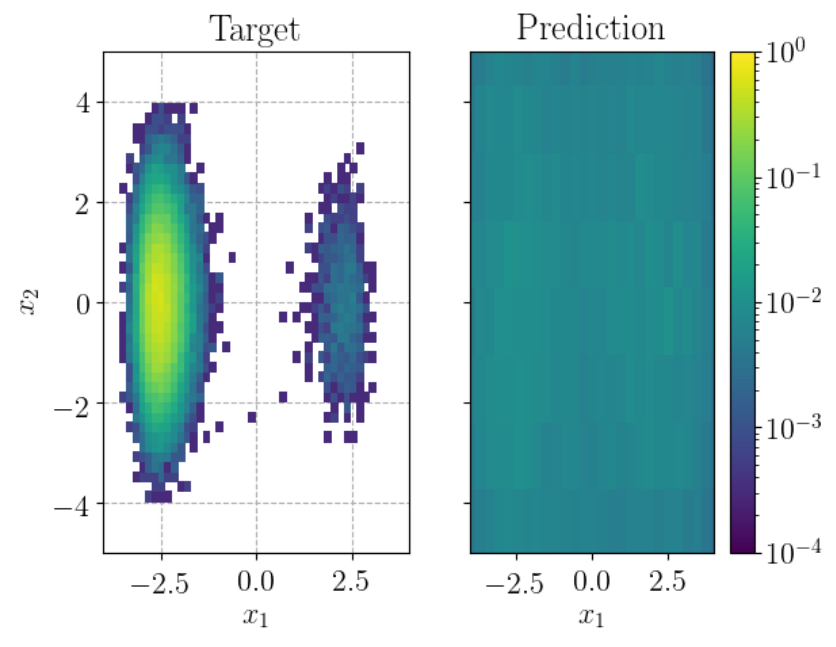


Figure 10: Target (left) and prediction (right) during learning.

Potential energy (e.g. at $x_2 = 0$)

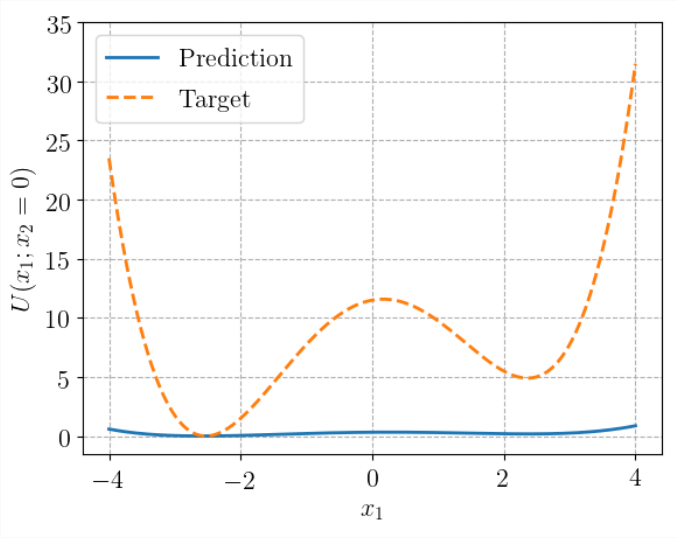


Figure 11: Reference potential energy $U(\mathsf{x})$ and potential energy estimated by predictive distribution $U_p(\mathsf{x}) \propto -\frac{1}{\beta} \log q(\mathsf{x})$ (noisy) at $\{\mathsf{x}|x_2 = 0\}$

mschoeberl@gmail.com

Predictive distribution $q(\mathsf{x}) = \int q(\mathsf{x}|\mathsf{z})q(\mathsf{z})d\mathsf{z}$

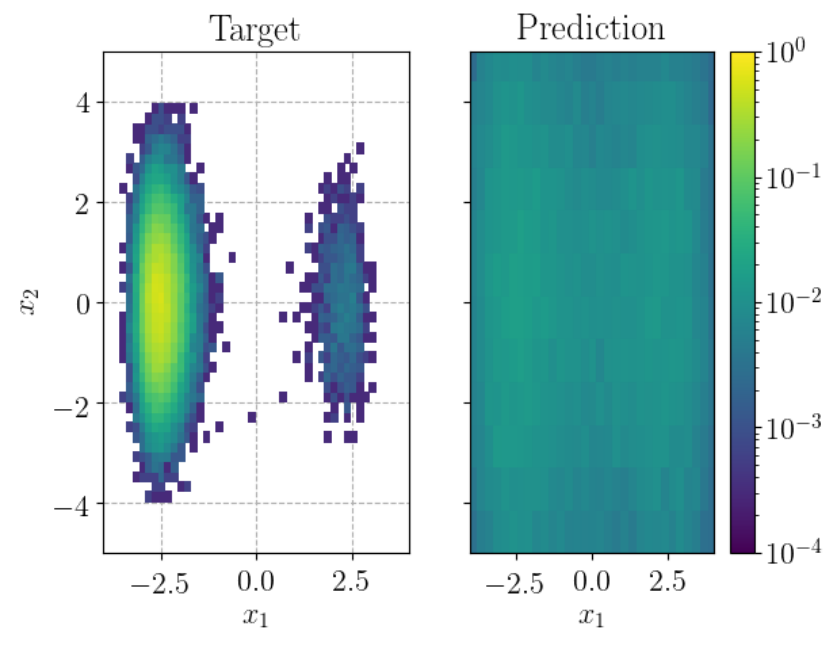


Figure 10: Target (left) and prediction (right) during learning.

Potential energy (e.g. at $x_2 = 0$)

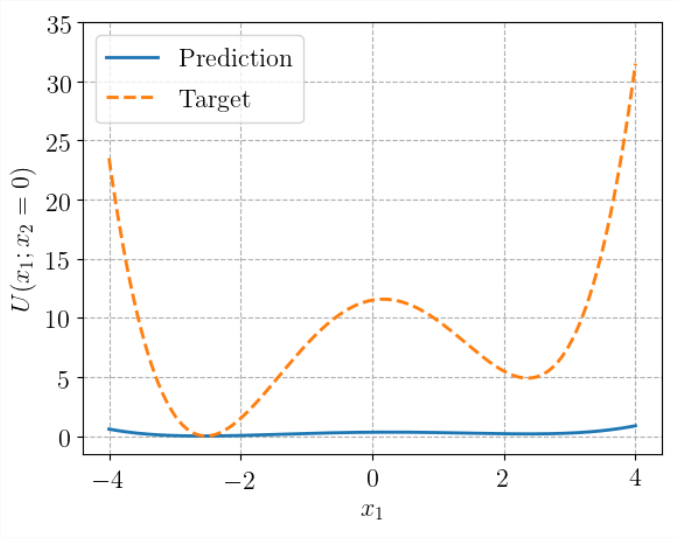


Figure 11: Reference potential energy $U(\mathsf{x})$ and potential energy estimated by predictive distribution $U_p(\mathsf{x}) \propto -\frac{1}{\beta} \log q(\mathsf{x})$ (noisy) at $\{\mathsf{x}|x_2 = 0\}$

Numerical illustration - Prediction

mschoeberl@gmail.com

Predictive distribution $q(\mathsf{x}) = \int q(\mathsf{x}|\mathsf{z})q(\mathsf{z})d\mathsf{z}$

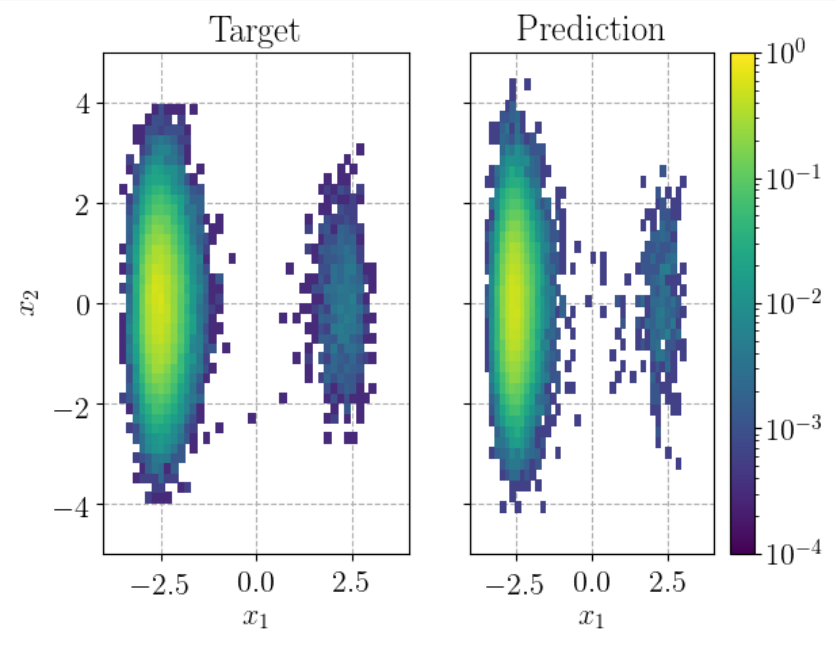


Figure 10: Target (left) and prediction (right) during learning.

Potential energy (e.g. at $x_2 = 0$)

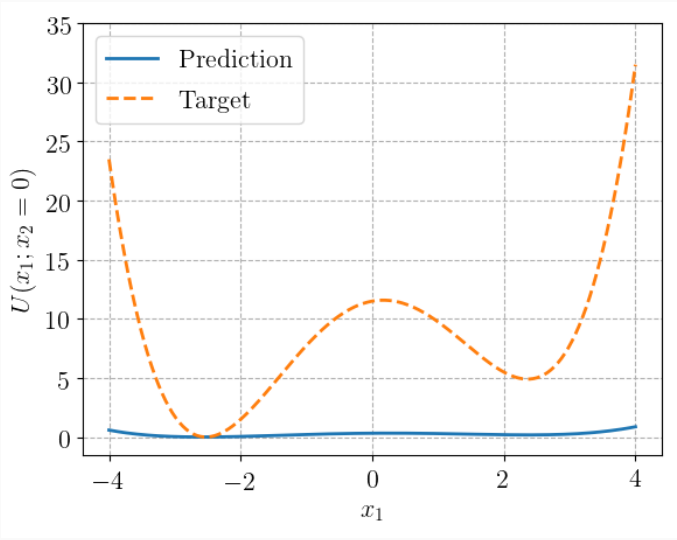


Figure 11: Reference potential energy $U(\mathsf{x})$ and potential energy estimated by predictive distribution $U_p(\mathsf{x}) \propto -\frac{1}{\beta} \log q(\mathsf{x})$ (noisy) at $\{\mathsf{x}|x_2 = 0\}$

Numerical illustration - Prediction

mschoeberl@gmail.com

Predictive distribution $q(\mathsf{x}) = \int q(\mathsf{x}|\mathsf{z})q(\mathsf{z})d\mathsf{z}$

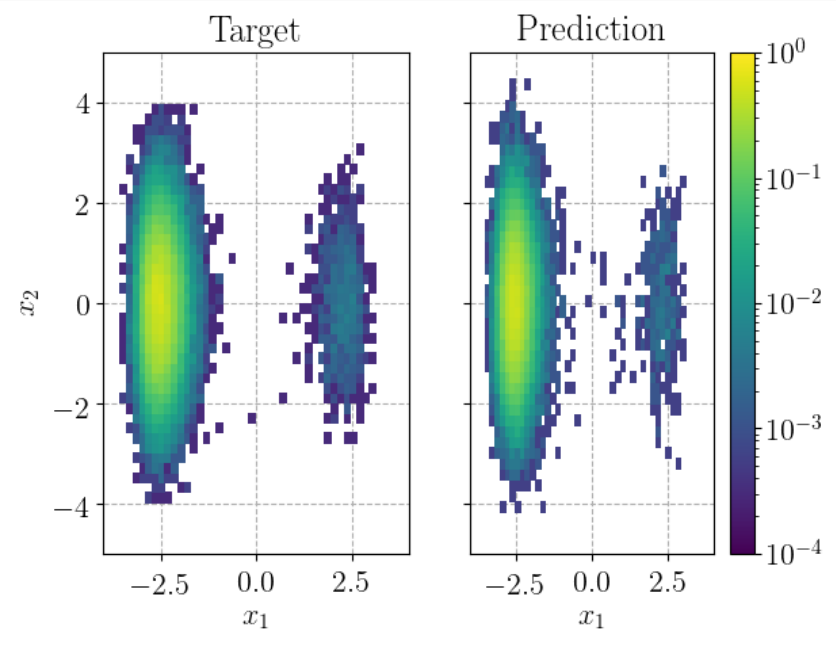


Figure 10: Target (left) and prediction (right) during learning.

Potential energy (e.g. at $x_2 = 0$)

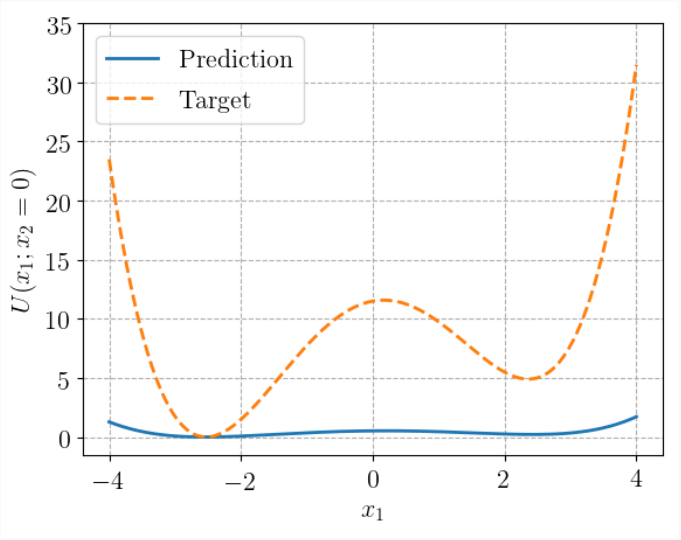


Figure 11: Reference potential energy $U(\mathsf{x})$ and potential energy estimated by predictive distribution $U_p(\mathsf{x}) \propto -\frac{1}{\beta} \log q(\mathsf{x})$ (noisy) at $\{\mathsf{x}|x_2 = 0\}$

Numerical illustration - Prediction

mschoeberl@gmail.com

Predictive distribution $q(\mathbf{x}) = \int q(\mathbf{x}|\mathbf{z})q(\mathbf{z})d\mathbf{z}$

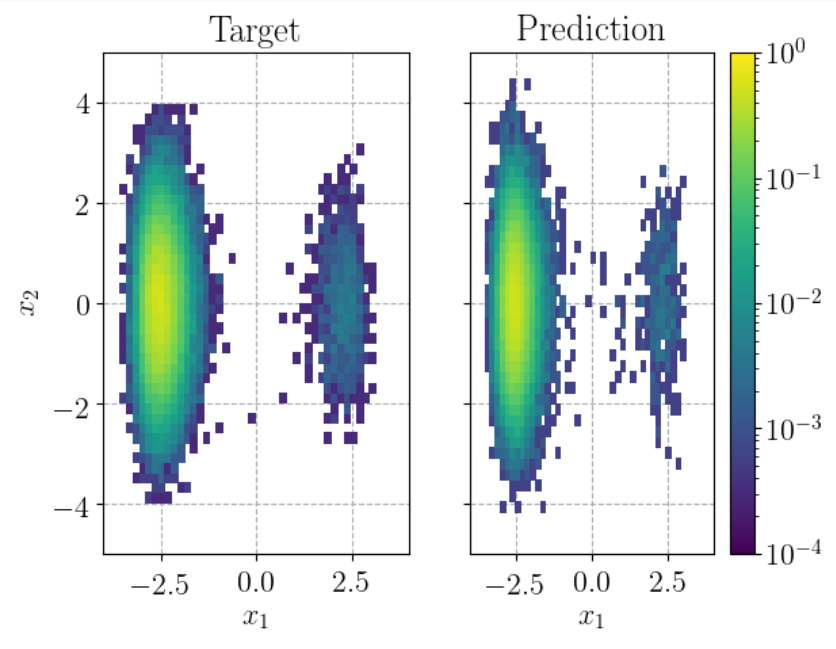


Figure 10: Target (left) and prediction (right) during learning.

Potential energy (e.g. at $x_2 = 0$)

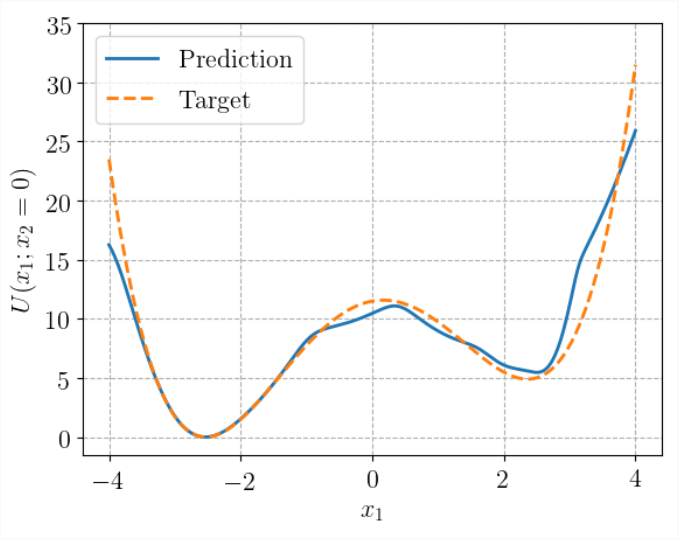


Figure 11: Reference potential energy $U(\mathbf{x})$ and potential energy estimated by predictive distribution $U_p(\mathbf{x}) \propto -\frac{1}{\beta} \log q(\mathbf{x})$ (noisy) at $\{\mathbf{x}|x_2 = 0\}$

Numerical illustration - What do we learn?

mschoeberl@gmail.com

What CVs are learned (without MD data)?

Latent embedding

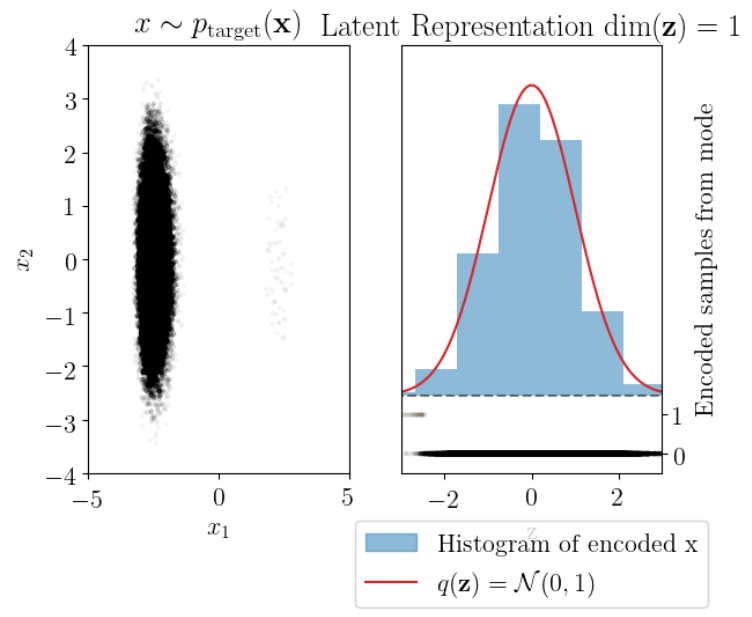


Figure 12: Reference samples encoded in the latent space.

CV-values **z** for *predicted* **x**

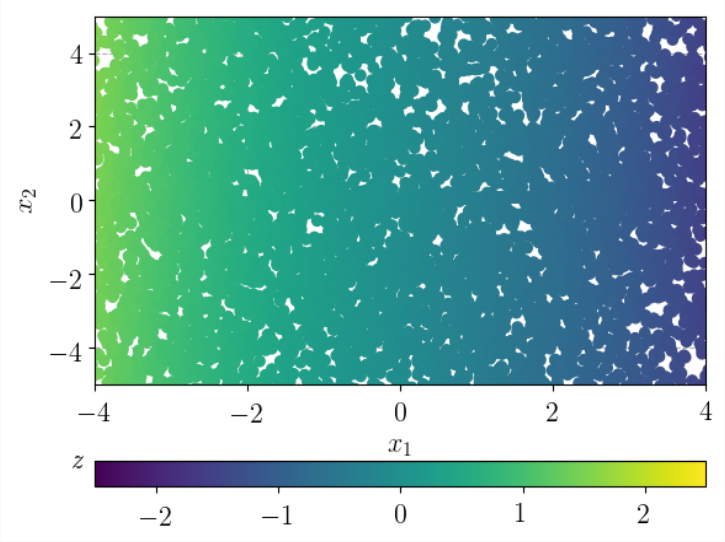


Figure 13: Samples $\mathbf{x}^{(i)} \sim q(\mathbf{x}; \boldsymbol{\theta})$ encoded with $r(\mathbf{z}|\mathbf{x}; \boldsymbol{\phi})$. Color of scatter indicate CV-values.

mschoeberl@gmail.com

What CVs are learned (without MD data)?

Latent embedding

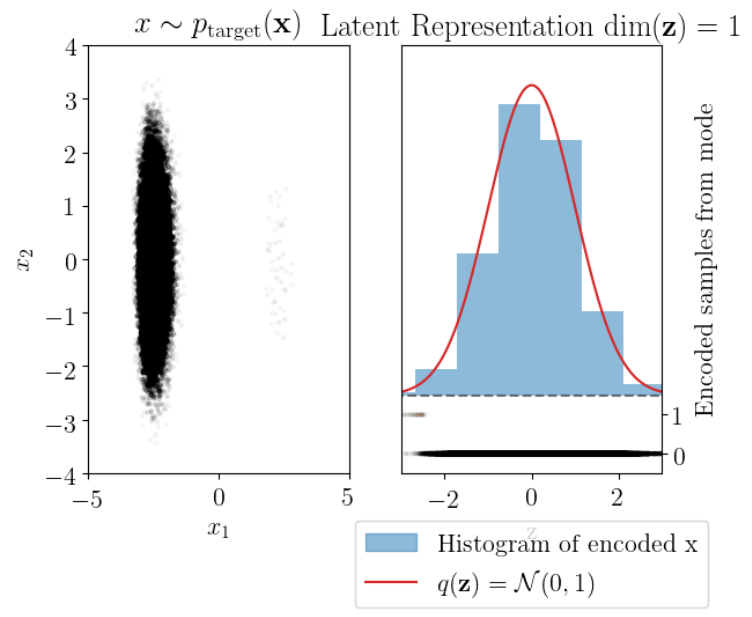


Figure 12: Reference samples encoded in the latent space.

CV-values **z** for *predicted* **x**

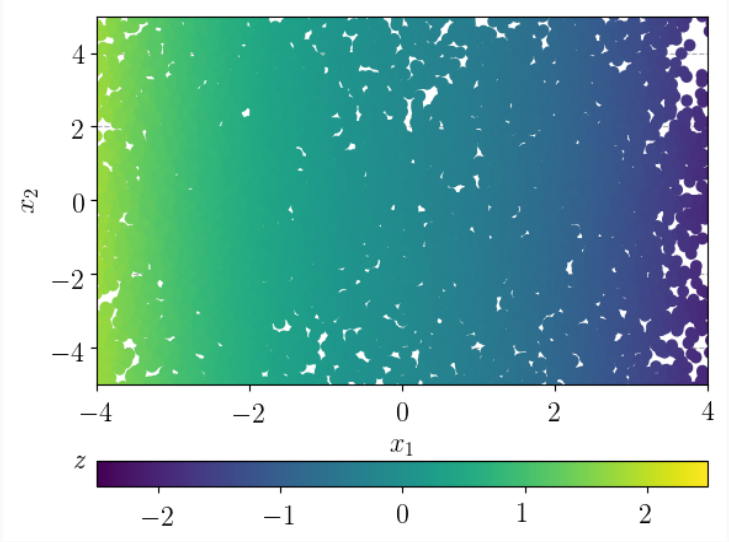


Figure 13: Samples $\mathbf{x}^{(i)} \sim q(\mathbf{x}; \boldsymbol{\theta})$ encoded with $r(\mathbf{z}|\mathbf{x}; \boldsymbol{\phi})$. Color of scatter indicate CV-values.

Numerical illustration - What do we learn?

mschoeberl@gmail.com

What CVs are learned (without MD data)?

Latent embedding

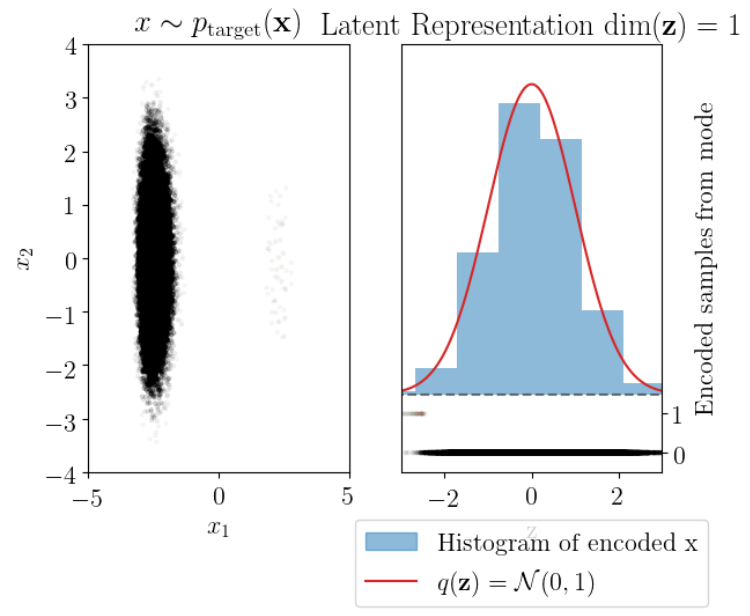


Figure 12: Reference samples encoded in the latent space.

CV-values **z** for *predicted* **x**

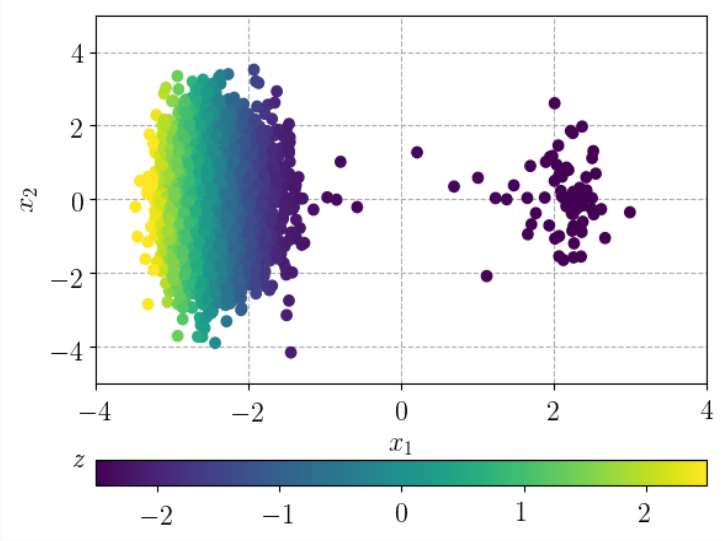


Figure 13: Samples $\mathbf{x}^{(i)} \sim q(\mathbf{x}; \boldsymbol{\theta})$ encoded with $r(\mathbf{z}|\mathbf{x}; \boldsymbol{\phi})$. Color of scatter indicate CV-values.

Without using data, one is able to learn CVs capturing multi-modality.

mschoeberl@gmail.com

What CVs are learned (without MD data)?

Latent embedding

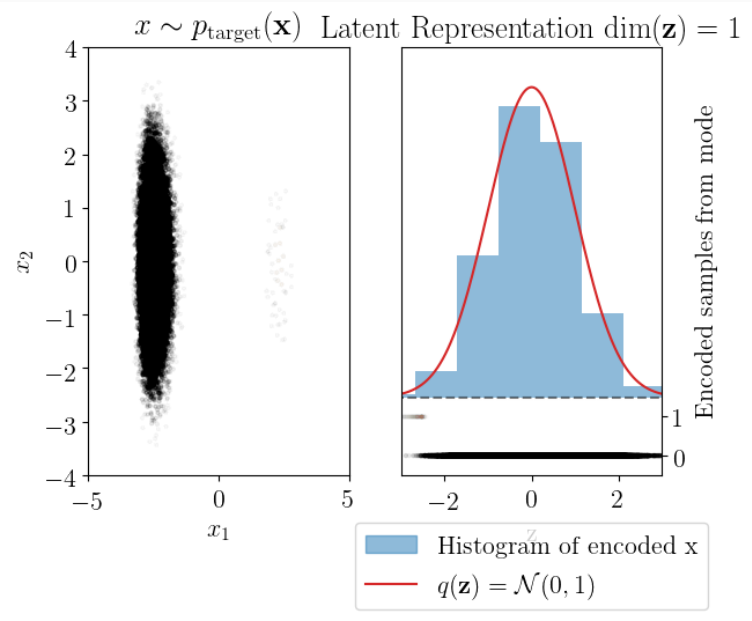


Figure 12: Reference samples encoded in the latent space.

CV-values **z** for *predicted* **x**

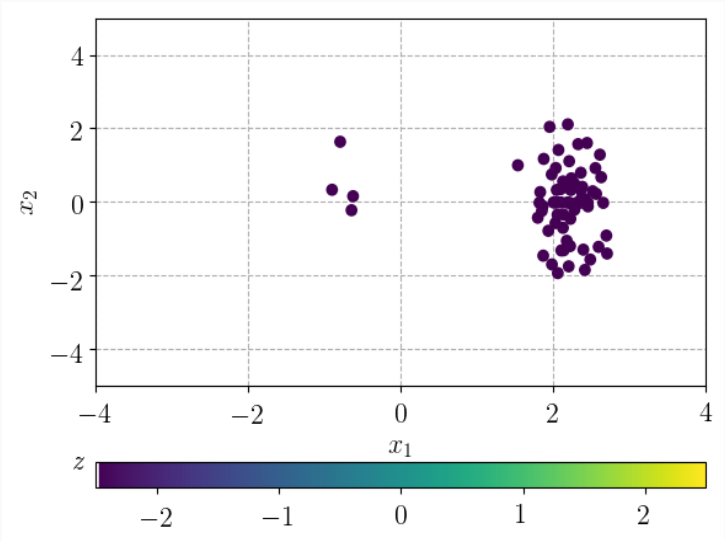


Figure 13: Samples $\mathbf{x}^{(i)} \sim q(\mathbf{x}; \boldsymbol{\theta})$ encoded with $r(\mathbf{z}|\mathbf{x}; \boldsymbol{\phi})$. Color of scatter indicate CV-values.

mschoeberl@gmail.com

What CVs are learned (without MD data)?

Latent embedding

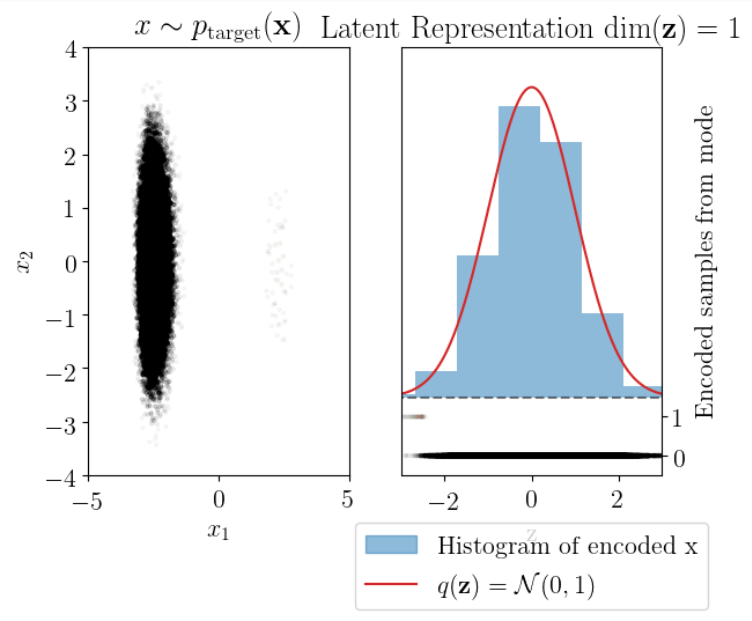


Figure 12: Reference samples encoded in the latent space.

CV-values **z** for *predicted* **x**

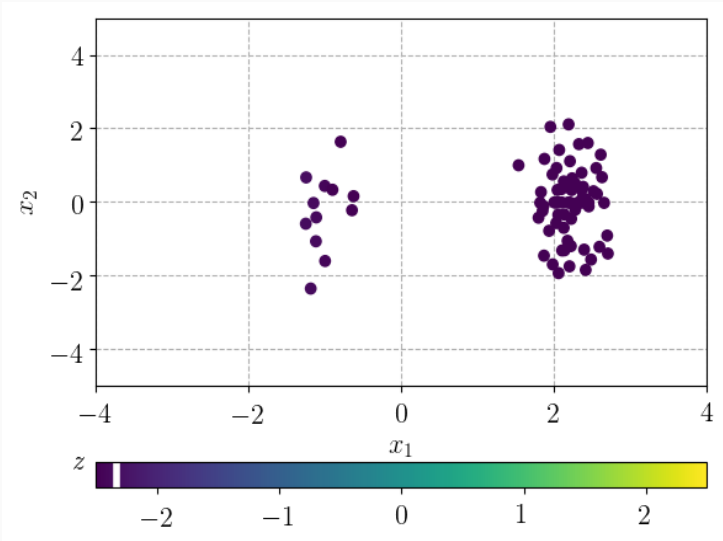


Figure 13: Samples $\mathbf{x}^{(i)} \sim q(\mathbf{x}; \boldsymbol{\theta})$ encoded with $r(\mathbf{z}|\mathbf{x}; \boldsymbol{\phi})$. Color of scatter indicate CV-values.

Numerical illustration - What do we learn?

mschoeberl@gmail.com

What CVs are learned (without MD data)?

Latent embedding

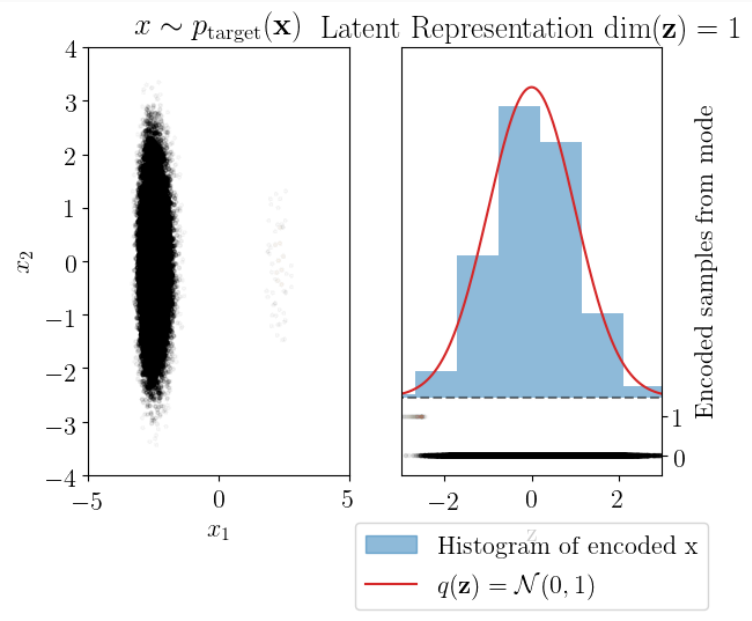


Figure 12: Reference samples encoded in the latent space.

CV-values **z** for *predicted* **x**

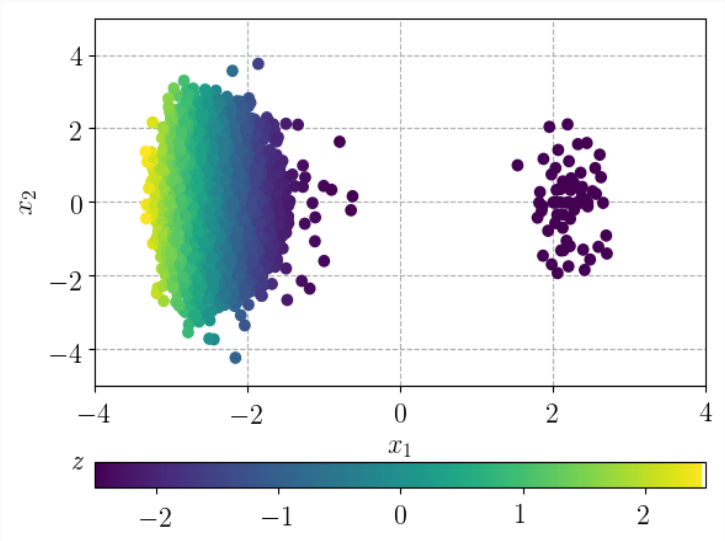


Figure 13: Samples $\mathbf{x}^{(i)} \sim q(\mathbf{x}; \boldsymbol{\theta})$ encoded with $r(\mathbf{z}|\mathbf{x}; \boldsymbol{\phi})$. Color of scatter indicate CV-values.

mschoeberl@gmail.com

Significant modes for ALA-2

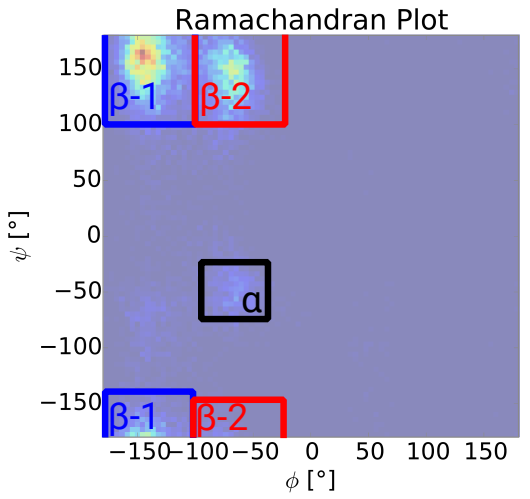
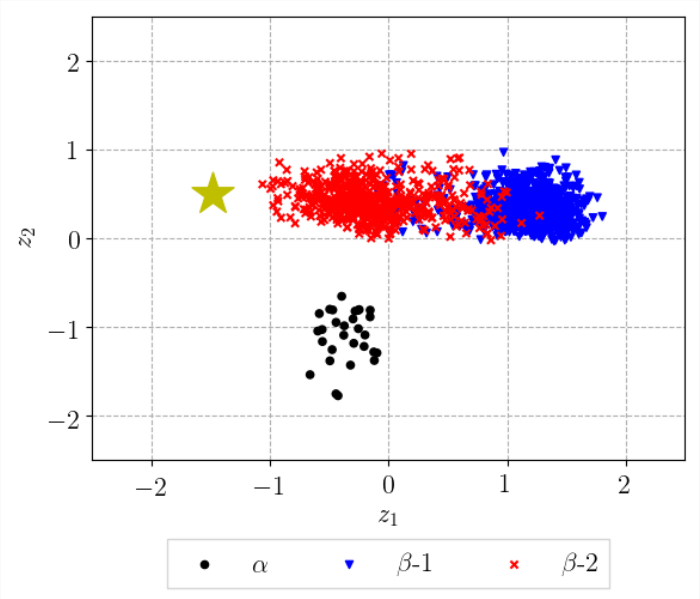


Figure 14: Reference modes

- What are relevant CVs of the system?
- The identification without data would facilitate the construction of enhanced sampling methods or biasing potentials.
- *Data-driven* approach: How to produce data accompanying whole configurational space while not being able to sample properly?

mschoeberl@gmail.com

Latent representation and prediction



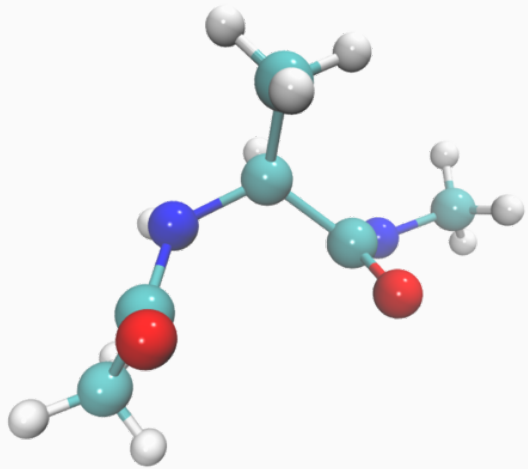
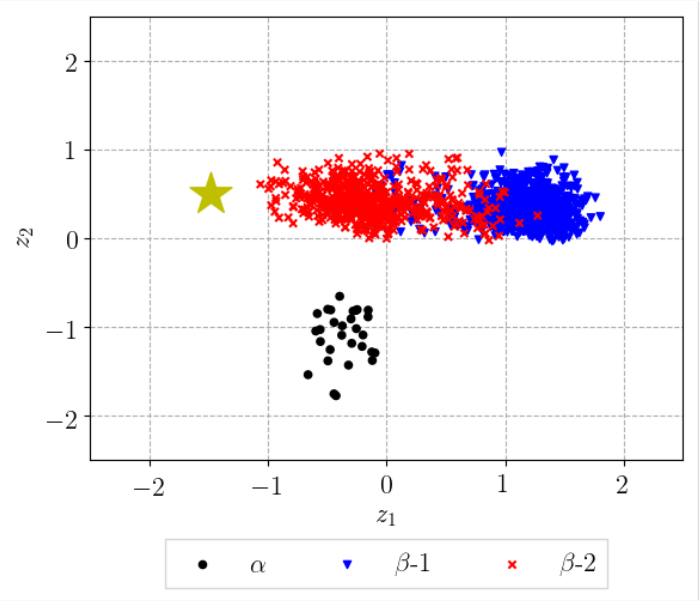


Figure 15: Latent representation of reference samples from multiple configurations (left). *Predicted* atomistic configurations (right), given the CV as indicated left.

mschoeberl@gmail.com

Latent representation and prediction



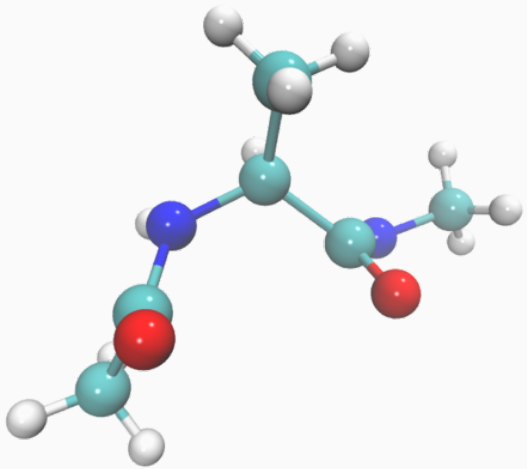
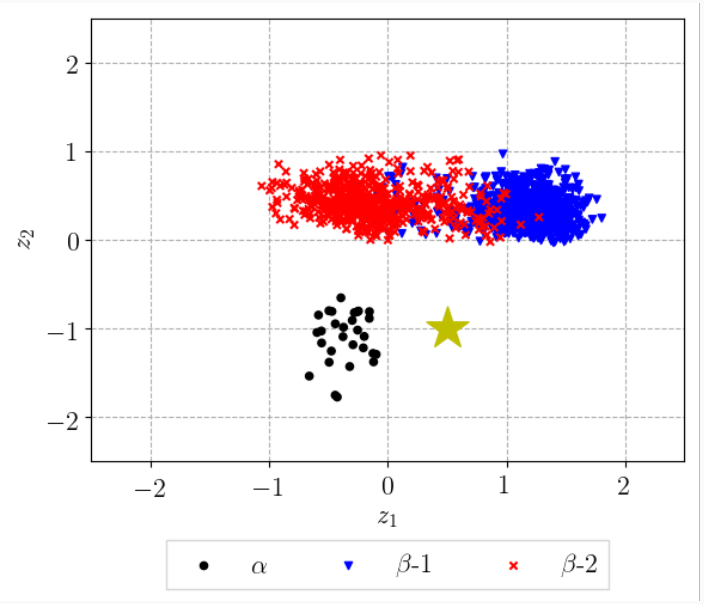


Figure 15: Latent representation of reference samples from multiple configurations (left). *Predicted* atomistic configurations (right), given the CV as indicated left.

mschoeberl@gmail.com

Latent representation and prediction



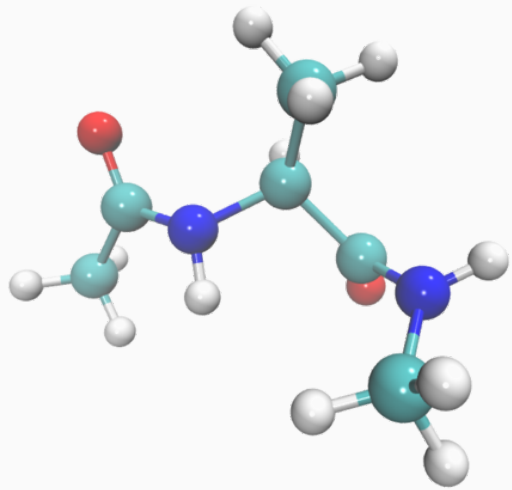


Figure 15: Latent representation of reference samples from multiple configurations (left). *Predicted* atomistic configurations (right), given the CV as indicated left.

Can we learn characteristics just by *evaluating* the force-field?

Characteristics (correlated to $\phi - \psi$ angles) are learned and yield atomistic configurations.

Summary

mschoeberl@gmail.com

Data-Driven

$$D_{\mathrm{KL}}\left(p_{\mathrm{target}}(\mathbf{x})||q(\mathbf{x})\right)$$

- Robust framework of CG in small data regime.
- Requires data from different modes.
- Generative approach.
- Quantification of epistemic uncertainty.

Embedded-Physics

$$D_{\mathrm{KL}}\left(q(\mathbf{x})||p_{\mathrm{target}}(\mathbf{x})\right)$$

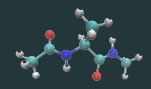
- Generative approach.
- No physical insight presumed. Instead: reveal insight while learning.
- Explorative capabilities - no data required.
- Problem independent as long as force-field is accessible.

Outline

- Uncertainty quantification for variational approach.
- Mixed formulation:

$$F(\boldsymbol{\phi}, \boldsymbol{\theta}) = \alpha D_{\mathrm{KL}}\left(p_{\mathrm{target}}(\mathbf{x})||q(\mathbf{x})\right) + (1-\alpha) D_{\mathrm{KL}}\left(q(\mathbf{x})||p_{\mathrm{target}}(\mathbf{x})\right)$$

Thank you!



mschoeberl@gmail.com

Encoding network

Linear layer	Input dimension	Output dimension	Activation layer	Activation function
$l_{\boldsymbol{\phi}}^{(1)}$	$\mathbf{dim(x)}$	d_1	$a^{(1)}$	SeLu[1]
$l_{\boldsymbol{\phi}}^{(2)}$	d_1	d_2	$a^{(2)}$	SeLu
$l_{\boldsymbol{\phi}}^{(3)}$	d_2	d_3	$a^{(3)}$	Log Sigmoid [2]
$l_{\boldsymbol{\phi}}^{(4)}$	d_3	$\mathbf{dim(z)}$	None	-
$l_{\boldsymbol{\phi}}^{(5)}$	d_3	$\mathbf{dim(z)}$	None	-

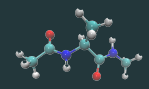
Table 1: Network specification of the encoding neural network with $d_1 = 50$, $d_2 = 100$, and $d_3 = 100$.

Decoding network

Linear layer	Input dimension	Output dimension	Activation layer	Activation function
$l_{\boldsymbol{\theta}}^{(1)}$	$\mathbf{dim(z)}$	d_3	$\tilde{a}^{(1)}$	Tanh
$l_{\boldsymbol{\theta}}^{(2)}$	d_3	d_2	$\tilde{a}^{(2)}$	Tanh
$l_{\boldsymbol{\theta}}^{(3)}$	d_2	d_1	$\tilde{a}^{(3)}$	Tanh
$l_{\boldsymbol{\theta}}^{(4)}$	d_1	$\mathbf{dim(x)}$	None	-

Table 2: Network specification of the decoding neural network with $d_{\{1,2,3\}}$ as defined in Table 1.

Numerical illustration - *Predictive* CVs



mschoeberl@gmail.com

- Predictions accounting for uncertainty in $\boldsymbol{\theta}$, with $\boldsymbol{\theta} \sim p(\boldsymbol{\theta}|\mathbf{X})$.
- Metropolis-within-Gibbs sampler [Mattei and Frellsen (2018)] corrects usage of approx. posterior $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$.

We illustrate this by computing the radius of gyration (Rg) [Fluitt and de Pablo (2015); Carmichael and Shell (2012)] given as,

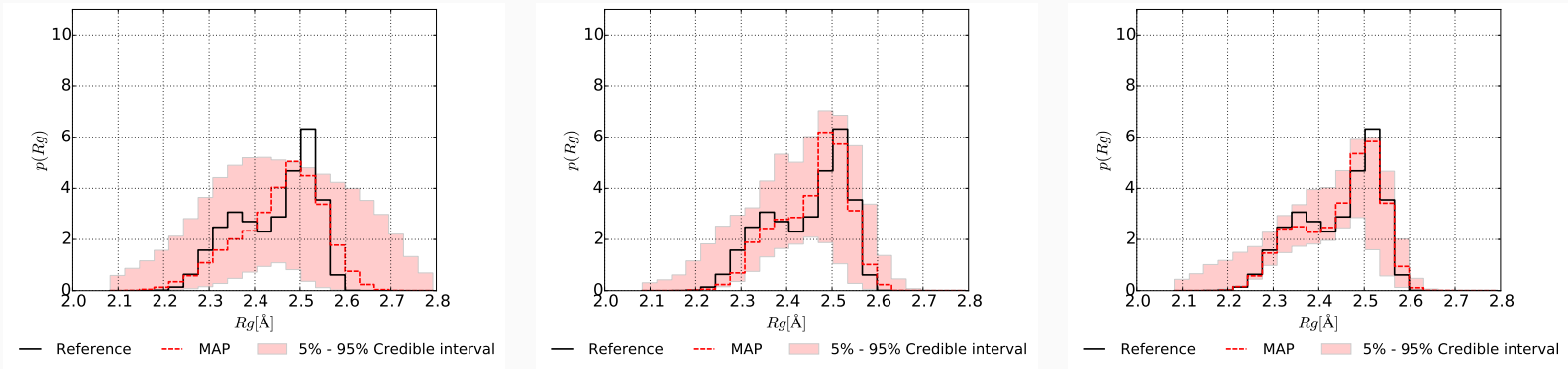
$$a_{\text{Rg}}(\mathbf{x}) = \sqrt{\frac{\sum_p m_p ||\mathbf{x}_p - \mathbf{x}_{\text{COM}}||^2}{\sum_p m_p}}.$$

- The sum considers atoms $p = 1, \ldots, P$.
- m_p and $\mathbf{x}_p$ denote the mass and the coordinates of each atom, respectively.
- $\mathbf{x}_{\text{COM}}$ denotes the center of mass of the peptide.

Numerical illustration - *Predictive* CVs

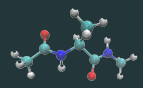
mschoeberl@gmail.com

- Predictions accounting for uncertainty in $\boldsymbol{\theta}$, with $\boldsymbol{\theta} \sim p(\boldsymbol{\theta}|\mathbf{X})$.
- Metropolis-within-Gibbs sampler [Mattei and Frellsen (2018)] corrects usage of approx. posterior $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$.



(a) $N = 50$. (b) $N = 200$. (c) $N = 500$.

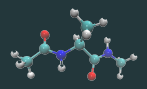
Figure 16: Predicted radius of gyration with $\mathbf{dim}(\mathbf{z}) = 2$ for various sizes N of the training dataset. The reference solution (black) is estimated by $N = 10000$. The shaded area represents the credible interval, reflecting the induced epistemic uncertainty from the limited amount of training data.



mschoeberl@gmail.com

Motivation

- How to avoid overfitting?
- How many $\boldsymbol{\theta}$s are actually required?
- Can one search *across models*?

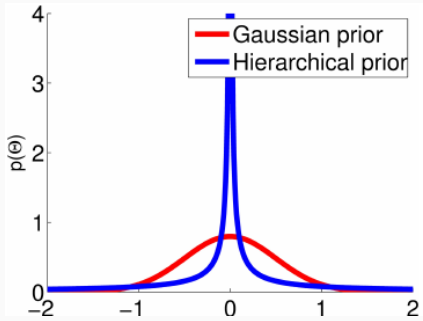

mschoeberl@gmail.com

Motivation

- How to avoid overfitting?
- How many $\boldsymbol{\theta}$s are actually required?
- Can one search *across models*?

Sparsity-enforcing hierarchical prior (ARD, [MacKay 1994])

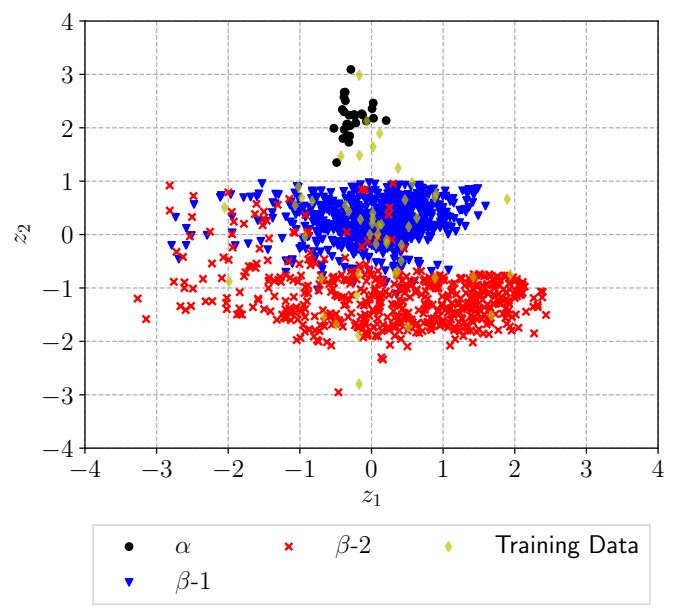
$$p(\boldsymbol{\theta}|\boldsymbol{\tau}) = \prod_j p(\boldsymbol{\theta}_j|\tau_j)$$
$$\theta_j \sim \mathcal{N}(0, \tau_j^{-1})$$
$$\tau_j \sim Gamma(a_0, b_0)$$



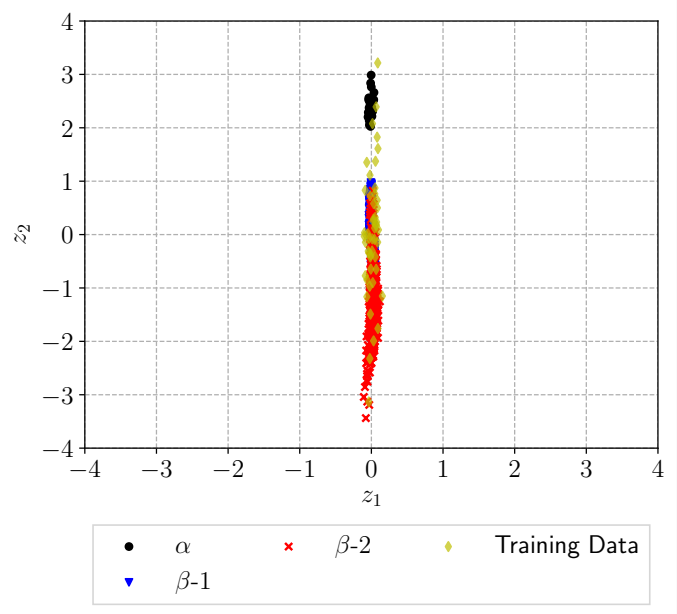
- Inner EM framework:
  - E-step: Estimate $< \tau_j >_{p(\tau_j|\theta_j)} = \frac{a_0+1/2}{b_0+\theta_j^2/2}$
  - M-step: Additive component to the derivative of the log-likelihood: $\frac{\partial}{\partial \tau_{c,j}} = - < \tau_j > \theta_j$

mschoeberl@gmail.com

Sparsity prior alleviates learning physically meaningful CVs given low data (e.g. N=50)



(a) Active ARD prior.



(b) Without ARD prior.

Figure 17: Representation of the $\mathbf{z}$-coordinates of the training data $\mathbf{X}$ with $N = 50$ in the CV space (yellow diamonds). Using the trained model and the mean of $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{z})$ we computed the $\mathbf{z}$-coordinates of 1527 test samples corresponding to different conformations of the alanine dipeptide to α (black), β-1 (blue), and β-2 (red).

mschoeberl@gmail.com

Lower bound the entropy $\mathbb{H}(q(\mathsf{x};\boldsymbol{\theta}))$

$$-\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta})\right] = -\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta}) + \underbrace{D_{KL}\left(q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})||q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})\right)}_{=0}\right]$$

mschoeberl@gmail.com

Lower bound the entropy $\mathbb{H}(q(\mathsf{x};\boldsymbol{\theta}))$

$$
\begin{aligned}
-\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta})\right] &= -\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta}) + \underbrace{D_{KL}\left(q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})||q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})\right)}_{=0}\right] \\
&\geq -\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta}) + D_{KL}\left(q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})||r(\mathsf{z}|\mathsf{x})\right)\right] \\
&= -\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\mathbb{E}_{q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})}\left[\log q(\mathsf{x};\boldsymbol{\theta}) + \log q_{\mathsf{P}}(\mathsf{z}|\mathsf{x}) - \log r(\mathsf{z}|\mathsf{x})\right]\right].
\end{aligned}
$$

mschoeberl@gmail.com

Lower bound the entropy $\mathbb{H}(q(\mathsf{x};\boldsymbol{\theta}))$

$$
\begin{aligned}
-\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta})\right] &= -\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta}) + \underbrace{D_{KL}\left(q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})||q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})\right)}_{=0}\right] \\
&\geq -\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta}) + D_{KL}\left(q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})||r(\mathsf{z}|\mathsf{x})\right)\right] \\
&= -\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\mathbb{E}_{q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})}\left[\log q(\mathsf{x};\boldsymbol{\theta}) + \log q_{\mathsf{P}}(\mathsf{z}|\mathsf{x}) - \log r(\mathsf{z}|\mathsf{x})\right]\right].
\end{aligned}
$$

Employ $\log q_{\mathsf{P}}(\mathsf{z}|\mathsf{x}) = \log q(\mathsf{z}) + \log q(\mathsf{x}|\mathsf{z}) - \log q(\mathsf{x})$:

$$
\begin{aligned}
-\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta})\right] &\geq -\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\mathbb{E}_{q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})}\left[\log q(\mathsf{x};\boldsymbol{\theta}) + \log q(\mathsf{z}) + \log q(\mathsf{x}|\mathsf{z}) - \log q(\mathsf{x}) - \right.\right. \\
&= -\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\mathbb{E}_{q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})}\left[\log q(\mathsf{z}) + \log q(\mathsf{x}|\mathsf{z}) - \log r(\mathsf{z}|\mathsf{x})\right]\right]
\end{aligned}
$$

mschoeberl@gmail.com

Lower bound the entropy $\mathbb{H}(q(\mathsf{x};\boldsymbol{\theta}))$

$$
\begin{aligned}
-\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta})\right] &= -\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta}) + \underbrace{D_{KL}\left(q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})||q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})\right)}_{=0}\right] \\
&\geq -\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta}) + D_{KL}\left(q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})||r(\mathsf{z}|\mathsf{x})\right)\right] \\
&= -\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\mathbb{E}_{q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})}\left[\log q(\mathsf{x};\boldsymbol{\theta}) + \log q_{\mathsf{P}}(\mathsf{z}|\mathsf{x}) - \log r(\mathsf{z}|\mathsf{x})\right]\right].
\end{aligned}
$$

Employ $\log q_{\mathsf{P}}(\mathsf{z}|\mathsf{x}) = \log q(\mathsf{z}) + \log q(\mathsf{x}|\mathsf{z}) - \log q(\mathsf{x})$:

$$
\begin{aligned}
-\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta})\right] &\geq -\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\mathbb{E}_{q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})}\left[\log q(\mathsf{x};\boldsymbol{\theta}) + \log q(\mathsf{z}) + \log q(\mathsf{x}|\mathsf{z}) - \log q(\mathsf{x}) - \right.\right. \\
&= -\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\mathbb{E}_{q_{\mathsf{P}}(\mathsf{z}|\mathsf{x})}\left[\log q(\mathsf{z}) + \log q(\mathsf{x}|\mathsf{z}) - \log r(\mathsf{z}|\mathsf{x})\right]\right]
\end{aligned}
$$

$$
-\mathbb{E}_{q(\mathsf{x};\boldsymbol{\theta})}\left[\log q(\mathsf{x};\boldsymbol{\theta})\right] \geq -\mathbb{E}_{q(\mathsf{x},\mathsf{z};\boldsymbol{\theta})}\left[\log q(\mathsf{z}) + \log q(\mathsf{x}|\mathsf{z}) - \log r(\mathsf{z}|\mathsf{x})\right]
$$

References

Reaction Rate Theory: Faraday Discussion 195 (Faraday Discussions). Royal Society of Chemistry, 2017. ISBN 178262483X. URL https://www.amazon.com/Reaction-Rate-Theory-Discussion-Discussions/dp/178262483X?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=178262483X.

Matthew J. Beal and Zoubin Ghahramani. Variational bayesian learning of directed graphical models with hidden variables. *Bayesian Anal.*, 1(4):793–831, 12 2006. doi: 10.1214/06-BA126. URL https://doi.org/10.1214/06-BA126.

mschoeberl@gmail.com

I. Bilionis and P.S. Koutsourelakis. Free energy computations by minimization of kullback–leibler divergence: An efficient adaptive biasing potential method for sparse representations. *Journal of Computational Physics*, 231(9):3849 – 3870, 2012. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2012.01.033. URL http://www.sciencedirect.com/science/article/pii/S0021999112000630.

Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016. ISSN 0027-8424. doi: 10.1073/pnas.1517384113. URL https://www.pnas.org/content/113/15/3932.

mschoeberl@gmail.com

Scott P. Carmichael and M. Scott Shell. A new multiscale algorithm and its application to coarse-grained peptide models for self-assembly. *The Journal of Physical Chemistry B*, 116(29): 8383–8393, 2012. doi: 10.1021/jp2114994. URL https://doi.org/10.1021/jp2114994. PMID: 22300263.

Maghesree Chakraborty, Chenliang Xu, and Andrew D. White. Encoding and selecting coarse-grain mapping operators with hierarchical graphs, 2018.

Pei-Yang Chen and Mark E. Tuckerman. Molecular dynamics based enhanced sampling of collective variables with very large time steps. *The Journal of Chemical Physics*, 148(2):024106, 2018. doi: 10.1063/1.4999447. URL https://doi.org/10.1063/1.4999447.

mschoeberl@gmail.com

Wei Chen and Andrew L Ferguson. Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration, 2017.

Eric Darve, David Rodríguez-Gómez, and Andrew Pohorille. Adaptive biasing force method for scalar and vector free energy calculations. *The Journal of Chemical Physics*, 128(14):144120, 2008. doi: 10.1063/1.2829861. URL https://doi.org/10.1063/1.2829861.

Andrew L. Ferguson. Bayeswham: A bayesian approach for free energy estimation, reweighting, and uncertainty quantification in the weighted histogram analysis method. *Journal of Computational Chemistry*, 38(18):1583–1605, 7 2017. ISSN 0192-8651. doi: 10.1002/jcc.24800.

Andrew L. Ferguson, Athanassios Z. Panagiotopoulos, Pablo G. Debenedetti, and Ioannis G. Kevrekidis. Integrating diffusion maps with umbrella sampling: Application to alanine dipeptide. *The Journal of Chemical Physics*, 134(13):135103, 2011. doi: 10.1063/1.3574394. URL https://doi.org/10.1063/1.3574394.

Aaron M. Fluitt and Juan J. de Pablo. An analysis of biomolecular force fields for simulations of polyglutamine in solution. *Biophysical Journal*, 109(5):1009 – 1018, 2015. ISSN 0006-3495. doi: https://doi.org/10.1016/j.bpj.2015.07.018. URL http://www.sciencedirect.com/science/article/pii/S0006349515007249.

mschoeberl@gmail.com

Markos A. Katsoulakis and José Trashorras. Information loss in coarse-graining of stochastic particle dynamics. *Journal of Statistical Physics*, 122(1):115–135, Jan 2006. ISSN 1572-9613. doi: 10.1007/s10955-005-8063-1. URL https://doi.org/10.1007/s10955-005-8063-1.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2013.

Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 971–980. Curran Associates, Inc., 2017. URL http://papers.nips.cc/paper/6698-self-normalizing-neural-networks.pdf.

mschoeberl@gmail.com

Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20): 12562–12566, 2002. ISSN 0027-8424. doi: 10.1073/pnas.202427399. URL http://www.pnas.org/content/99/20/12562.

Pierre-Alexandre Mattei and Jes Frellsen. Leveraging the exact likelihood of deep latent variable models, 2018.

W. G. Noid. Perspective: Coarse-grained models for biomolecular systems. *The Journal of Chemical Physics*, 139(9):090901, 2013. doi: 10.1063/1.4818908. URL https://doi.org/10.1063/1.4818908.

Frank Noé and Hao Wu. Boltzmann generators - sampling equilibrium states of many-body systems with deep learning, 2018.

mschoeberl@gmail.com

Rajesh Ranganath, Dustin Tran, and David M. Blei. Hierarchical variational models. In *ICML*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 324–333. JMLR.org, 2016.

Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, pages 1278–1286, 2014. URL http://jmlr.org/proceedings/papers/v32/rezende14.html.

Romelia Salomon-Ferrer, David A. Case, and Ross C. Walker. An overview of the amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 3(2): 198–210. doi: 10.1002/wcms.1121. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1121.

mschoeberl@gmail.com

M. Scott Shell. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *The Journal of Chemical Physics*, 129(14):144108, 2008. doi: 10.1063/1.2992060. URL https://doi.org/10.1063/1.2992060.

W. Clark Still, Anna Tempczyk, Ronald C. Hawley, and Thomas Hendrickson. Semianalytical treatment of solvation for molecular mechanics and dynamics. *Journal of the American Chemical Society*, 112(16):6127–6129, Aug 1990. ISSN 0002-7863. doi: 10.1021/ja00172a038. URL https://doi.org/10.1021/ja00172a038.

mschoeberl@gmail.com

José Trashorras and Dimitrios Tsagkarogiannis. From mesoscale back to microscale: Reconstruction schemes for coarse-grained stochastic lattice systems. *SIAM Journal on Numerical Analysis*, 48 (5):1647–1677, 2010. doi: 10.1137/080722382. URL https://hal.archives-ouvertes.fr/hal-00275802.

Christoph Wehmeyer and Frank Noé. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *The Journal of Chemical Physics*, 148(24):241703, 2018. doi: 10.1063/1.5011399. URL https://doi.org/10.1063/1.5011399.