Bayesian Coarse-Graining in Atomistic Simulations:

Adaptive Identification of the Dimensionality and Salient Features

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

[1] Continuum Mechanics Group
Technical University of Munich

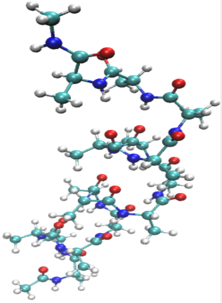
[2] Institute for Advanced Study
Technical University of Munich

[3] Institute of Computational Sciences and Informatics
University of Notre Dame

SIAM CSE, Atlanta, GA, USA
March 1st 2017

Problem Definition - Equilibrium Statistical Mecha

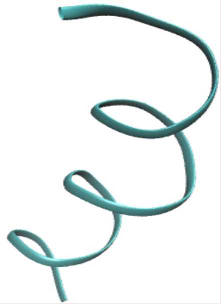
Fine-Grained Model (FG)



$$\boxed{p_f(\boldsymbol{x}) \propto e^{-\beta U_f(\boldsymbol{x})}}$$

- $\boldsymbol{x} \in \mathcal{M}$: fine-scale DOFs
- $U_f(\boldsymbol{x})$: atomistic potential
- Observables:
  $\mathbb{E}_{p_f(\boldsymbol{x})}[a] = \int a(\boldsymbol{x})\, p_f(\boldsymbol{x})\, d\boldsymbol{x}$

Coarse-Grained Model (CG)



$$\boldsymbol{X} = \mathcal{R}(\boldsymbol{x}), \quad dim(\boldsymbol{X}) << dim(\boldsymbol{x})$$

- $\boldsymbol{X}$: *coarse*-scale dofs
- $\mathcal{R}$: *restriction* operator (mapping *fine* $\rightarrow$ *coarse*)

Questions

- What are good coarse-grained variables $\boldsymbol{X}$ (how many, how are they related to the FG description?)
- What is the right CG model?
- Given a good CG model for X, how much can one predict about the whole x (reconstruction)?
- How much information is lost during coarse-graining and how does this affect predictions produced by the CG model?
- Given finite simulation data at the fine-scale, how (un)certain can one be in their predictions?

Questions

- What are good coarse-grained variables $\boldsymbol{X}$ (how many, how are they related to the FG description?)
- What is the right CG model?
- Given a good CG model for X, how much can one predict about the whole x (reconstruction)?
- How much information is lost during coarse-graining and how does this affect predictions produced by the CG model?
- Given finite simulation data at the fine-scale, how (un)certain can one be in their predictions?

Motivation

Questions

- What are good coarse-grained variables $\boldsymbol{X}$ (how many, how are they related to the FG description?)
- What is the right CG model?
- Given a good CG model for $\mathbf{X}$, how much can one predict about the whole $\boldsymbol{x}$ (reconstruction)?
- How much information is lost during coarse-graining and how does this affect predictions produced by the CG model?
- Given finite simulation data at the fine-scale, how (un)certain can one be in their predictions?

Questions

- What are good coarse-grained variables $\boldsymbol{X}$ (how many, how are they related to the FG description?)
- What is the right CG model?
- Given a good CG model for $\mathbf{X}$, how much can one predict about the whole $\boldsymbol{x}$ (reconstruction)?
- How much information is lost during coarse-graining and how does this affect predictions produced by the CG model?
- Given finite simulation data at the fine-scale, how (un)certain can one be in their predictions?

Motivation

Questions

- What are good coarse-grained variables $\boldsymbol{X}$ (how many, how are they related to the FG description?)
- What is the right CG model?
- Given a good CG model for $\mathbf{X}$, how much can one predict about the whole $\boldsymbol{x}$ (reconstruction)?
- How much information is lost during coarse-graining and how does this affect predictions produced by the CG model?
- Given finite simulation data at the fine-scale, how (un)certain can one be in their predictions?

Motivation

Two roads in CG:

1.) Variational (Mean Field, and many others)

$$min_{\bar{p}_f(\boldsymbol{x})} \; KL(\bar{p}_f(\boldsymbol{x}) \, || p_f(\boldsymbol{x}))$$

2.) **Data-driven** (e.g. Relative Entropy [Shell (2008)]):

$$\boxed{min_{\bar{p}_f(\boldsymbol{x})} \; KL(p_f(\boldsymbol{x}) \, || \bar{p}_f(\boldsymbol{x}))}$$

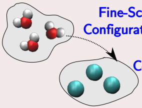
where:

- $\bar{p}_f(\boldsymbol{x})$: approximation
- $p_f(\boldsymbol{x}) \propto e^{-\beta U_f(\boldsymbol{x})}$: exact

Motivation

Existing methods

$$\underbrace{p_f(\boldsymbol{x})}_{\textit{fine}} \xrightarrow{\mathcal{R}(\boldsymbol{x})=\boldsymbol{X}} \underbrace{\bar{p}_c(\boldsymbol{X})}_{\textit{coarse}}$$



Proposed (Generative model)

$$\underbrace{p_c(\boldsymbol{X})}_{\textit{coarse}} \xrightarrow{p_{cf}(\boldsymbol{x}|\boldsymbol{X})} \underbrace{\bar{p}_f(\boldsymbol{x}) = \int p_{cf}(\boldsymbol{x}|\boldsymbol{X})\; p_c(\boldsymbol{X})\; d\boldsymbol{X}}_{\textit{fine}}$$

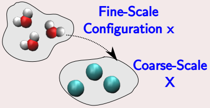
Notes

- No restriction operator (fine-to-coarse $\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{X}$).
- A probabilistic *coarse-to-fine* map $p_{cf}(\boldsymbol{x}|\boldsymbol{X})$ is prescribed
- The coarse model $p_c(\boldsymbol{X})$ is not the marginal of $\boldsymbol{X}$ (given $\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{X}$)

Motivation

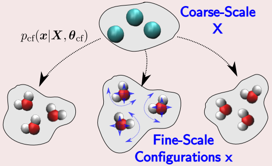
Existing methods

$$\underbrace{p_f(\boldsymbol{x})}_{fine} \xrightarrow{\mathcal{R}(\boldsymbol{x})=\boldsymbol{X}} \underbrace{\bar{p}_c(\boldsymbol{X})}_{coarse}$$



Proposed (Generative model)

$$\underbrace{p_c(\boldsymbol{X})}_{coarse} \xrightarrow{p_{cf}(\boldsymbol{x}|\boldsymbol{X})} \underbrace{\bar{p}_f(\boldsymbol{x}) = \int p_{cf}(\boldsymbol{x}|\boldsymbol{X})\; p_c(\boldsymbol{X})\; d\boldsymbol{X}}_{fine}$$



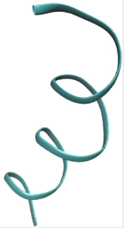
Notes

- No restriction operator (fine-to-coarse $\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{X}$).
- A probabilistic *coarse-to-fine* map $p_{cf}(\boldsymbol{x}|\boldsymbol{X})$ is prescribed
- The coarse model $p_c(\boldsymbol{X})$ is not the marginal of $\boldsymbol{X}$ (given $\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{X}$)

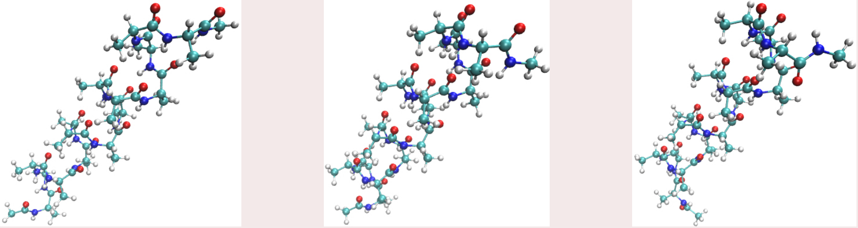
Motivation

Given $p_c(\boldsymbol{X})$ and $p_{cf}(\boldsymbol{x}|\boldsymbol{X})$:

1) Draw $\mathbf{X}$ from $p_c(\boldsymbol{X})$ (i.e. simulate CG model)



2) Draw $\boldsymbol{x}$ from $p_{cf}(\boldsymbol{x}|\boldsymbol{X})$



Proposed Probabilistic Generative model

- Parametrize: $\underbrace{p_c(\boldsymbol{X}|\boldsymbol{\theta}_c)}_{\text{coarse model}}, \quad \underbrace{p_{cf}(\boldsymbol{x}|\boldsymbol{X},\ \boldsymbol{\theta}_{cf})}_{\text{coarse}\longrightarrow\text{fine map}}$

- Optimize:

$$
\begin{aligned}
&\min_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} KL(p_f(\boldsymbol{x}) \ || \ \bar{p}_f(\boldsymbol{x}|\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf})) \\
\leftrightarrow &\min_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} -\int p_f(\boldsymbol{x}) \log \frac{\int p_{cf}(\boldsymbol{x}|\boldsymbol{X},\boldsymbol{\theta}_{cf})\ p_c(\boldsymbol{X}|\boldsymbol{\theta}_c)\ d\boldsymbol{X}}{p_f(\boldsymbol{x})}\ d\boldsymbol{x} \\
\leftrightarrow &\max_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} \int p_f(\boldsymbol{x}) \left(\log \int p_{cf}(\boldsymbol{x}|\boldsymbol{X},\boldsymbol{\theta}_{cf})\ p_c(\boldsymbol{X}|\boldsymbol{\theta}_c)\ d\boldsymbol{X}\right)\ d\boldsymbol{x} \\
\leftrightarrow &\max_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} \sum_{i=1}^{N} \log \int p_{cf}(\boldsymbol{x}^{(i)}|\boldsymbol{X},\boldsymbol{\theta}_{cf})\ p_c(\boldsymbol{X}|\boldsymbol{\theta}_c)\ d\boldsymbol{X} \\
\leftrightarrow &\max_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} \mathcal{L}(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}), \quad \text{(MLE)}
\end{aligned}
$$

- MAP estimate: $\max_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} \mathcal{L}(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}) + \underbrace{\log p(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf})}_{log-prior}$

- Fully Bayesian i.e. posterior: $p(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}|x^{(1:N)}) \propto exp\{\mathcal{L}(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf})\ p(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf})\}$

Learning

Proposed Probabilistic Generative model

- Parametrize: $\underbrace{p_c(\boldsymbol{X}|\boldsymbol{\theta}_c)}_{\text{coarse model}}, \quad \underbrace{p_{cf}(\boldsymbol{x}|\boldsymbol{X},\ \boldsymbol{\theta}_{cf})}_{\text{coarse}\rightarrow\text{fine map}}$

- Optimize:

$$
\begin{aligned}
&\min_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} KL(p_f(\boldsymbol{x}) \mid\mid \bar{p}_f(\boldsymbol{x}|\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf})) \\
\leftrightarrow &\min_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} -\int p_f(\boldsymbol{x}) \log \frac{\int p_{cf}(\boldsymbol{x}|\boldsymbol{X},\boldsymbol{\theta}_{cf})\ p_c(\boldsymbol{X}|\boldsymbol{\theta}_c)\ d\boldsymbol{X}}{p_f(\boldsymbol{x})}\ d\boldsymbol{x} \\
\leftrightarrow &\max_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} \int p_f(\boldsymbol{x}) \left(\log \int p_{cf}(\boldsymbol{x}|\boldsymbol{X},\boldsymbol{\theta}_{cf})\ p_c(\boldsymbol{X}|\boldsymbol{\theta}_c)\ d\boldsymbol{X}\right)\ d\boldsymbol{x} \\
\leftrightarrow &\max_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} \sum_{i=1}^{N} \log \int p_{cf}(\boldsymbol{x}^{(i)}|\boldsymbol{X},\boldsymbol{\theta}_{cf})\ p_c(\boldsymbol{X}|\boldsymbol{\theta}_c)\ d\boldsymbol{X} \\
\leftrightarrow &\max_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} \mathcal{L}(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}), \quad \text{(MLE)}
\end{aligned}
$$

- MAP estimate: $\max_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} \mathcal{L}(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}) + \underbrace{\log p(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf})}_{log-prior}$

- Fully Bayesian i.e. posterior: $p(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}|\boldsymbol{x}^{(1:N)}) \propto exp\{\mathcal{L}(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf})\ p(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf})\}$

Proposed Probabilistic Generative model

- Parametrize: $\underbrace{p_c(\boldsymbol{X}|\boldsymbol{\theta}_c)}_{\text{coarse model}}, \quad \underbrace{p_{cf}(\boldsymbol{x}|\boldsymbol{X}, \ \boldsymbol{\theta}_{cf})}_{\text{coarse}\rightarrow\text{fine map}}$

- Optimize:

$$\begin{aligned}
&\min_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} KL(p_f(\boldsymbol{x}) \ || \ \bar{p}_f(\boldsymbol{x}|\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf})) \\
\leftrightarrow &\min_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} - \int p_f(\boldsymbol{x}) \log \frac{\int p_{cf}(\boldsymbol{x}|\boldsymbol{X},\boldsymbol{\theta}_{cf}) \ p_c(\boldsymbol{X}|\boldsymbol{\theta}_c) \ d\boldsymbol{X}}{p_f(\boldsymbol{x})} \ d\boldsymbol{x} \\
\leftrightarrow &\max_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} \int p_f(\boldsymbol{x}) \left(\log \int p_{cf}(\boldsymbol{x}|\boldsymbol{X},\boldsymbol{\theta}_{cf}) \ p_c(\boldsymbol{X}|\boldsymbol{\theta}_c) \ d\boldsymbol{X}\right) \ d\boldsymbol{x} \\
\leftrightarrow &\max_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} \sum_{i=1}^N \log \int p_{cf}(\boldsymbol{x}^{(i)}|\boldsymbol{X},\boldsymbol{\theta}_{cf}) \ p_c(\boldsymbol{X}|\boldsymbol{\theta}_c) \ d\boldsymbol{X} \\
\leftrightarrow &\max_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} \ \mathcal{L}(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}), \quad \text{(MLE)}
\end{aligned}$$

- MAP estimate: $\max_{\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}} \ \mathcal{L}(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}) + \underbrace{\log p(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf})}_{log-prior}$

- Fully Bayesian i.e. posterior: $p(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}|\boldsymbol{x}^{(1:N)}) \propto exp\{\mathcal{L}(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf}) \ p(\boldsymbol{\theta}_c,\boldsymbol{\theta}_{cf})\}$

Stochastic VB-Expectation-Maximization [Beal & Ghahramani 2003]

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}_c, \boldsymbol{\theta}_{cf}) &= \sum_{i=1}^N \log \int p_{cf}(\boldsymbol{x}^{(i)}|\boldsymbol{X}^{(i)}, \boldsymbol{\theta}_{cf})\, p_c(\boldsymbol{X}^{(i)}|\boldsymbol{\theta}_c)\, d\boldsymbol{X}^{(i)} \\
&= \sum_{i=1}^N \log \int q(\boldsymbol{X}^{(i)}) \frac{p_{cf}(\boldsymbol{x}^{(i)}|\boldsymbol{X}^{(i)}, \boldsymbol{\theta}_{cf})\, p_c(\boldsymbol{X}^{(i)}|\boldsymbol{\theta}_c)}{q(\boldsymbol{X}^{(i)})}\, d\boldsymbol{X}^{(i)} \\
&\geq \sum_{i=1}^N \int q(\boldsymbol{X}^{(i)}) \log \frac{p_{cf}(\boldsymbol{x}^{(i)}|\boldsymbol{X}^{(i)}, \boldsymbol{\theta}_{cf})\, p_c(\boldsymbol{X}^{(i)}|\boldsymbol{\theta}_c)}{q(\boldsymbol{X}^{(i)})}\, d\boldsymbol{X}^{(i)} \\
&= \sum_{i=1}^N \mathcal{F}_i(q(\boldsymbol{X}^{(i)}),\ \boldsymbol{\theta}_c, \boldsymbol{\theta}_{cf}) = \mathcal{F}(\boldsymbol{q},\ \boldsymbol{\theta}_c, \boldsymbol{\theta}_{cf})
\end{aligned}$$

- E-step: Approximate $q_i^{opt}(\boldsymbol{X}^{(i)})$ using a multivariate Gaussians:

$$q_i(\boldsymbol{X}^{(i)}) = \mathcal{N}(\boldsymbol{\mu}_i^{opt}, \boldsymbol{\Sigma}_i^{opt})$$

- M-step: Compute gradients $\sum_{i=1}^N \nabla_{\boldsymbol{\theta}_c}\mathcal{F}$, $\sum_{i=1}^N \nabla_{\boldsymbol{\theta}_{cf}}\mathcal{F}$, (and Hessian) and update $(\boldsymbol{\theta}_c, \boldsymbol{\theta}_{cf})$

Stochastic VB-Expectation-Maximization [Beal & Ghahramani 2003]

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}_c, \boldsymbol{\theta}_{cf}) &= \sum_{i=1}^N \log \int p_{cf}(\boldsymbol{x}^{(i)}|\boldsymbol{X}^{(i)}, \boldsymbol{\theta}_{cf})\ p_c(\boldsymbol{X}^{(i)}|\boldsymbol{\theta}_c)\ d\boldsymbol{X}^{(i)} \\
&= \sum_{i=1}^N \log \int q(\boldsymbol{X}^{(i)}) \frac{p_{cf}(\boldsymbol{x}^{(i)}|\boldsymbol{X}^{(i)}, \boldsymbol{\theta}_{cf})\ p_c(\boldsymbol{X}^{(i)}|\boldsymbol{\theta}_c)}{q(\boldsymbol{X}^{(i)})}\ d\boldsymbol{X}^{(i)} \\
&\geq \sum_{i=1}^N \int q(\boldsymbol{X}^{(i)}) \log \frac{p_{cf}(\boldsymbol{x}^{(i)}|\boldsymbol{X}^{(i)}, \boldsymbol{\theta}_{cf})\ p_c(\boldsymbol{X}^{(i)}|\boldsymbol{\theta}_c)}{q(\boldsymbol{X}^{(i)})}\ d\boldsymbol{X}^{(i)} \\
&= \sum_{i=1}^N \mathcal{F}_i(q(\boldsymbol{X}^{(i)}),\ \boldsymbol{\theta}_c, \boldsymbol{\theta}_{cf}) = \mathcal{F}(\boldsymbol{q},\ \boldsymbol{\theta}_c, \boldsymbol{\theta}_{cf})
\end{aligned}
$$

- E-step: Approximate $q_i^{opt}(\boldsymbol{X}^{(i)})$ using a multivariate Gaussians:

$$q_i(\boldsymbol{X}^{(i)}) = \mathcal{N}(\boldsymbol{\mu}_i^{opt}, \boldsymbol{\Sigma}_i^{opt})$$

- M-step: Compute gradients $\sum_{i=1}^N \nabla_{\boldsymbol{\theta}_c}\mathcal{F}$, $\sum_{i=1}^N \nabla_{\boldsymbol{\theta}_{cf}}\mathcal{F}$, (and Hessian) and update $(\boldsymbol{\theta}_c, \boldsymbol{\theta}_{cf})$

Essential Ingredient: Stochastic Optimization

ADAptive Moment estimation (ADAM, [Kingma & Ba 2014])

Learning

- Exponential-family distributions:

$$p_c(\boldsymbol{X}|\theta_c) = \exp\left\{ \underbrace{\theta_c^T \phi(\boldsymbol{X})}_{\text{CG potential } U_c} - A(\theta_c) \right\} \qquad (e^{A(\boldsymbol{\theta}_c)} = \int e^{\boldsymbol{\theta}_c^T \phi(\boldsymbol{X})} \, d\boldsymbol{X})$$

$$p_{cf}(\boldsymbol{x}|\boldsymbol{X}, \boldsymbol{\theta}_{cf}) = \exp\{\boldsymbol{\theta}_{cf}^T \psi(\boldsymbol{x}, \boldsymbol{X}) - B(\boldsymbol{X}, \boldsymbol{\theta}_{cf})\} \quad (e^{B(\boldsymbol{X}, \boldsymbol{\theta}_{cf})} = \int e^{\boldsymbol{\theta}_{cf}^T \psi(\boldsymbol{x}, \boldsymbol{X})} \, d\boldsymbol{x})$$

- Gradients:

$$\nabla_{\boldsymbol{\theta}_c} \mathcal{F} = \sum_{i=1}^N < \phi(\boldsymbol{X}^{(i)}) >_{q_i(\boldsymbol{X}^{(i)})} - N < \phi(\boldsymbol{X}) >_{p_c(\boldsymbol{X}|\boldsymbol{\theta}_c)}$$

$$\nabla_{\boldsymbol{\theta}_{cf}} \mathcal{F} = \sum_{i=1}^N (< \psi(\boldsymbol{x}^{(i)}, \boldsymbol{X}^{(i)}) >_{q_i(\boldsymbol{X}^{(i)})} - < \psi(\boldsymbol{x}, \boldsymbol{X}^{(i)}) >_{p_{cf}(\boldsymbol{x}|\boldsymbol{X}^{(i)}, \boldsymbol{\theta}_{cf}) q_i(\boldsymbol{X}^{(i)})})$$

- Hessian:

$$\left. \begin{array}{l} \nabla^2_{\boldsymbol{\theta}_c} \mathcal{F} = -N \, Cov_{p_c(\boldsymbol{X}|\boldsymbol{\theta}_c)}[\phi(\boldsymbol{X})] \\ \nabla^2_{\boldsymbol{\theta}_{cf}} \mathcal{F} = -\sum_{i=1}^N Cov_{p_{cf}(\boldsymbol{x}|\boldsymbol{X}^{(i)}, \boldsymbol{\theta}_{cf}) q_i(\boldsymbol{X}^{(i)})}[\psi(\boldsymbol{x}, \boldsymbol{X}^{(i)})] \end{array} \right\} \longrightarrow \text{Concave}$$

Learning

- Exponential-family distributions:

$$p_c(\boldsymbol{X}|\theta_c) = \exp\left\{ \underbrace{\boldsymbol{\theta}_c^T \phi(\boldsymbol{X})}_{\text{CG potential } U_c} - A(\boldsymbol{\theta}_c) \right\} \qquad (e^{A(\boldsymbol{\theta}_c)} = \int e^{\boldsymbol{\theta}_c^T \phi(\boldsymbol{X})} \, d\boldsymbol{X})$$

$$p_{cf}(\boldsymbol{x}|\boldsymbol{X}, \boldsymbol{\theta}_{cf}) = \exp\{\boldsymbol{\theta}_{cf}^T \psi(\boldsymbol{x}, \boldsymbol{X}) - B(\boldsymbol{X}, \boldsymbol{\theta}_{cf})\} \quad (e^{B(\boldsymbol{X}, \boldsymbol{\theta}_{cf})} = \int e^{\boldsymbol{\theta}_{cf}^T \psi(\boldsymbol{x}, \boldsymbol{X})} \, d\boldsymbol{x})$$

- Gradients:

$$\nabla_{\boldsymbol{\theta}_c} \mathcal{F} = \sum_{i=1}^N < \phi(\boldsymbol{X}^{(i)}) >_{q_i(\boldsymbol{X}^{(i)})} - N < \phi(\boldsymbol{X}) >_{p_c(\boldsymbol{X}|\boldsymbol{\theta}_c)}$$
$$\nabla_{\boldsymbol{\theta}_{cf}} \mathcal{F} = \sum_{i=1}^N (< \psi(\boldsymbol{x}^{(i)}, \boldsymbol{X}^{(i)}) >_{q_i(\boldsymbol{X}^{(i)})} - < \psi(\boldsymbol{x}, \boldsymbol{X}^{(i)}) >_{p_{cf}(\boldsymbol{x}|\boldsymbol{X}^{(i)}, \boldsymbol{\theta}_{cf}) q_i(\boldsymbol{X}^{(i)})})$$

- Hessian:

$$\left.\begin{array}{l} \nabla^2_{\boldsymbol{\theta}_c} \mathcal{F} = -N \, Cov_{p_c(\boldsymbol{X}|\boldsymbol{\theta}_c)}[\phi(\boldsymbol{X})] \\ \nabla^2_{\boldsymbol{\theta}_{cf}} \mathcal{F} = -\sum_{i=1}^N Cov_{p_{cf}(\boldsymbol{x}|\boldsymbol{X}^{(i)}, \boldsymbol{\theta}_{cf}) q_i(\boldsymbol{X}^{(i)})}[\psi(\boldsymbol{x}, \boldsymbol{X}^{(i)})] \end{array}\right\} \longrightarrow \text{Concave}$$

Learning

- MAP-estimates:

$$\max_{\boldsymbol{\theta}_c, \boldsymbol{\theta}_{cf}} \mathcal{L}(\boldsymbol{\theta}_c, \boldsymbol{\theta}_{cf}) + \underbrace{\log p(\boldsymbol{\theta}_c, \boldsymbol{\theta}_{cf})}_{log-prior}$$

- Approximate Bayesian posterior using Laplace approximation

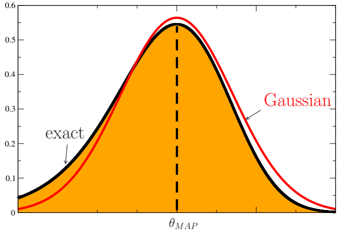


Figure : Laplace approximation: $p(\boldsymbol{\theta}|\boldsymbol{x}^{(1:N)}) \approx \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{S})$

where:

- $\boldsymbol{\mu} = \boldsymbol{\theta}_{,MAP}$
- $\boldsymbol{S}^{-1} = \begin{bmatrix} N Cov_{p_c(\boldsymbol{X}|\boldsymbol{\theta}_c)}[\phi(\boldsymbol{X}), \phi_l(\boldsymbol{X})] & 0 \\ 0 & \sum_{i=1}^{N} Cov_{p_{cf}(\mathbf{q}|\boldsymbol{X}^{(i)}, \boldsymbol{\theta}_{cf}) q_i(\boldsymbol{X}^{(i)})}[\psi(\mathbf{q}, \boldsymbol{X})] \end{bmatrix}$

Adaptive feature learning

Which feature functions $\phi(\boldsymbol{X})$ one use?

$$\text{CG potential:}\quad U_c(\mathbf{X}) = \theta_c^T \phi(\boldsymbol{X}) \rightarrow p_c(\mathbf{X}) \propto e^{U_c(\mathbf{X})}$$

- Option 1: Use as many as possible in combination with a sparsity-enforcing prior [Schöberl et al, JCP 2017].
    - Suitable when $\mathbf{X}$ have a clear, physical meaning.
- Option 2: Consider a parametrized family $\Phi_z = \{\phi(\mathbf{X};\boldsymbol{z})\}$ and greedily add the best member of this (i.e. optimize $\boldsymbol{z}$). I.e. suppose:

$$U_c(\mathbf{X}) = \theta_c^T \phi(\boldsymbol{X}) + \theta_{c,new}\phi(\mathbf{X};\boldsymbol{z})$$

Then, the largest expected decrease in $KL(p_f(\boldsymbol{x})\,||\bar{p}_f(\boldsymbol{x}))$ is:

$$\arg\max_{\boldsymbol{z}} \left(\sum_{i=1}^{N} < \phi(\mathbf{X};\boldsymbol{z}) >_{q_i(\mathbf{X})} - < \phi(\mathbf{X};\boldsymbol{z}) >_{p_c(\mathbf{X})} \right)^2$$

Adaptive feature learning

Which feature functions $\phi(\boldsymbol{X})$ one use?

$$\text{CG potential: } U_c(\mathbf{X}) = \boldsymbol{\theta}_c^T \phi(\boldsymbol{X}) \rightarrow p_c(\mathbf{X}) \propto e^{U_c(\mathbf{X})}$$

- Option 1: Use as many as possible in combination with a sparsity-enforcing prior [Schöberl et al, JCP 2017].
  - Suitable when $\mathbf{X}$ have a clear, physical meaning.
- Option 2: Consider a parametrized family $\Phi_z = \{\phi(\mathbf{X}; \boldsymbol{z})\}$ and greedily add the best member of this (i.e. optimize $\boldsymbol{z}$). I.e. suppose:

$$U_c(\mathbf{X}) = \boldsymbol{\theta}_c^T \phi(\boldsymbol{X}) + \boldsymbol{\theta}_{c,new} \phi(\mathbf{X}; \boldsymbol{z})$$

Then, the largest expected decrease in $KL(p_f(\boldsymbol{x}) \,||\, \bar{p}_f(\boldsymbol{x}))$ is:

$$\arg\max_{\boldsymbol{z}} \left(\sum_{i=1}^{N} < \phi(\mathbf{X}; \boldsymbol{z}) >_{q_i(\mathbf{X})} - < \phi(\mathbf{X}; \boldsymbol{z}) >_{p_c(\mathbf{X})} \right)^2$$

Which feature functions $\phi(\boldsymbol{X})$ one use?

$$\text{CG potential:} \quad U_c(\mathbf{X}) = \boldsymbol{\theta}_c^T \phi(\boldsymbol{X}) \rightarrow p_c(\mathbf{X}) \propto e^{U_c(\mathbf{X})}$$

- Option 1: Use as many as possible in combination with a sparsity-enforcing prior [Schöberl et al, JCP 2017].
  - Suitable when $\mathbf{X}$ have a clear, physical meaning.
- Option 2: Consider a parametrized family $\Phi_z = \{\phi(\mathbf{X}; \boldsymbol{z})\}$ and greedily add the best member of this (i.e. optimize $\boldsymbol{z}$). I.e. suppose:

$$U_c(\mathbf{X}) = \boldsymbol{\theta}_c^T \phi(\boldsymbol{X}) + \boldsymbol{\theta}_{c,new} \phi(\mathbf{X}; \boldsymbol{z})$$

Then, the largest expected decrease in $KL(p_f(\boldsymbol{x}) \,||\bar{p}_f(\boldsymbol{x}))$ is:

$$\arg\max_{\boldsymbol{z}} \left(\sum_{i=1}^{N} < \phi(\mathbf{X}; \boldsymbol{z}) >_{q_i(\mathbf{X})} - < \phi(\mathbf{X}; \boldsymbol{z}) >_{p_c(\mathbf{X})} \right)^2$$

Numerical Illustrations - Alanine Dipeptide

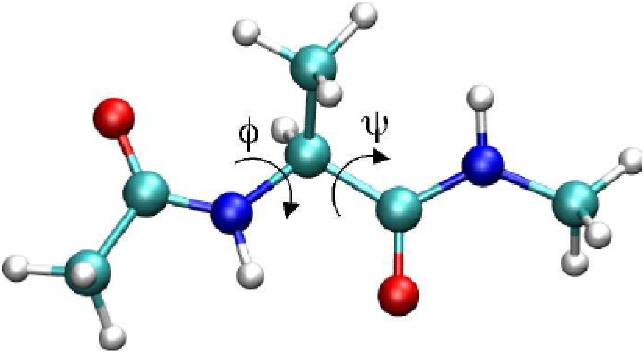


Figure : Alanine Dipeptide [Bonomi et al,CPC , 2009]

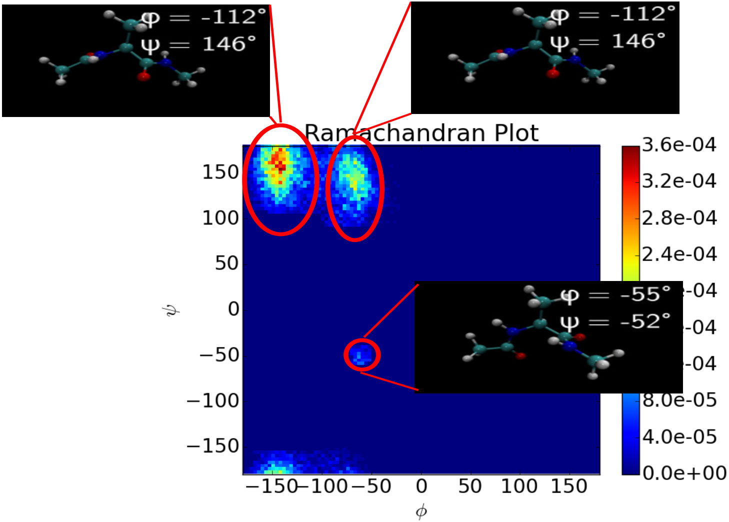
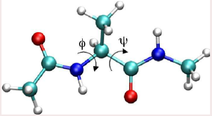


Figure : Ramachandran plot for Alanine Dipeptide with respect to dihedral angles ϕ, ψ.

Numerical Illustrations - Alanine Dipeptide

Proposed *Global* Model



- **Coarse-Grained model:**

$$\boxed{p_c(\mathbf{X}) \propto e^{-\beta U_c(\mathbf{X})}, \quad U_c(\mathbf{X}) = \boldsymbol{\theta}_c^T \phi(\boldsymbol{X})}$$

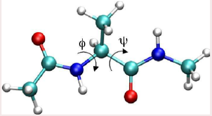
We assume:

- $\mathbf{X} \in [0,1]^{n_X}, \; n_X = dim(\mathbf{X})$
- radial basis functions $\phi(\mathbf{X}; \boldsymbol{z}) = e^{-\sum_{k=1}^{n_X} \tau_k (X_k - X_{0,k})^2}$ where $\boldsymbol{z} = \{\tau_k, X_{0,k}\}_{j=1}^{n_X}$

- **Coarse-to-Fine map:**

$$\boxed{p_{cf}(\mathbf{x}|\mathbf{X}) = \mathcal{N}(\boldsymbol{\mu} + \mathbf{W}\mathbf{X}, \; \mathbf{S})}, \quad \boldsymbol{\theta}_{cf} = \{\boldsymbol{\mu}, \mathbf{W}, \mathbf{S}\}$$

Proposed *Global* Model



- **Coarse-Grained model:**

$$\boxed{p_c(\mathbf{X}) \propto e^{-\beta U_c(\mathbf{X})}, \quad U_c(\mathbf{X}) = \boldsymbol{\theta}_c^T \phi(\boldsymbol{X})}$$

We assume:

- $\mathbf{X} \in [0,1]^{n_X}, \; n_X = dim(\mathbf{X})$
- radial basis functions $\phi(\mathbf{X}; \boldsymbol{z}) = e^{-\sum_{k=1}^{n_X} \tau_k (X_k - X_{0,k})^2}$ where $\boldsymbol{z} = \{\tau_k, X_{0,k}\}_{j=1}^{n_X}$

- **Coarse-to-Fine map:**

$$\boxed{p_{cf}(\boldsymbol{x}|\mathbf{X}) = \mathcal{N}(\boldsymbol{\mu} + \boldsymbol{W}\mathbf{X}, \; \boldsymbol{S})}, \quad \boldsymbol{\theta}_{cf} = \{\boldsymbol{\mu}, \boldsymbol{W}, \boldsymbol{S}\}$$

Numerical Illustrations - Alanine Dipeptide

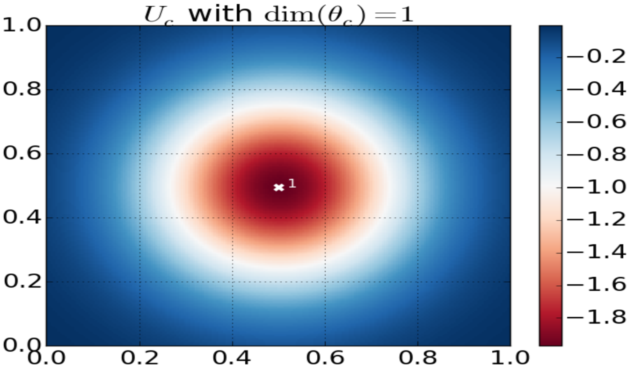


Figure : $L = 1$

$$\boxed{U_c = \sum_{l=1}^{L} \theta_{c,l}\ \phi_l(\mathbf{X}),\ \ dim(\mathbf{X}) = 2}$$

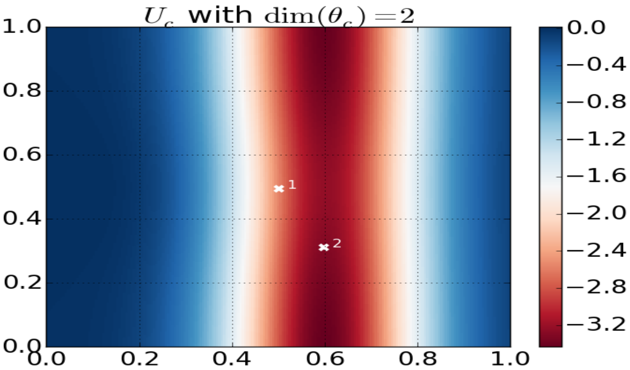


Figure : $L = 2$

$$\boxed{U_c = \sum_{l=1}^{L} \theta_{c,l}\ \phi_l(\mathbf{X}),\ dim(\mathbf{X}) = 2}$$

Numerical Illustrations - Alanine Dipeptide

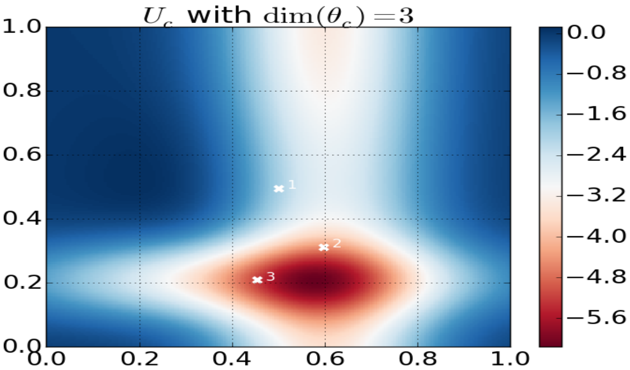


Figure : $L = 3$

$$\boxed{U_c = \sum_{l=1}^{L} \theta_{c,l}\ \phi_l(\mathbf{X}),\ dim(\mathbf{X}) = 2}$$

Numerical Illustrations - Alanine Dipeptide

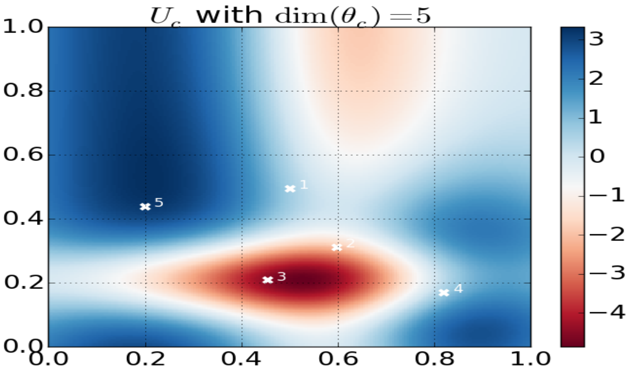


Figure : $L = 5$

$$\boxed{U_c = \sum_{l=1}^{L} \theta_{c,l}\ \phi_l(\mathbf{X}),\ \ dim(\mathbf{X}) = 2}$$

Numerical Illustrations - Alanine Dipeptide

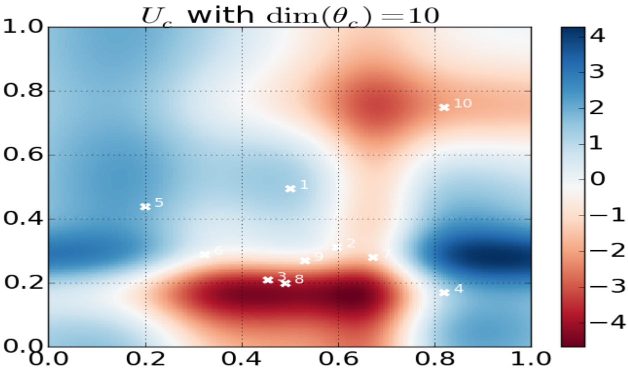


Figure : $L = 10$

$$\boxed{U_c = \sum_{l=1}^{L} \theta_{c,l}\ \phi_l(\mathbf{X}),\ \ dim(\mathbf{X}) = 2}$$

Numerical Illustrations - Alanine Dipeptide

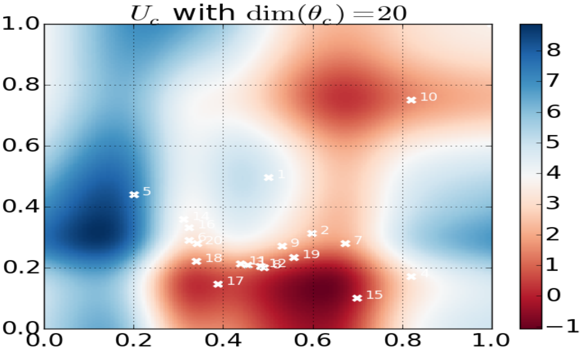


Figure : $L = 20$

$$\boxed{U_c = \sum_{l=1}^{L} \theta_{c,l}\ \phi_l(\mathbf{X}),\ \ dim(\mathbf{X}) = 2}$$

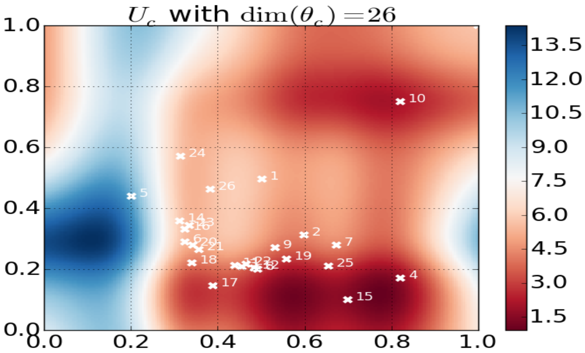


Figure : $L = 26$

$$\boxed{U_c = \sum_{l=1}^{L} \theta_{c,l}\ \phi_l(\mathbf{X}),\ \ dim(\mathbf{X}) = 2}$$

Numerical Illustrations - Alanine Dipeptide

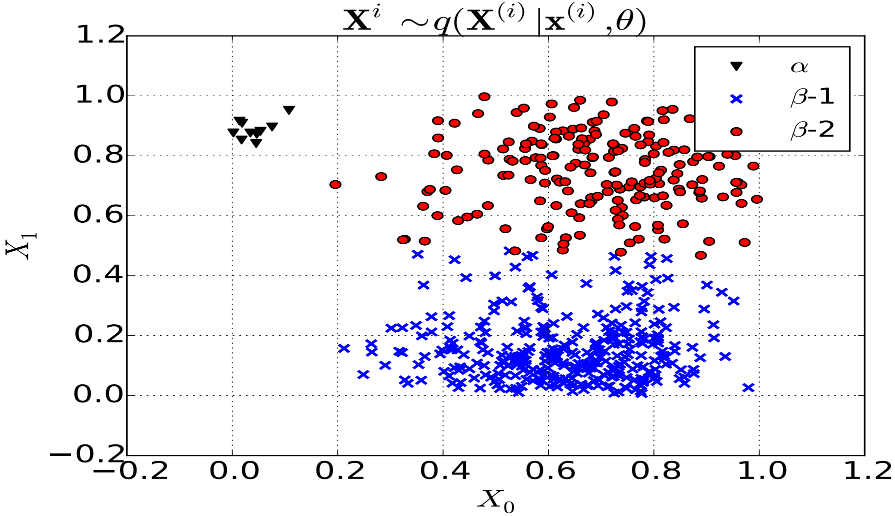


Figure : Visualization in (Latent) CG-variable Space **X**

Probabilistic Predictions of Macroscopic Properties

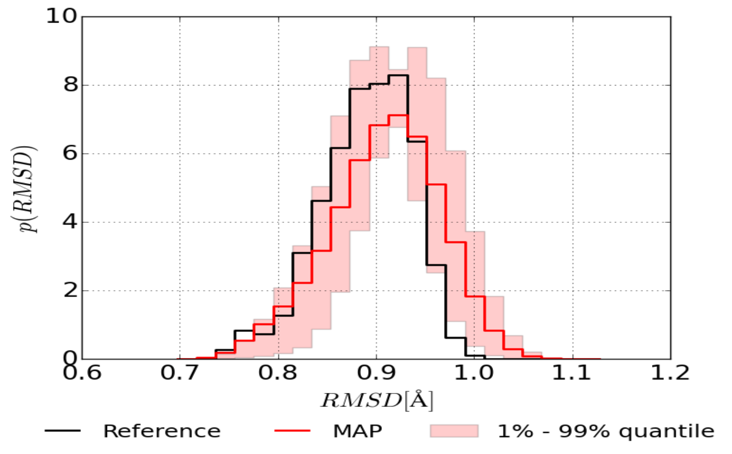


Figure : Root-mean-squared (RMSD) deviation from an α-helical conformation

Conclusions

Summary

- A *generative probabilistic model is proposed*
- It consists of a CG-density and a probabilistic *coarse* $\rightarrow$ *fine* map.
- Can account for information loss due to CG
- Can quantify *predictive uncertainty* in fine-scale observables.
- Can be used for *model* selection.

Conclusions

Summary

- A *generative probabilistic model is proposed*
- It consists of a CG-density and a probabilistic *coarse* $\rightarrow$ *fine* map.
- Can account for information loss due to CG
- Can quantify *predictive uncertainty* in fine-scale observables.
- Can be used for *model* selection.

Outlook

- Explore alternative definitions of coarse variables $\boldsymbol{X}$ and alternative *coarse* $\rightarrow$ *fine* maps p_{cf} e.g.:
  - Discrete states indicating Free-Energy wells
  - Hierachical *coarse-graining*:

$$\bar{p}_f(\boldsymbol{x}) = \int p_{cf}(\boldsymbol{x}|\boldsymbol{X}_1)\; p_c(\boldsymbol{X}_1|\boldsymbol{X}_2)\; p_c(\boldsymbol{X}_2|\boldsymbol{X}_3)\ldots p_c(\boldsymbol{X}_M)\; d\boldsymbol{X}_1\ldots\boldsymbol{X}_M$$

- Fully Bayesian or Variational Bayesian