Predictive Collective Variable Discovery

with Deep Bayesian Models in Atomistic Systems.

M. Schöberl[1,2,3], N. Zabaras[1], P.-S. Koutsourelakis[2]

meeting venue
meeting date

[1]Center for Informatics and Computational Science, University of Notre Dame, Indiana, USA.
[2]Continuum Mechanics Group, Technical University of Munich, Germany.
[3]Fellow, Hanns-Seidel-Foundataion, Germany.

Predictive Collective Variable Discovery

mschoeberl@gmail.com

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

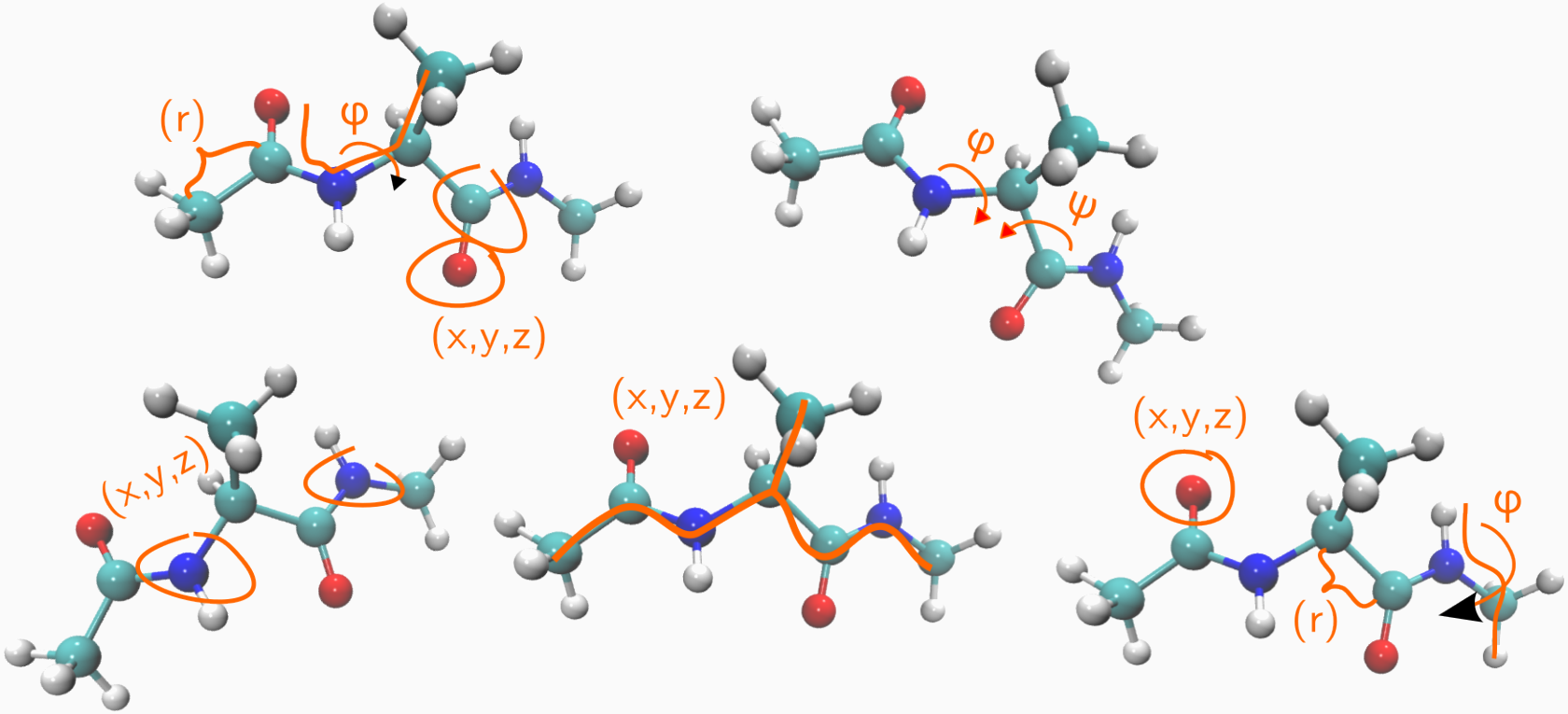


Figure 1: Alanine Dipeptide. Exemplary collective variables.

Predictive Collective Variable Discovery

mschoeberl@gmail.com

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

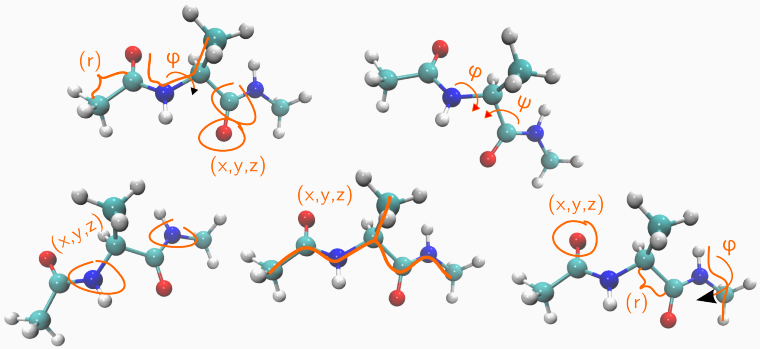


Figure 1: Alanine Dipeptide. Exemplary collective variables.

- Relevant coordinates are highly clustered around a set of *lower dimensional collective variables*.

  [Banon]

- Vast combinatoric possibilities for *choosing* CVs. [Chakraborty et al. (2018)]

Motivation

mschoeberl@gmail.com

Questions we address

- How to identify good CVs?
- What are *good* CVs?
- How to quantify "*good*"?
- Are identified CVs physically interpretable?

Motivation

mschoeberl@gmail.com

Questions we address

- How to identify good CVs?
- What are *good* CVs?
- How to quantify "*good*"?
- Are identified CVs physically interpretable?

Probabilistic predictions

- How to use utilize identified CVs for *predictive* purposes?
- How to quantify predictive *uncertainty*?

Motivation - Available approaches

mschoeberl@gmail.com

Methodologies for identifying collective variables at our disposal:

1. Intuition-, path-, state-based coordinates.
2. Linear dimensionality reduction methods.
   - Principal component analysis (PCA) [F.R.S. (1901); Hotelling (1933)] .
   - Multidimensional scaling (MDS) [Troyer and Cohen] .

Motivation - Available approaches

mschoeberl@gmail.com

Methodologies for identifying collective variables at our disposal:

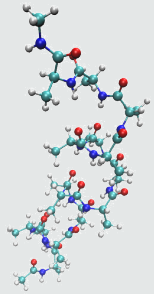
1. Intuition-, path-, state-based coordinates.
2. Linear dimensionality reduction methods.
   - Principal component analysis (PCA) [F.R.S. (1901); Hotelling (1933)] .
   - Multidimensional scaling (MDS) [Troyer and Cohen] .
3. Nonlinear dimensionality reduction methods.
   - Isometric feature map (Isomap) [Tenenbaum et al. (2000)] .
   - Sketch map [Ceriotti et al. (2011)] .
   - Diffusion maps (and locally scaled diffusion maps) [Coifman et al. (2005); Nadler et al. (2006)] .
   - Kinetic map [Noé and Clementi (2015)] and Commute maps [Noé et al. (2016)] .

Physical intuition does not suffice for identifying good collective variables [Rohrdanz et al. (2013)] .

Problem definition - Equilibrium statistical mechanics

mschoeberl@gmail.com

Atomistic model



$$p(\mathsf{x}) \propto e^{-\beta U(\mathsf{x})}$$

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

CV representation

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

- z: (reduced) collective variables.
- $\mathcal{R}$: operator (mapping onto CV space).

Methodology - Probabilistic generative model

mschoeberl@gmail.com

Introduce $p(\mathbf{z})$ and $p(\mathbf{x}|\mathbf{z})$, thus the joint $p(\mathbf{x}, \mathbf{z})$:

Methodology - Probabilistic generative model

mschoeberl@gmail.com

Introduce $p(\mathbf{z})$ and $p(\mathbf{x}|\mathbf{z})$, thus the joint $p(\mathbf{x}, \mathbf{z})$:

Which latent CVs give rise to the observations?

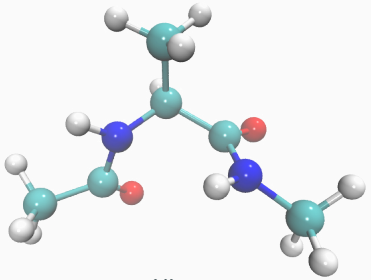
CV

Unobserved - - - - - - - - - - - - - - - - - $\mathbf{z} \sim p(\mathbf{z}^{(i)})$

Generate / Decode

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

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



Data $\mathbf{x}^{(i)} \sim p(\mathbf{x})$

Methodology - Probabilistic generative model

mschoeberl@gmail.com

Introduce $p(\mathbf{z})$ and $p(\mathbf{x}|\mathbf{z})$, thus the joint $p(\mathbf{x}, \mathbf{z})$:

Which latent CVs give rise to the observations?

CV

Unobserved - - - - - - - - - - - - - - - - $\mathbf{z} \sim p(z^{(i)})$

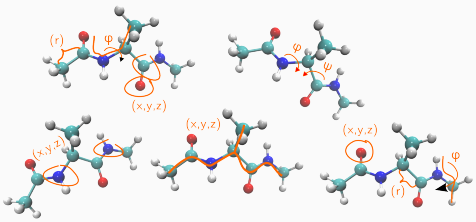
Generate / Decode

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

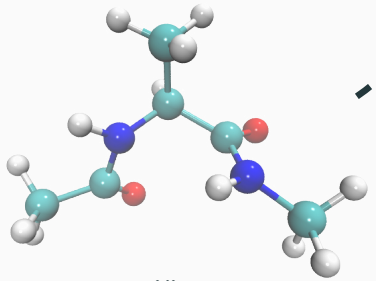
Encode $\mathbf{x}$ with

$p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{x}^{(i)}|\mathbf{z})p(\mathbf{z})}{p(\mathbf{x})}$

Code / Dictionary



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



Data $x^{(i)} \sim p(\mathbf{x})$

Methodology - Probabilistic generative model

mschoeberl@gmail.com

Proposed generative model [Jordan (1999)]

adopted from Predictive Coarse-Graining [Schöberl, Zabars, and Koutsourelakis (2017)]

[Felsberger and Koutsourelakis, Wed.]

$$\underbrace{p(\mathbf{z})}_{\text{CV.}} \xrightarrow{p(\mathbf{x}|\mathbf{z})} \bar{p}(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{z})\, p(\mathbf{z})\, d\mathbf{z}$$

mschoeberl@gmail.com

Proposed generative model [Jordan (1999)]

adopted from Predictive Coarse-Graining [Schöberl, Zabars, and Koutsourelakis (2017)]

[Felsberger and Koutsourelakis, Wed.]

$$\underbrace{p(\mathbf{z})}_{\text{CV.}} \xrightarrow{p(\mathbf{x}|\mathbf{z})} \bar{p}(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{z}) \; p(\mathbf{z}) \; d\mathbf{z}$$

- CVs $\mathbf{z}$ as latent generator for atomistic coordinates $\mathbf{x}$.

mschoeberl@gmail.com

Proposed generative model [Jordan (1999)]

adopted from Predictive Coarse-Graining [Schöberl, Zabars, and Koutsourelakis (2017)]

[Felsberger and Koutsourelakis, Wed.]

$$\underbrace{p(\mathbf{z})}_{\text{CV.}} \xrightarrow{p(\mathbf{x}|\mathbf{z})} \bar{p}(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{z})\; p(\mathbf{z})\; d\mathbf{z}$$

- CVs $\mathbf{z}$ as latent generator for atomistic coordinates $\mathbf{x}$.
- *Probabilistic* map from CVs $\mathbf{z}$ to the atomistic coordinates $\mathbf{x}$, via $p(\mathbf{x}|\mathbf{z})$.

mschoeberl@gmail.com

Proposed generative model [Jordan (1999)]

adopted from Predictive Coarse-Graining [Schöberl, Zabars, and Koutsourelakis (2017)]

[Felsberger and Koutsourelakis, Wed.]

$$\underbrace{p(\mathbf{z})}_{\text{CV space.}} \xrightarrow{p(\mathbf{x}|\mathbf{z})} \bar{p}(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{z})\, p(\mathbf{z})\, d\mathbf{z}$$

- CVs $\mathbf{z}$ as latent generator for atomistic coordinates $\mathbf{x}$.
- *Probabilistic* map from CVs $\mathbf{z}$ to the atomistic coordinates $\mathbf{x}$, via $p(\mathbf{x}|\mathbf{z})$.
- Posterior distribution $p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{p(\mathbf{x})}$ represents code/dictionary for mapping atomistic coordinates $\mathbf{x}$ into its lower dimensional CV embedding $\mathbf{z}$.
  (intractable, e.g. use approximate inference thus approximate posterior with $q(\mathbf{z}|\mathbf{x})$).

Methodology - Learning

mschoeberl@gmail.com

Parametrize $\underbrace{p_{\boldsymbol{\theta}}(\mathsf{z}|\boldsymbol{\theta})}_{\text{Model for CVs}}$, $\underbrace{p_{\boldsymbol{\theta}}(\mathsf{x}|\mathsf{z},\ \boldsymbol{\theta})}_{\text{Mapping given CVs to atomistic coordinates}}$

Methodology - Learning

mschoeberl@gmail.com

Parametrize $\underbrace{p_{\boldsymbol{\theta}}(\mathsf{z}|\boldsymbol{\theta})}_{\text{Model for CVs}}$, $\underbrace{p_{\boldsymbol{\theta}}(\mathsf{x}|\mathsf{z},\ \boldsymbol{\theta})}_{\text{Mapping given CVs to atomistic coordinates}}$

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

$$\log p(\mathsf{x}^{(i)}, \cdots, \mathsf{x}^{(N)}) = \sum_{i=1}^{N} \log p_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}),$$

with $\log p_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})$ for one datapoint $\boldsymbol{x}^{(i)}$:

Lower bound the log-likelihood

[Dempster et al. (1977)]

$$\begin{aligned}
\log p_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) &= \log \int p_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}|\boldsymbol{z}^{(i)}) p_{\boldsymbol{\theta}}(\boldsymbol{z}^{(i)})\, d\mathsf{z} \\
&= \log \int q_{\boldsymbol{\phi}}(\boldsymbol{z}^{(i)}|\boldsymbol{x}^{(i)}) \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}|\boldsymbol{z}^{(i)}) p_{\boldsymbol{\theta}}(\boldsymbol{z}^{(i)})}{q_{\boldsymbol{\phi}}(\boldsymbol{z}^{(i)}|\boldsymbol{x}^{(i)})}\, d\mathsf{z} \\
&\geq \int q_{\boldsymbol{\phi}}(\boldsymbol{z}^{(i)}|\boldsymbol{x}^{(i)}) \log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}|\boldsymbol{z}^{(i)}) p_{\boldsymbol{\theta}}(\boldsymbol{z}^{(i)})}{q_{\boldsymbol{\phi}}(\boldsymbol{z}^{(i)}|\boldsymbol{x}^{(i)})}\, d\mathsf{z} = \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \boldsymbol{x}^{(i)})
\end{aligned}$$

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)]

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

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)]

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

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

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

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)]

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

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

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

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)]

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

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

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

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)]

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

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

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

Discovery of CVs as approximate Bayesian inference, i.e. identify the code/dictionary $q_{\phi}(\mathrm{z}|\mathrm{x})$.

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

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

- MAP estimate

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

- Approximate Bayesian posterior of decoding parameters $\boldsymbol{\theta}$, $p(\boldsymbol{\theta}|\mathbf{X})$, with Laplace approximation.

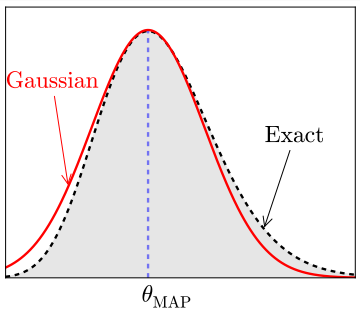
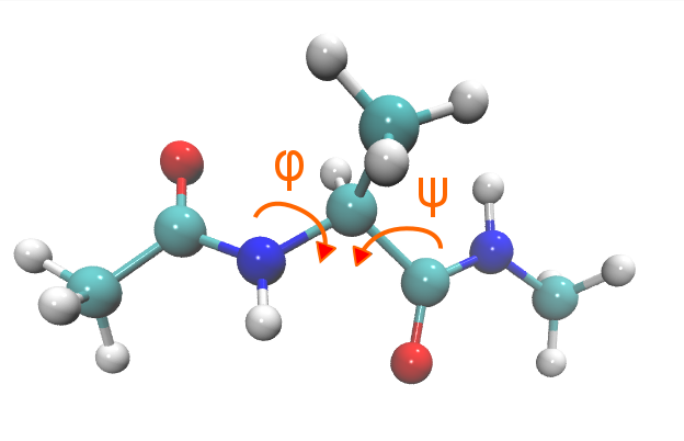


Figure 2: Laplace approximation: $p(\boldsymbol{\theta}|\mathbf{X}) \approx \mathcal{N}(\boldsymbol{\mu}, S)$

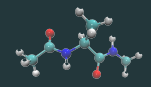
- $p(\boldsymbol{\theta}|\mathbf{X}) \approx \mathcal{N}(\boldsymbol{\mu}, S)$
- $\boldsymbol{\mu} = \boldsymbol{\theta}_{MAP}$
- $S^{-1} = -\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{X})}{\partial \theta_k \partial \theta_l} - \frac{\partial^2 \log p(\boldsymbol{\theta})}{\partial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

Numerical Illustration - Alanine Dipeptide

mschoeberl@gmail.com



- 22 atoms $\rightarrow \dim(\mathbf{x}) = 66$.
- Implicit solvent.
- *No* pre-processing of data the data: $\boldsymbol{x}^{(i)}$ are the cartesian coordinates of all atoms in the system.



mschoeberl@gmail.com

Auto-Encoding Variational Bayes

According [Kingma and Welling (2013)] .

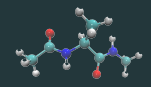
$$\bar{p}(\mathsf{x}|\boldsymbol{\theta}) = \int p(\mathsf{x}|\mathsf{z},\boldsymbol{\theta})\, p(\mathsf{z}|\boldsymbol{\theta})\, d\mathsf{z}$$

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

$$p(\mathsf{x}|\mathsf{z},\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}(\mathsf{z};\boldsymbol{\theta}),\ \boldsymbol{\sigma}^2 I)$$

With,

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



mschoeberl@gmail.com

Auto-Encoding Variational Bayes

According [Kingma and Welling (2013)] .

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

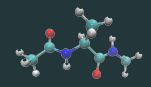
Mapping $\mathsf{z} \to \mathsf{x}$ (decoder):

$$p(\mathsf{x}|\mathsf{z},\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}(\mathsf{z};\boldsymbol{\theta}),\ \boldsymbol{\sigma}^2 I)$$

With,

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

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



mschoeberl@gmail.com

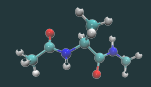
Auto-Encoding Variational Bayes

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

We assume CVs are normal distributed according,

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

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



mschoeberl@gmail.com

Auto-Encoding Variational Bayes

The approximate posterior is of the following form,

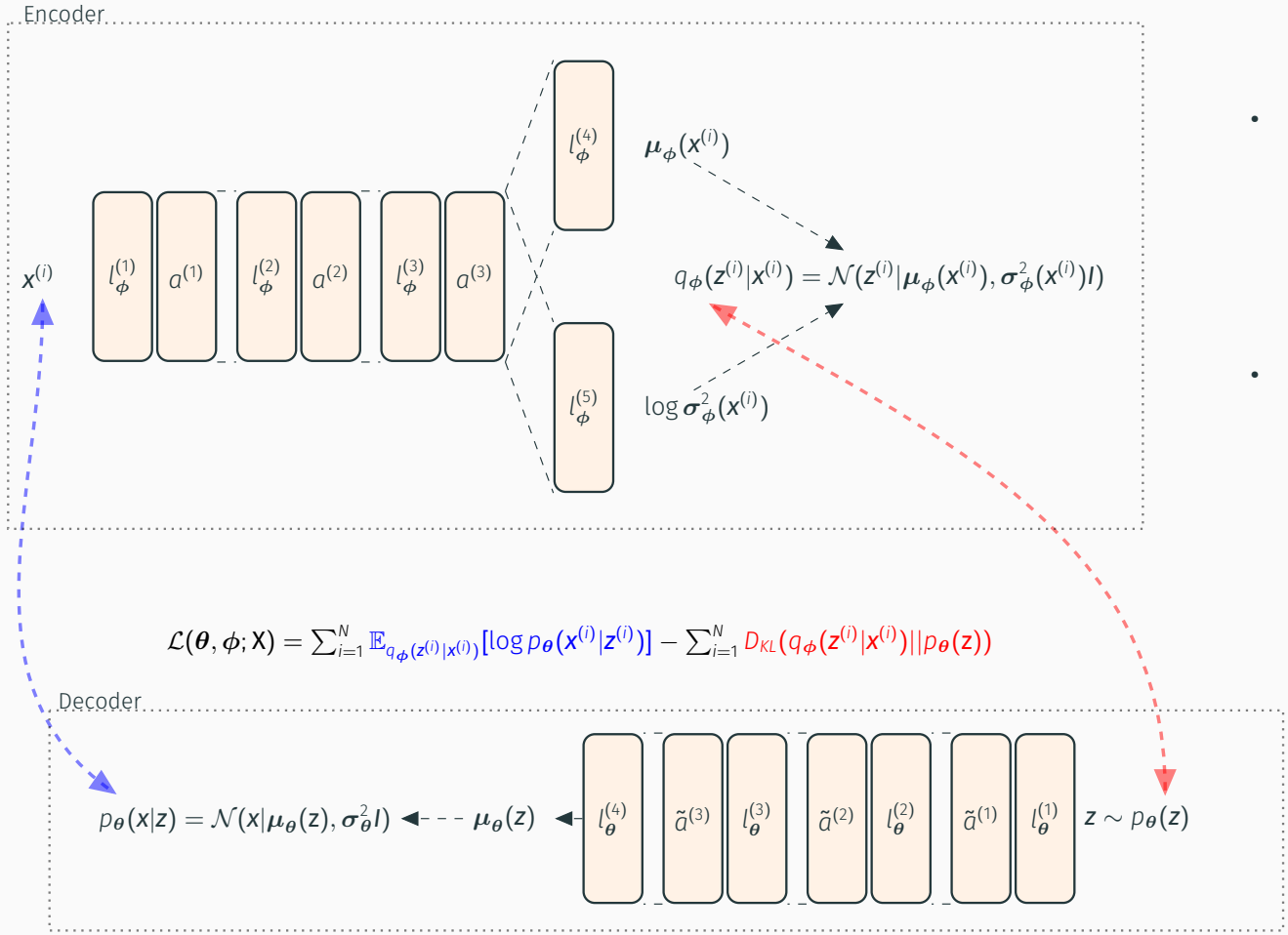
$$q(\mathsf{z}|\mathsf{x},\phi) = \mathcal{N}(\boldsymbol{\mu}(\mathsf{x};\phi), \boldsymbol{\sigma}(\mathsf{x})^2 I).$$

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 $\boldsymbol{\sigma}^2(\mathsf{x};\phi)$.

Numerical Illustration - Model

mschoeberl@gmail.com



- $\dim(\mathbf{x}) = 66$, $\dim(\mathbf{z}) = 2$
- Output dimensions of linear layers $l^i_{\phi,\psi} = \{100, 50, 100\}$
- Activation functions $a^i = \{selu, logsig, logsig\}$, $\tilde{a}^i = \{selu, tanh, tanh\}$

Figure 3: Schematic draft of the Bayesian variational autoencoder employed [Kingma and Welling (2013); Rezende et al. (2014)] .

Numerical Illustration - Training

mschoeberl@gmail.com

Training with $N = 1500$ datapoints, a batch size of $M = 64$ for 1500 epochs, $\dim(\mathbf{z}) = 2$

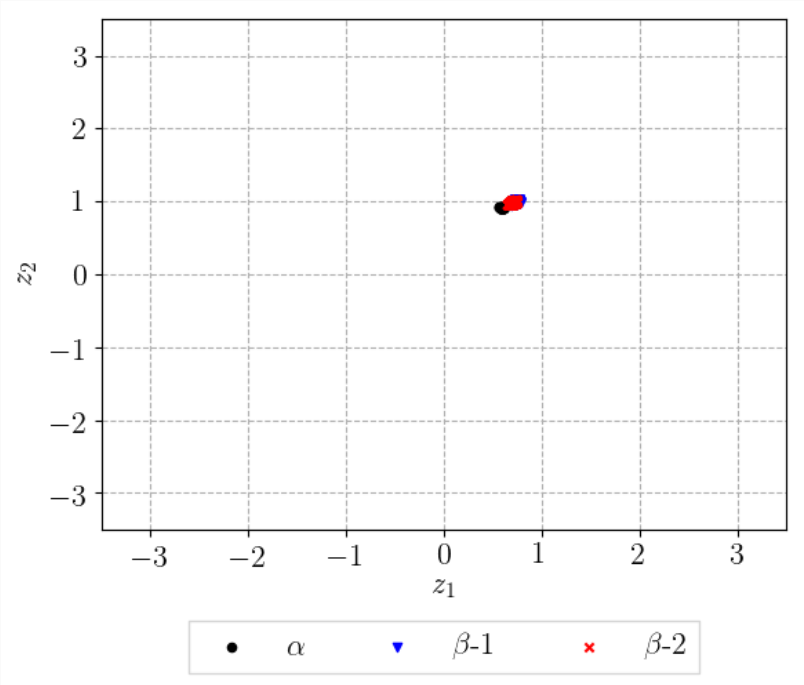


Figure 4: CVs assigned to training data (latent representation of training data). Discovered CVs separate conformational modes well.

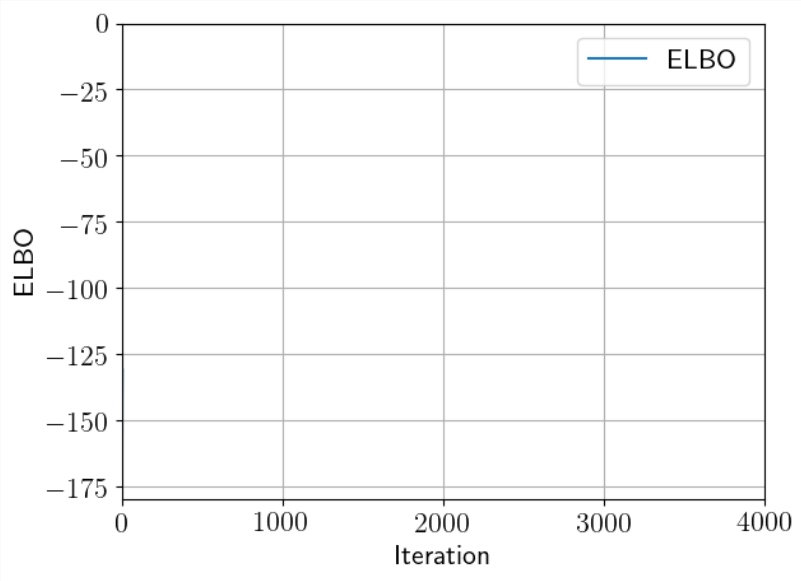


Figure 5: Lower bound.

Numerical Illustration - Prediction

mschoeberl@gmail.com

Prediction for $\{\mathbf{z}|z_1 = [-4, 4], z_2 = 0\}$.

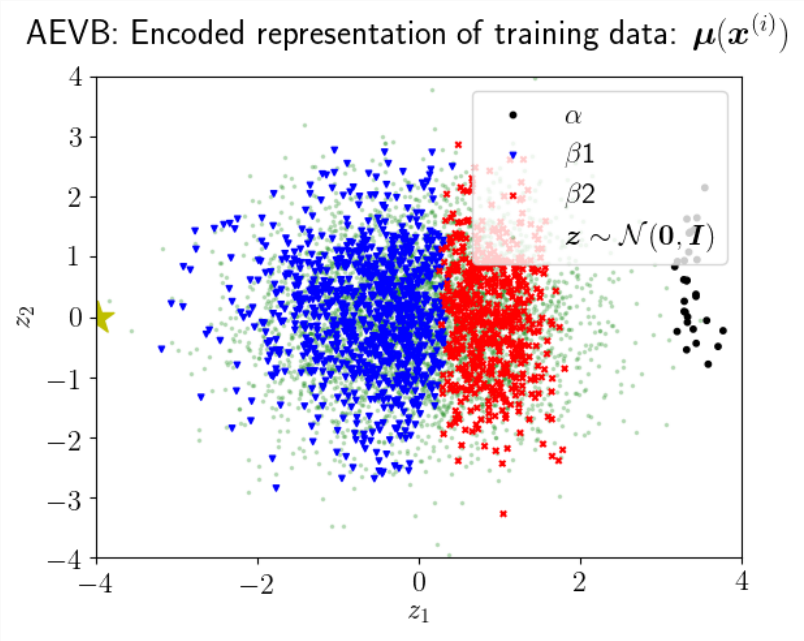


Figure 6: CVs are highly correlated with different conformations.

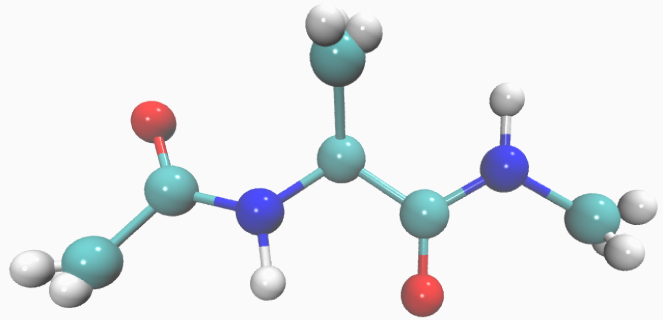


Figure 7: Mean prediction for $\mathbf{x}$ with decoder $p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})$, given $\{\mathbf{z}|z_1 = [-4, 4], z_2 = 0\}$.

Numerical Illustration - Prediction

mschoeberl@gmail.com

Prediction for $\{\mathbf{z}|z_1 = [-4, 4], z_2 = 0\}$.

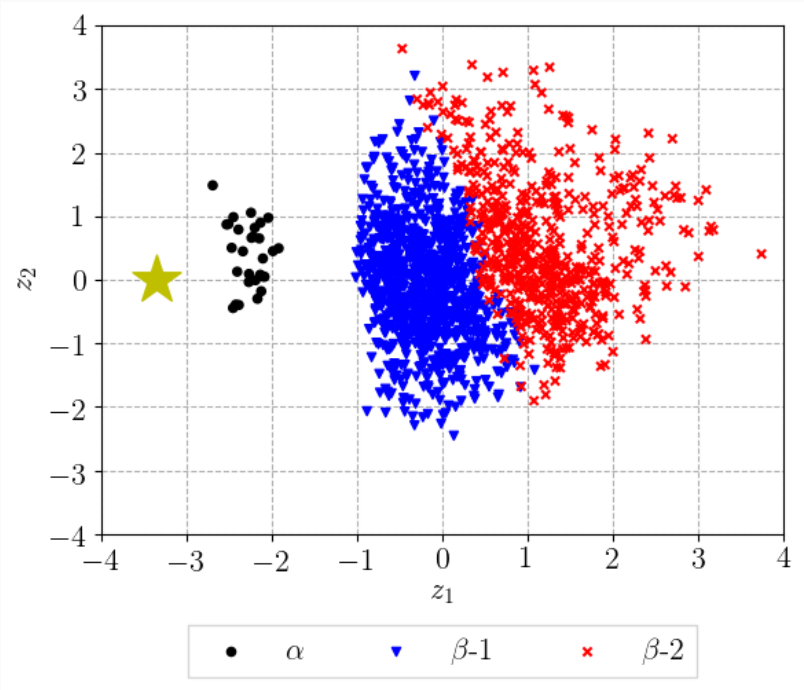


Figure 6: CVs are highly correlated with different conformations.

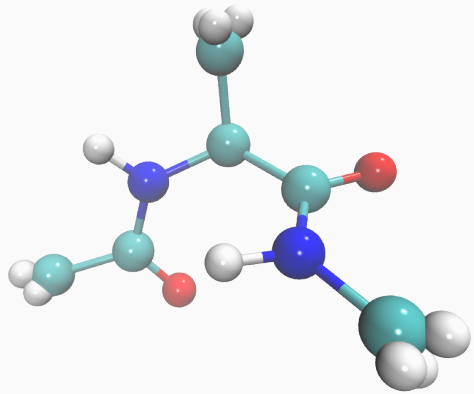


Figure 7: Mean prediction for $\mathbf{x}$ with decoder $p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})$, given $\{\mathbf{z}|z_1 = [-4, 4], z_2 = 0\}$.

mschoeberl@gmail.com

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

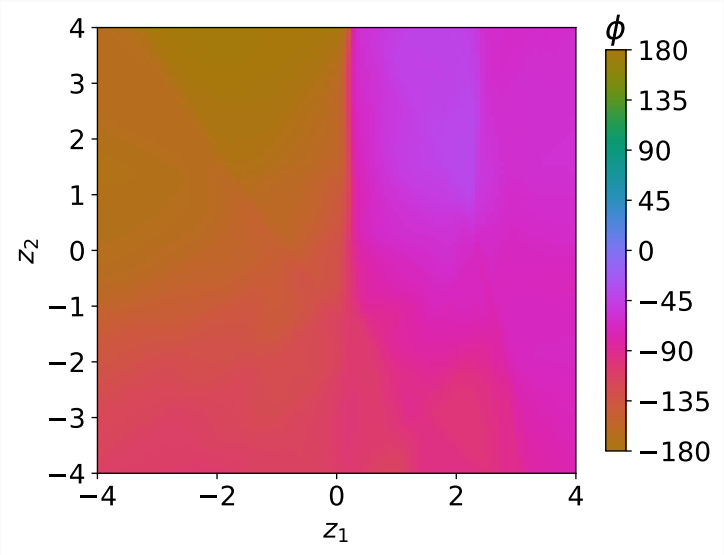


Figure 8: Assessing dihedral angles ϕ given the CVs $\mathbf{z}$.

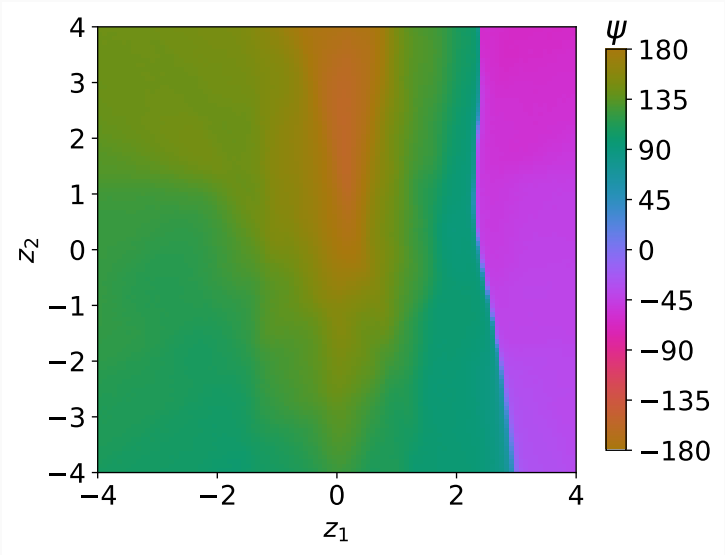
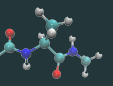


Figure 9: Assessing dihedral angles ψ given the CVs $\mathbf{z}$.

Numerical Illustration - *Physical* interpretation?



mschoeberl@gmail.com

- $(z_1, z_2) \mapsto (\phi, \psi)$

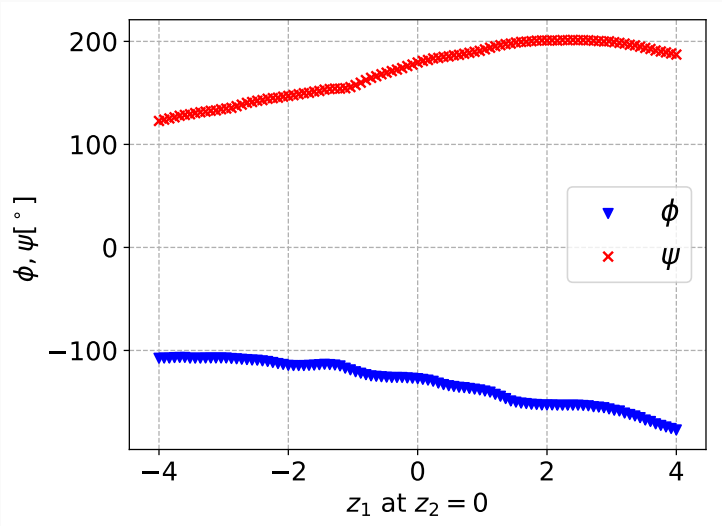


Figure 10: Assessing dihedral angles (ϕ, ψ) given the CVs $\mathbf{z}$ at $z_2 = 0$.

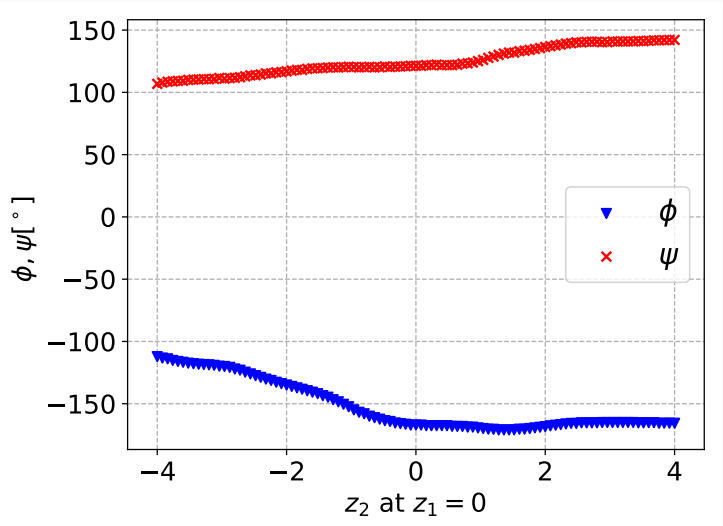
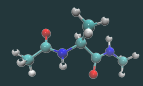


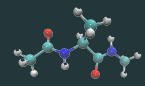
Figure 11: Assessing dihedral angles (ϕ, ψ) given the CVs $\mathbf{z}$ at $z_1 = 0$.



mschoeberl@gmail.com

Motivation

- Network structure highly individual. $\dim(\boldsymbol{\theta})$ grows exponentially.
- What are the actually required parameters $\boldsymbol{\theta}$?
- Can one search *across models*?



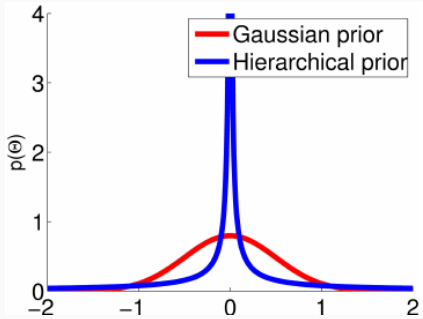
mschoeberl@gmail.com

Motivation

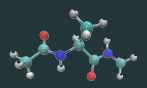
- Network structure highly individual. $\dim(\boldsymbol{\theta})$ grows exponentially.
- What are the actually required parameters $\boldsymbol{\theta}$?
- Can one search *across models*?

Sparsity-enforcing - Hierarchical priors (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 of the derivative of the log-likelihood: $\frac{\partial}{\partial \tau_{c,j}} = - < \tau_j > \theta_j$

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 $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$.

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 $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$.

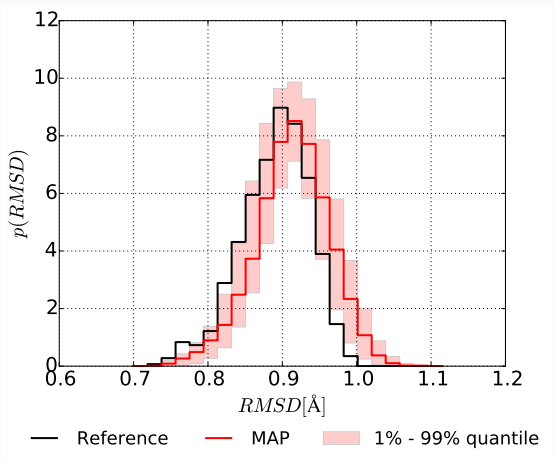


Figure 12: Root-mean-square deviation.

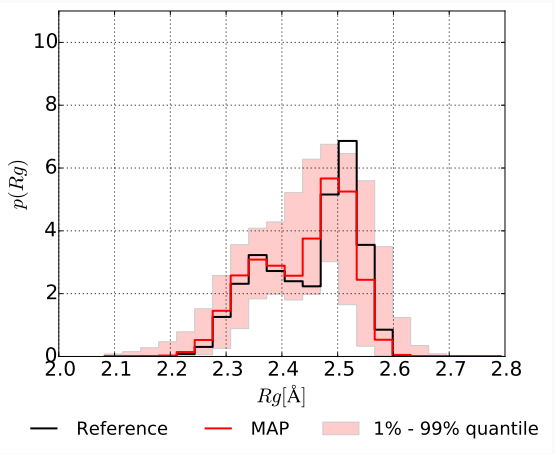


Figure 13: Radius of gyration.

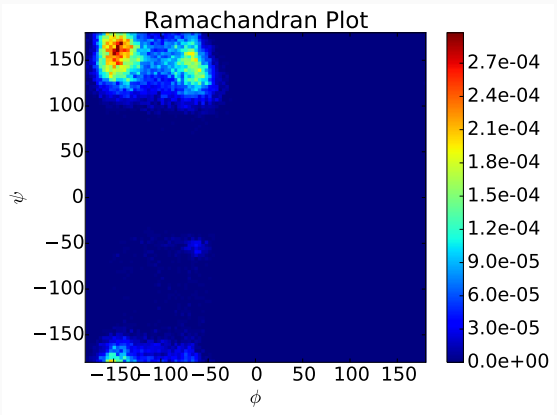
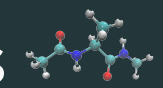


Figure 14: Ramachandran plot

ARD-prior identifies a parsimonious representation: 60% of the $\boldsymbol{\theta}$'s set to zero.

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 $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$.
- $N = \{526, 1527, 4004\}$

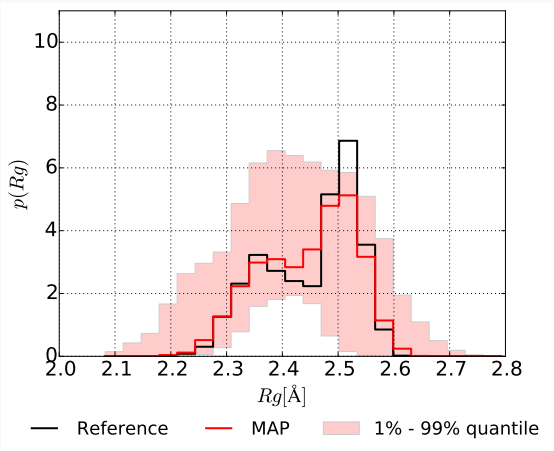


Figure 15: Radius of gyration. $N = 526$.

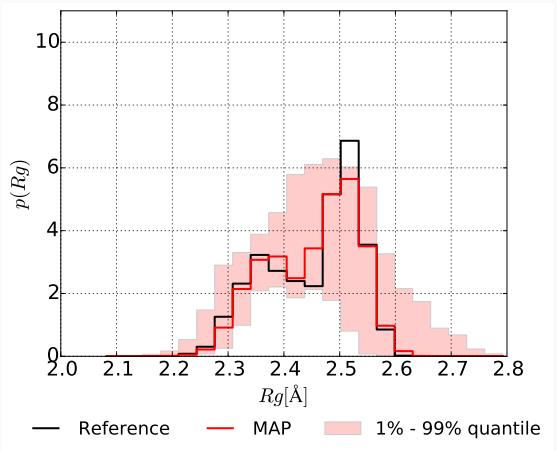


Figure 16: Radius of gyration. $N = 1527$.

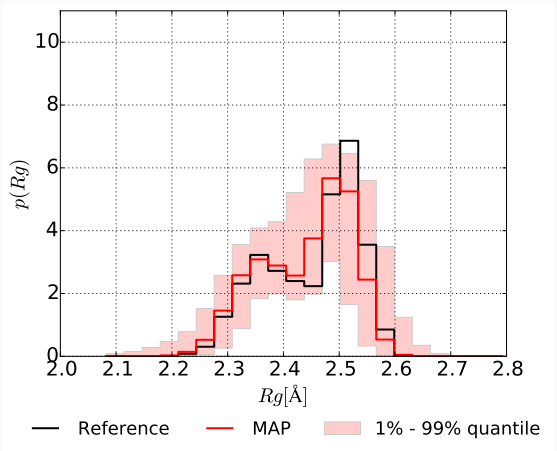


Figure 17: Radius of gyration. $N = 4004$.

Adequate predictions with a dataset of $N = 526$. *Low* data instead of big data.

Summary

mschoeberl@gmail.com

Conclusion

- CV discovery as *Bayesian* inference.
- Identified CVs reveal physics based on low data.
- *Physics learned from data*: discovered CVs are attached with physicochemical meaning.
- *Predictive* model obtained for property estimation and uncertainty quantification.

Outlook

- Sparse automatic determination of network structure.
- VI on network parameters $\boldsymbol{\theta}, \boldsymbol{\phi}$; Fully Bayesian approach.
- Utilize identified CVs for enhanced sampling methods (iteratively add data obtained from enhanced sampling).
- Addressing robustness and dependency of network [Mattei and Frellsen (2018)] .

Questions?

Thank you!

References

Banon.

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.

Michele Ceriotti, Gareth A. Tribello, and Michele Parrinello. Simplifying the representation of complex free-energy landscapes using sketch-map. *Proceedings of the National Academy of Sciences*, 108(32):13023–13028, 2011. ISSN 0027-8424. doi: 10.1073/pnas.1108486108. URL http://www.pnas.org/content/108/32/13023.

References II

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

R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences*, 102(21): 7426–7431, 2005. ISSN 0027-8424. doi: 10.1073/pnas.0500334102. URL http://www.pnas.org/content/102/21/7426.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977. ISSN 00359246. URL http://www.jstor.org/stable/2984875.

References III

Karl Pearson F.R.S. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901. doi: 10.1080/14786440109462720. URL https://doi.org/10.1080/14786440109462720. PCA beginnings.

H. Hotelling. Analysis of a complex of statistical variables into principal components. *J. Educ. Psych.*, 24, 1933.

Michael I. Jordan, editor. *Learning in Graphical Models*. MIT Press, Cambridge, MA, USA, 1999. ISBN 0-262-60032-3.

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

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

References IV

Boaz Nadler, Stéphane Lafon, Ronald R. Coifman, and Ioannis G. Kevrekidis. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Applied and Computational Harmonic Analysis*, 21(1):113 – 127, 2006. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2005.07.004. URL http://www.sciencedirect.com/science/article/pii/S1063520306000534. Special Issue: Diffusion Maps and Wavelets.

Frank Noé and Cecilia Clementi. Kinetic distance and kinetic maps from molecular dynamics simulation. *Journal of Chemical Theory and Computation*, 11(10):5002–5011, 2015. doi: 10.1021/acs.jctc.5b00553. URL https://doi.org/10.1021/acs.jctc.5b00553. PMID: 26574285.

References V

Frank Noé, Ralf Banisch, and Cecilia Clementi. Commute maps: Separating slowly mixing molecular configurations for kinetic modeling. *Journal of Chemical Theory and Computation*, 12(11): 5620–5630, 2016. doi: 10.1021/acs.jctc.6b00762. URL https://doi.org/10.1021/acs.jctc.6b00762. PMID: 27696838.

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.

References VI

Mary A. Rohrdanz, Wenwei Zheng, and Cecilia Clementi. Discovering mountain passes via torchlight: Methods for the definition of reaction coordinates and pathways in complex macromolecular reactions. *Annual Review of Physical Chemistry*, 64(1):295–316, 2013. doi: 10.1146/annurev-physchem-040412-110006. URL https://doi.org/10.1146/annurev-physchem-040412-110006. PMID: 23298245.

Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000. ISSN 0036-8075. doi: 10.1126/science.290.5500.2319. URL http://science.sciencemag.org/content/290/5500/2319.

mschoeberl@gmail.com

John M. Troyer and Fred E. Cohen. Protein conformational landscapes: Energy minimization and clustering of a long molecular dynamics trajectory. *Proteins: Structure, Function, and Bioinformatics*, 23(1):97–110. doi: 10.1002/prot.340230111. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.340230111.