Predictive Coarse-Graining

M. Schöberl\textsuperscript{1}, N. Zabaras\textsuperscript{2,3}, P.S. Koutsourelakis\textsuperscript{1}

\textsuperscript{1}Continuum Mechanics Group
Technische Universität München

\textsuperscript{2}Institute for Advanced Study
Technische Universität München

\textsuperscript{3}Warwick Centre of Predictive Modeling
University of Warwick

November 27, 2015
Motivation Coarse-Graining

Atomistic simulation for obtaining insights of chemical and physical process of complex systems.

**Difficulty**

- Complex interactions
- Long-range interactions
- Small time- and length-scales

→ Exceeding computational tractability

**A coarse description allows us**

- to evaluate larger systems during larger time intervals
- to gain understanding of physics of the system.

Figure: Coarse-graining water
Motivation

Approach

- Describing system with less degrees of freedom
- Determining optimal parameter set for given parametrization of a coarse-grained potential
- Leading to point estimates of its parametrization and thus also in predictions
- Predictions performed on coarse scale

How can we quantify the uncertainty induced by a coarse description and the loss of information?

How can we reconstruct fine configurations given a coarse configuration?
General coarse-graining problem

Fine-scale degrees of freedom \( x \in \mathcal{M} \) with \( \mathcal{M} \subset \mathbb{R}^n \), \( n \gg 1 \) in equilibrium described by a Boltzmann-type PDF:

\[
p_f(x|\beta) = \frac{\exp \{- \beta U(x)\}}{Z(\beta)}
\]

- Potential \( U(x) \)
- Inverse Temperature \( \beta = \frac{1}{k_b T} \), with temperature \( T \)
- Partition function \( Z(\beta) = \int_{\mathcal{M}} \exp \{- \beta U(x)\} \, dx \)
General coarse-graining problem

Coarse-scale description

\[ p_c(X|\beta) = \frac{\exp\{-\beta U_c(X, \theta_c)\}}{Z_c(\theta_c, \beta)} \]

- Coarse variables \( X \in M_c, M_c \subseteq \mathbb{R}^{n_c}, n_c \ll n \)
- Coarse-grained potential selected \( \hat{U}_c \) selected out of candidate potentials \( \hat{U}_c \in U_c \)
- Parametrization and shape selected by assuming \( U_c(X, \theta_c) \) but any potential possible

How are fine and coarse configurations connected?
Connecting fine variables \( x \) with coarse variables \( X \) with coarse-graining map

\[ \xi : M \rightarrow M_c \]

\[ X = \xi(x) \]
How to determine the effective coarse potential $U_c$?

**Analytical solution** for the effective coarse-grained potential:

$$U_c(X) = -\beta^{-1} \ln \int_M \exp \{-\beta U(x)\} \delta(\xi(x) - X) dx$$

Potential of mean force with respect to coarse-grained coordinates $\xi(x)$. → intractable!

**Numerical** coarse-graining strategies:

- iterative Boltzmann inversion [5]
- inverse Monte Carlo [8]
- force matching [6]
- variational mean-field theory [9]
- relative entropy [10]
The relative entropy method\[10\]

→ minimize a distance metric between $p_f(X|\beta)$ and $p_c(X|\beta)$ with the PDF $p_f(X|\beta)$ sampling configuration $x$ that maps to configuration $X$

$$p_f(X|\beta) \propto \int_{\mathcal{M}} p_f(x|\beta) \delta(\xi(x) - X) dx$$

Relative entropy method\[10, 4, 2\] (or KL-divergence)

$$\text{KL}(p_f||p_c) = \int_{\mathcal{M}_c} p_f(X|\beta) \ln \left( \frac{p_f(X|\beta)}{p_c(X|\beta)} \right) dX$$

Formulation in $\mathcal{M}$:

$$\text{KL}(p_f||p_c) = \int_{\mathcal{M}} p_f(x|\beta) \ln \left( \frac{p_f(x|\beta)}{p_c(\xi(x)|\beta)} \right) dx + S_{\text{map}}(\xi(x))$$

▷ with $S_{\text{map}} = \left\langle \int_{\mathcal{M}} \delta(\xi(x) - X) dx \right\rangle_{p_f(x|\beta)}$ measures the information loss induced by mapping $\xi(x)$. 
Relative entropy method

$S_{\text{map}}(\xi(x))$ is independent of $\hat{U}_c$, the minimization of the KL-divergence is equivalent to minimization of

$$\mathcal{F}(\hat{U}_c) = \left\langle \ln \frac{p_f(x|\beta)}{p_c(\xi(x)|\beta)} \right\rangle_{p_f(x|\beta)}$$

For any given $m$-dimensional parametrization $\theta_c \in \Theta_c \subset \mathbb{R}^m$ of the potential, the optimization problem follows:

$$\theta_c^* = \arg \min_{\theta_c \in \Theta_c} \mathcal{F}(U_c(X, \theta_c))$$
Relative entropy method

Model Restrictions

- Relative entropy method has no capability to predict fine-scale configurations (no mapping coarse to fine)
  - Eg. for predicting correlations within fine configurations
- Predicting quantities of interest (QoI) by evaluating on fine-scale
  \[
  \langle f(x) \rangle_{p_f(x|\beta)} = \int_{\mathcal{M}} f(x) p_f(x|\beta) dx
  \]
- With coarse-grained description \( p_c(X|\beta) \) predictions in \( \mathcal{M}_c \) but not in \( \mathcal{M} \)
  - Difficulty:
    - How to define observable \( g(X) \) for desired QoI in \( \mathcal{M}_c \)?
    - Are you able to predict same QoI with \( \langle g(X) \rangle_{p_c(X|\beta)} = \int_{\mathcal{M}_c} g(X) p_c(X|\beta) dX \)?
- No quantification of epistemic uncertainty: determine \( S_{\text{map}}(\xi(x)) \)
Proposed Model

Framework following **generative probabilistic models** and builds upon:

1. Coarse-scale PDF \( p_c(X|\theta_c, \beta) \propto \exp\{ -\beta U_c(X, \theta_c) \} \)
   - Describing statistics of coarse variables \( X \)
   - Parametrized by \( \theta_c \), no restrictions in shape of \( p_c \)

2. Probabilistic mapping from **coarse to fine** \( p_{cf}(x|X, \theta_{cf}) \)
   - Conditional PDF of \( x \) given the coarse variables \( X \)
   - Parametrization of the probabilistic mapping by \( \theta_{cf} \in \Theta_{cf} \subset \mathbb{R}^{m_{cf}} \)
   - Deterministic mapping \( \xi(x) \) is not invertible (many to one map)
     → Probabilistic relation necessary
   - \( S_{map} \) not fixed for per definition. → Optimization with respect to \( \theta_{cf} \) possible.

3. Prior PDF for model parametrization
   - \( p(\theta_c) \)
   - \( p(\theta_{cf}) \)
Proposed Model

- Proposed model describes a **Bayesian perspective** of coarse-graining equilibrium atomistic ensembles
- Model is data driven by $N$ data points $x^{(i)}$ in the fine-scale, denoted as $x^{(1:N)} = \{x^{(1)}, \ldots, x^{(N)}\}$

Given the data $x^{(1:N)}$ a posterior distribution on the model parameters $\theta_{cf}, \theta_c$ and the **latent** variables $X^{(1:N)}$ can be defined.

- Coarse variables seen as **latent**
- To each observable $x^{(i)}$ one latent Variable $X^{(i)}$ is assigned
- $X^{(i)}$ represents pre-image of the fine-scale observation $x^{(i)}$
Bayesian coarse-graining - posterior

Posterior distribution

\[ p(\theta_{cf}, \theta_c, X^{(1:N)} | x^{(1:N)}) \propto p(x^{(1:N)} | \theta_{cf}, \theta_c, X^{(1:N)}) p(\theta_{cf}, \theta_c, X^{(1:N)}) \]

\[ = \left( \prod_{i=1}^{N} p_{cf}(x^{(i)} | \theta_{cf}, X^{(i)}) p_c(X^{(i)} | \theta_c) \right) p(\theta_{cf}) p(\theta_c) \]

Use given data \( x^{(1:N)} \) for inferring the posterior distribution or for determining MAP estimates of the model parameters \( \theta_{cf}^*, \theta_c^* \).

- Bayesian formulation answers the questions of
  - model validation and
  - prediction
Bayesian coarse-graining

The posterior distribution leads to the predictive distribution for fine-scale configurations $x$:

**Predictive distribution**

$$p(x|x^{(1:N)}) = \int p_{cf}(x|X, \theta_{cf}) p_c(X|\theta_c) p(\theta_c, \theta_{cf}|x^{(1:N)}) dX d\theta_{cf} d\theta_{cf}$$

**Simulate fine-scale system**

1. Draw sample $\theta^*_c, \theta^*_{cf} \sim p(\theta_c, \theta_{cf}|x^{(1:N)})$
2. Draw sample of coarse-scale description $X^* \sim p_c(X|\theta^*_c)$
3. Draw sample of fine-scale description $x^* \sim p_{cf}(x^*|X^*, \theta^*_{cf})$
Bayesian coarse-graining: ensemble averages

Approximating ensemble averages by

\[
\langle f(x) \rangle_{p_f(x)} = \int_{\mathcal{M}} f(x) p_f(x) dx
\]

\[
= E[f(x)] \approx E[f(x) | x^{(1:N)}]
\]

\[
= \int_{\mathcal{M}} f(x) p(x | x^{(1:N)}) dx
\]

\[
= \int_{\mathcal{M}} f(x) p_{cf}(x | \mathbf{X}, \theta_{cf}) p_c(\mathbf{X} | \theta_c) p(\theta_c, \theta_{cf} | x^{(1:N)}) dx d\mathbf{X} d\theta_c d\theta_{cf}
\]

Error arising from discrepancy between \(E[f(x)]\) and \(E[f(x) | x^{(1:N)}]\)
Bayesian coarse-graining: coarse-scale model $p_c$

Coarse-scale model

$$p_c(X|\theta_{cf}) = \frac{\exp\{-\beta U_c(X, \theta_c)\}}{Z_c(\theta_c)}$$

Requirements

- Allow sufficient flexibility of $p_c(X)$
- $U_c$ built from high-order interactions affording flexibility in $p_c(X)$
Bayesian coarse-graining: reconstruction map $p_{cf}$

Map from coarse to fine: $p_{cf}(x|X, \theta_{cf})$

- Several fine-configurations map to same coarse-configuration
  - Probabilistic relation between a given coarse-configuration and a fine-configuration takes account of it. [7]
- Fast reconstruction ([7, 1]) of fine-scale configurations $x$ given a coarse-configuration $X$ desired.

Influence of $p_{cf}$

- What is expected from the coarse variables $X$ in terms of predicting the given data $x^{(1:N)}$?
- Adjusting $p_c(X|\theta_c)$ so that $X$ agrees best with data $x^{(1:N)}$ connected with probabilistic mapping $p_{cf}$.

Figure: Configurations $x$ given $X$
Optimization

For learning the model parameters $\theta_c$ and $\theta_{cf}$ the expectation maximization algorithm is applied.

- Point estimates for maximizing the likelihood or posterior distribution are obtained.
- Approximation of posterior on $\theta_{cf}$ by Laplace (validated by full posterior sampling)

MAP estimate

$$\theta_{MAP} = \arg \max_{\theta} \left[ \log p(\theta_c, \theta_{cf}, x^{(1:N)}) \right]$$

$$= \arg \max_{\theta} \left[ \log p(x^{(1:N)}|\theta_{cf}, \theta_{cf}) + \log p(\theta_c) + \log p(\theta_{cf}) \right]$$

- In the following the ML-estimate is shown and will be extended by prior distributions $p(\theta)$ in a next step by adding log-priors $\log p(\theta_c) + \log p(\theta_{cf})$. 
Optimization

To obtain point estimates with MLE for the model parameters $\theta_\text{c}$ and $\theta_\text{cf}$ we maximize the log-likelihood by using the EM algorithm.

The likelihood is defined as

$$p(x^{(1:N)}|\theta_\text{c}, \theta_\text{cf}) = \int p(x^{(1:N)}, x^{(1:N)}|\theta_\text{c}, \theta_\text{cf}) dx^{(1:N)}$$

$$= \int p_\text{cf}(x^{(1:N)}|x^{(1:N)}, \theta_\text{cf}) p_\text{c}(x^{(1:N)}|\theta_\text{c}) dx^{(1:N)}$$

$$= \int \prod_{i=1}^{N} p_\text{cf}(x^{(i)}|x^{(i)}, \theta_\text{cf}) p_\text{c}(x^{(i)}|\theta_\text{c}) dx^{(1:N)}$$

augmenting the log-likelihood with an arbitrary density $q(x^{(1:N)})$

$$\log p(x^{(1:N)}|\theta_\text{c}, \theta_\text{cf}) = \log \int q(x^{(1:N)}) \frac{p_\text{cf}(x^{(1:N)}|x^{(1:N)}, \theta_\text{cf}) p_\text{c}(x^{(1:N)}|\theta_\text{c})}{q(x^{(1:N)})} dx^{(1:N)}$$

$$\geq \int q(x^{(1:N)}) \log \frac{p_\text{cf}(x^{(1:N)}|x^{(1:N)}, \theta_\text{cf}) p_\text{c}(x^{(1:N)}|\theta_\text{c})}{q(x^{(1:N)})} dx^{(1:N)}$$

$$= \mathcal{L}(q(x^{(1:N)}), \theta_\text{c}, \theta_\text{cf})$$
Expectation maximization

It holds the following decomposition for the lower bound:

\[
\mathcal{L}(q(X^{(1:N)}), \theta_c, \theta_{cf}) = \int q(X^{(1:N)}) \log \frac{p(x^{(1:N)}|X^{(1:N)}, \theta_{cf})p(X^{(1:N)}|\theta_c)}{q(X^{(1:N)})} dX \\
= -\text{KL}(q(X^{(1:N)})||p(X^{(1:N)}|x^{(1:N)}, \theta_{cf}, \theta_{cf})) + \log p(x^{(1:N)})
\]

**Decomposition of the log likelihood**

\[
\log p(x^{(1:N)}|\theta_c, \theta_{cf}) = \mathcal{L}(q(X^{(1:N)}), \theta_c, \theta_{cf}) + \text{KL}(q(X^{(1:N)})||p(X^{(1:N)}|x^{(1:N)}, \theta_{cf}, \theta_{cf})).
\]

Maximizing \( \mathcal{L}(q(X^{(1:N)}), \theta_c, \theta_{cf}) \) is equal to minimizing \( \text{KL}(q(X^{(1:N)})||p(X^{(1:N)}|x^{(1:N)}, \theta_{cf}, \theta_{cf})) \)

\( p(X^{(1:N)}|x^{(1:N)}, \theta_{cf}, \theta_{cf}) \) PDF over pre-images for \( X \) for the given data \( x^{(1:N)} \) and parameters \( \theta \).
For given parameters $\theta_{\text{cf}}$, $\theta_{\text{c}}$ the lower bound is maximized if we choose

$$q(X^{(1:N)}) \propto p_{\text{cf}}(x^{(1:N)}|X^{(1:N)}, \theta_{\text{cf}})p_c(X^{(1:N)}|\theta_{\text{c}}).$$

Therefore,

$$\Rightarrow KL(q(X^{(1:N)})||p(X^{(1:N)}|x^{(1:N)}, \theta_{\text{c}}, \theta_{\text{cf}})) = 0.$$
First order derivatives

\[
\frac{\partial L}{\partial \theta_c} = \beta \left( N \left\langle \frac{\partial U_c(\mathbf{X}, \theta_c)}{\partial \theta_c} \right\rangle_{p_c(\mathbf{X}^{(i)}|\theta^t_c)} - \sum_{i=1}^{N} \left\langle \frac{\partial U_c(\mathbf{X}^{(i)}, \theta_c)}{\partial \theta_c} \right\rangle_{q(\mathbf{X}^{(i)}|\mathbf{X}^{(i)}, \theta^t_c, \theta^t_c)} \right)
\]

\[
\frac{\partial L}{\partial \theta_{cf}} = \left\langle \frac{1}{p_{cf}(\mathbf{X}^{(1:N)}|\mathbf{X}, \theta_{cf})} \frac{\partial p_{cf}(\mathbf{X}^{(1:N)}|\mathbf{X}, \theta_{cf})}{\partial \theta_{cf}} \right\rangle_{q(\mathbf{X}^{(1:N)}|\mathbf{X}^{(1:N)}, \theta^t_{cf}, \theta^t_c)}
\]

Second order derivatives

\[
\frac{\partial^2 L}{\partial \theta_k \partial \theta_l} = -\beta \sum_{i=1}^{N} \left\langle \frac{\partial^2 U_c(\mathbf{X}^{(i)}, \theta_c)}{\partial \theta_k \partial \theta_l} \right\rangle_{q(\mathbf{X}^{(i)}|\mathbf{X}^{(i)}, \theta^t_{cf}, \theta^t_c)} + \beta N \left\langle \frac{\partial^2 U_c(\mathbf{X}, \theta_c)}{\partial \theta_k \partial \theta_l} \right\rangle_{p_c(\mathbf{X}|\theta_c)}
\]

\[
- \beta^2 N \left\langle \frac{\partial U_c(\mathbf{X}, \theta_c)}{\partial \theta_k} \frac{\partial U_c(\mathbf{X}, \theta_c)}{\partial \theta_l} \right\rangle_{p_c(\mathbf{X}|\theta_c)} + \beta^2 N \left\langle \frac{\partial U_c(\mathbf{X}, \theta_c)}{\partial \theta_k} \right\rangle_{p_c(\mathbf{X}|\theta_c)} \left\langle \frac{\partial U_c(\mathbf{X}, \theta_c)}{\partial \theta_l} \right\rangle_{p_c(\mathbf{X}|\theta_c)}
\]
Second order derivatives

\[
\frac{\partial^2 \mathcal{L}(q(X^{(1:N)}), \theta_c, \theta_{cf})}{\partial \theta_{cf}^2} = \sum_{i=1}^{N} \left( \frac{1}{p_{cf}(x^{(i)}|X^{(i)}, \theta_{cf})} \cdot \frac{\partial^2 p_{cf}(x^{(i)}|X^{(i)}, \theta_{cf})}{\partial \theta_{cf}^2} \right) - \frac{1}{p_{cf}(x^{(i)}|X^{(i)}, \theta_{cf})^2} \left( \frac{\partial p_{cf}(x^{(i)}|X^{(i)}, \theta_{cf})}{\partial \theta_{cf}} \right)^2
\]
Robbins-Monro Optimization

Gradients are sample averages → noise afflicted.

Robbins-Monro Algorithm

\[ \theta^{t+1} = \theta^t + \alpha_t \nabla_{\theta} \mathcal{L}(\theta^t) \]

with

\[ \alpha_t = \frac{\alpha}{(t + A)^\rho}, \text{ with } \frac{1}{2} < \rho \leq 1. \]

- Convergence is guaranteed to the true optimum \( \theta^* \) if infinite steps performed.
- \( \alpha_t \) is a sequence of real numbers and has to fulfill:

\[ \sum_{t=1}^{\infty} \alpha_t = \infty \quad \text{and} \quad \sum_{t=1}^{\infty} \alpha_t^2 < \infty \]
Laplace Approximation

Approximating a density function \( p(z) = \frac{1}{Z} f(z) \) by

\[
q(z) = \mathcal{N}(z|z_0, \Sigma)
\]

with \( \Sigma = (-\nabla\nabla f(z)|_{z=z_0})^{-1} \) and \( \nabla f(z)|_{z=z_0} = 0 \).

For the given problem using a noninformative prior it follows

**Laplace approximation**

\[
\frac{\partial^2 \log p(\theta_{cf}|x^{(1:N)})}{\partial \theta^2_{cf}} = \sum_i^N \left\langle \frac{\partial^2 \log p_{cf}(x^{(i)}|X^{(i)}, \theta_{cf})}{\partial \theta^2_{cf}} \right\rangle_{q(X^{(i)})}
\]
Predictive uncertainty

Propagate uncertainty induced by mapping

1. $\theta_{cf}^i \sim \mathcal{N}(\theta_{cf}^{MAP}, \Sigma)$ with $\Sigma = \left(-\frac{\partial^2 \log p(\theta_{cf}|x^{(1:N)})}{\partial \theta_{cf}^2}\right)^{-1}$

2. $X^{ij} \sim p_c(X|\theta_{c}^{MAP})$

3. $x^{ij}_k \sim p_{cf}(x|X^{ij}, \theta_{cf}^i)$

Predictive uncertainty of properties:

$\langle f|_{\theta_c, \theta_{cf}'} = \rangle = \int f(x)p_{cf}(x|X, \theta_{cf}^i)p_c(X|\theta_{c}^{MAP})dx dX$
Example Problem: Coarse-graining 1D Ising Model

- Applying proposed method ('predCg') for coarse-graining a one dimensional Ising Model
- Comparison between 'predCg' and the deterministic formulation of minimizing the KL-divergence ('relEntr') by Shell (2008)
- Assessed by predictive capabilities for magnetization and correlation
The Ising-model regarded in the following discussion is one dimensional. It has the properties:

- System size $n_f$ in fine- and $n_c$ in coarse-scale
- Level of coarse-graining $l_c = \frac{n_f}{n_c}$
- Inverse temperature $\beta$
- External field $\mu$
- Regarded interaction length $L_f$ in fine- and $L_c$ in coarse scale, including all interactions $k \leq L$
- $J_0 = 1$ overall interaction strength
Ising Model - Potential in Fine-Scale

Fine variables take the values $x_i \in \{-1, 1\}$ following

$$p_f(x|\beta) \propto \exp(-\beta U(x, J_k, \mu))$$

**Fine-scale potential**

$$U(x, J_k, \mu) = -\frac{1}{2} \sum_{k=1}^{L_f} J_k \sum_{|i-j|=k} x_i x_j - \mu \sum_{i=1}^{n_f} x_i$$

with $i, j \in \{1, \ldots, n_f\}$ having $n_f$ lattice sites.

- Maximal interactions of $L_f$ sites apart are regarded in the potential.
- $|i-j|=k$ interpreted as summation over neighbors $k$ over all sites $i$ sites apart
- $J_k$, strength of the $k$-th interaction.

with $J_k$ following a power law for a given overall strength $J_0$ and exponent $a$,

$$J_k = \frac{K}{Lk^a}$$

with,

$$K = J_0 L^{1-a} \sum_{k=1}^{L} k^{-a}$$

in order to normalize the interaction strength [1].
Coarse variables take the values $X_i \in \{-1, 1\}$, $p_c \propto \exp\{-\beta U_c(X, \theta_c, \mu)\}$

### Coarse-scale potential

$$U_c(X, \theta_c, \mu) = -\frac{1}{2} (\theta_{\text{lin}}^c \sum_{i=1}^{n_c} X_i$$

$$+ \sum_{k=1}^{L_c} \theta_{\text{two}}^c \sum_{|i-j|=k} X_i X_j$$

$$+ \sum_{i=1}^{n_c} \theta_{\text{trip}}^c \sum_{m=1}^{n} \sum_{n=1}^{n} X_i X_{i \pm m} X_{i \pm m \pm n})$$

$$- \mu \sum_{i=1}^{n_c} X_i$$

with $i, j \in \{1, \ldots, n_c\}$ and $n_c \ll n_f$ lattice sites.
Each coarse variable $X_r, \ r = \{1, \ldots, R\}$ describes $S$ fine-scale variables $x_{r,s}$ with $s = \{1, \ldots, S\}$. The mapping from fine-scale variables $x_{r,s}$ to coarse variable $X_r$ is defined as,

$$X_r = \begin{cases} 
+1, & \frac{1}{S} \sum_{s} x_{r,s} > 0 \\
-1, & \frac{1}{S} \sum_{s} x_{r,s} < 0 \\
U(-1, +1), & \text{otherwise}
\end{cases}$$

**Figure**: Mapping from fine-scale $x$ to coarse-scale $X$
Coarse-to-fine mapping \( p_{cf}(x|X, \theta_{cf}) \)

\[
p_{cf}(x_r,s|X_r) = \theta_{cf} \left( \frac{1+x_{r,s}X_r}{2} \right) \left( 1 - \theta_{cf} \right) \left( \frac{1-x_{r,s}X_r}{2} \right)
\]

Assumption: \( x_i \) conditionally independent for given \( X \):

\[
p(x^{1:N}|X^{1:N}, \theta_{cf}) = N \prod_{i=1}^{N} p_{cf}(x^{(t)}|X^{(t)}, \theta_{cf})
\]

\[
= \prod_{i=1}^{N} \prod_{r=1}^{R} \prod_{s=1}^{S} \left( \theta_{cf} \frac{1+x_{r,s}X_r}{2} \left( 1 - \theta_{cf} \right) \frac{1-x_{r,s}X_r}{2} \right)
\]

\[
= \theta_{cf}^{\sum_{i=1}^{N} \sum_{r=1}^{R} \sum_{s=1}^{S} \frac{1+x_{r,s}X_r}{2}}
\]

\[
\left( 1 - \theta_{cf} \right)^{\sum_{i=1}^{N} \sum_{r=1}^{R} \sum_{s=1}^{S} \frac{1-x_{r,s}X_r}{2}}
\]

Figure: Probabilistic mapping form coarse to fine: \( p_{cf}(x|X, \theta_{cf}) \)
Which properties to predict?

The magnetization $m$ is given:

**Magnetization in the fine-scale**

$$m_{\text{fine}} = \int \frac{1}{n} \sum_{i} x_i \ p_f(x) \, dx$$

**Magnetization calculated with proposed method ('predcg')**

$$m_{\text{ML,fine}} = \int \frac{1}{n} \sum_{i} x_i \ p_{\text{cf}}(x|X, \theta_{\text{cf}}) p_c(X|\theta_c) p(\theta_{\text{cf}}, \theta_c|x^{(1:N)}) \ d\theta_{\text{cf}} d\theta_c dX dx$$
Quantity of Interest: Magnetization

Magnetization calculated with relative entropy method ‘relEntr’

(deterministic mapping fine to coarse: $X = \xi(x)$)

- For each $\mu$ calculate optimal model parameter $\theta_{cg, \det}^*$
- Since we want to compare the magnetization on the same (fine) scale a mapping is introduced
  - if $X_r = -1$, select randomly a possible configuration $x_{r,s}$ where either all $x_{r,s} = -1$ or one is $x_{r,s} = -1$ and the other $x_{r,s} = 1$ (for given $r$: $s \in \{1 \ldots l_c\}$)
  - if $X_r = +1$, select randomly a possible configuration $x_{r,s}$ where either all $x_{r,s} = +1$ or one is $x_{r,s} = +1$ and the other $x_{r,s} = -1$ (for given $r$: $s \in 1, 2$)
- For a given coarse-state the selected fine-states $x_{\det}$ are used for calculating the magnetization

Relative error in magnetization,

$$\text{err}_{mag} = \frac{\|m - m_{\text{truth}}\|}{\|m_{\text{truth}}\|}.$$
The correlation $R_l$ of all sites being separated by $l$ sites is measured as,

$$R_l = \frac{1}{n} \sum_{|i-j|=l}^{n} < x_i x_j > .$$

Relative correlation error,

$$\text{err}_\text{corr} = \frac{\| R - R_{\text{truth}} \|}{\| R_{\text{truth}} \|} .$$
1: for $\mu = \mu_{\text{min}}$ to $\mu_{\text{max}}$ do
2:   set initial fine-scale parameter: $J_0$, $n_f$, $\mu$
3:   create data set: $x^{(1:S_f)} \sim p(x^{(1:S_f)} | J_0, \mu, \beta)$ with $S_f$ samples
4:   set initial coarse-graining parameter: $\theta_{c0}$, $\theta_{cf0}$, $n_c = \frac{n_f}{l_c}$, $\mu$
5:   $\theta^t_c \leftarrow \theta_{c0}$
6:   $\theta^t_{cf} \leftarrow \theta_{cf0}$
7: while !(Convergence check passed) do
8:   function E-STEP
9:      $X \sim q(X|x^{(1:N)}, \theta^t_c, \theta^t_{cf}, \mu) \propto p_{cf}(x^{(1:N)}|X, \theta^t_{cf})p_c(X|\theta^t_c, \mu)$ with $S_q$ per data point $x^{(i)}$
10: function M-STEP
11:      $\frac{\partial L}{\partial \theta^t_c} = -\beta \left< \frac{\partial}{\partial \theta^t_c} U_c(X, \theta_t^c) \right> q(x^{(1:N)}|x, \theta^t_c, \theta^t_{cf}) + \beta \left< \frac{\partial}{\partial \theta^t_c} U_c(X, \theta_t^c) \right> p_c(X|\theta^t_c)$
12:      $\frac{\partial L}{\partial \theta^t_{cf}} = \left< \frac{1}{p_{cf}(x^{(1:N)}|X, \theta_{cf})} \frac{\partial p_{cf}(x^{(1:N)}|X, \theta_{cf})}{\partial \theta_{cf}} \right> q(x^{(1:N)}|x, \theta^t_c, \theta^t_{cf})$
13:      $\theta^{t+1}_c \leftarrow \theta^t_c + \alpha \theta^t_c \frac{\partial L}{\partial \theta^t_c}$
14:      $\theta^{t+1}_{cf} \leftarrow \theta^t_{cf} + \alpha \theta^t_{cf} \frac{\partial L}{\partial \theta^t_{cf}}$
15: end while
16: $x^{\text{pred}} \sim \tilde{p}(x_{\text{new}}|\theta^t_c, \theta^t_{cf})$ with $S^*_{\text{pred}} \cdot S_c$ samples
17: function EVALMAG($x^{\text{pred}}, x^{(1:N)}$)
18: function EVALCORR($x^{\text{pred}}, x^{(1:N)}$)
19: end for=0
Ising Model - Posterior and Prediction

Coarse-Scale Sample
\( X \sim p(X^i | x^i, \theta_c, \theta_{cf}) \)

Predicted Fine-Scale Sample
\( x_{pred}^i \sim p_{cf}(x^i | X^i, \theta_{cf}) \)

Fine-Scale Data Point \( x^i \)
Overview

1. Behavior of parameters with respect to $\mu$
2. Influence of available data $S_f$ for training the model
3. Level of coarse-graining
4. Aspects of coarse-grained models
Behavior of parameters with respect to $\mu$

**Constant attributes:**

- $S_c = 200$ (samples $\mathbf{X} \sim p_c(\mathbf{X}|\theta_c)$)
- $S_q = 50$ (samples $\mathbf{X} \sim q(\mathbf{X}^{(i)}|\mathbf{x}^{(i)}, \theta_c, \theta_{cf})$)
- $J_0 = 1$ (overall interaction strength)
- $S_{pred} = 100S_c$ (samples for prediction step after learning the model)

Furthermore:

- $n_f = 32$, system size fine-scale
- $L_f = L_c = 1$, in fine- and coarse scale nearest-neighbor interactions are regarded
Behavior of parametrization with respect to $\mu$

The potentials follows to,

$$U(J, \mu) = -\frac{1}{2} J \sum_{|i-j|=1} x_i x_j - \mu \sum_{i=1}^{n_f} x_i$$

$$U_c(\theta_c, \mu) = -\frac{1}{2} \theta_c \sum_{|i-j|=1} X_i X_j - \mu \sum_{i=1}^{n_c} X_i.$$

- Learn two parameters $\theta_c$ and $\theta_{cf}$ of the mapping $p_{cf}$
- Learning for every evaluated $\mu$: $\mu = [-4.5, 4.5]$ with a step size of 0.6
Figure: $\theta_c$: Parameter of potential $U_c$

Figure: $\theta_{cf}$: Parameter of the mapping $p_{cf}(x | X, \theta_{cf})$
Key aspects

- True magnetization based on data set and predicted by 'predCg' coincides.
- Method 'relEntr' not able to predict data set due to missing mapping capabilities from coarse to fine.
- Parameters $\theta_c$ and $\theta_{cf}$ behave symmetrical with respect to $\mu = 0$.
- Stronger interaction $\theta_c$ for bigger $|\mu|$ but also higher probability $\theta_{cf}$ that the coarse configuration reflects the fine configuration.
Behavior with respect to size of data-set

- Learn $\theta_c$ and $\theta_{cf}$ at each evaluated $\mu$
- Given different size of data-set $S_f$ with $S_f \in \{5, 10, 20, 50\}$.
- Interaction length in fine-scale $L_f = 10$ with exponential decay of overall strength $J_0 = 1.5$

Potentials

$U_f(J, \mu) = -\frac{1}{2} \sum_{k}^{L_f} J_k \sum_{|i-j|=k} x_i x_j - \mu \sum_{i=1}^{n_f} x_i$

$U_c(\theta_c, \mu) = -\frac{1}{2} \theta_c \sum_{|i-j|=1} x_i x_j - \frac{1}{2} \theta_{c}^{lin} \sum_{i}^{n_c} x_i - \mu \sum_{i=1}^{n_c} x_i$. 
Behavior with respect to size of data-set

Figure: Uncertainty in magnetization $S_f = 5$

Figure: Uncertainty in magnetization $S_f = 10$
Behavior with respect to size of data-set

Figure: Uncertainty in magnetization $S_f = 20$

Figure: Uncertainty in magnetization $S_f = 50$
Behavior with respect to size of data-set

**Figure**: Relative mean error of magnetization with respect to increasing amount of data $S_f$, $S_f \in \{5, 10, 20, 50, 100\}$

**Figure**: Relative mean error of correlation with respect to increasing amount of data $S_f$, $S_f \in \{5, 10, 20, 50, 100\}$
Behavior with respect to size of data-set

Uncertainty decreasing with increasing amount of data available

By using $S_f = 5$ predictions resembling true magnetization well

Figure: Correlation $R_l$ for $\mu = 0$ with $S_f \in \{5, 10, 20, 50, 100\}$
Predictability by various levels of coarse-gaining

- Choose various level of coarse-graining $l_c$
- Mapping $p_{cf}(x|X, \theta_{cf})$ responsible for $l_c \in \{2, 4, 8, 16\}$ fine variables
- Coarse-grained potential $U_c(X, \theta_c)$ as given before
Predictability by various levels of coarse gaining

Figure: Uncertainty in magnetization $l_c = 16$

Figure: Uncertainty in magnetization $l_c = 8$
Predictability by various levels of coarse gaining

Figure: Uncertainty in magnetization $l_c = 4$

Figure: Uncertainty in magnetization $l_c = 2$
Predictability by various levels of coarse gaining

Figure: Correlation $R_l$ at $\mu = 0$ with $l_c \in \{2, 4, 8, 16\}$

- Applying $l_c = 8$ leads already to good predictions in magnetization.
- Predicting correlations needs finer resolution in coarse-scale.
Model Comparison

- Assumption before: Model is given for potential $U_c(X, \theta_c)$
- Using different order of interactions and interaction lengths
Model Comparison

Model [abc]

- (a) $a = 0$ no linear term, $a = 1$ linear term
- (b) $b = 0$ no 2nd order term, $b = x$ 2nd order term up to interaction length $L_c = x$
- (c) $c = 0$ no third order term, $c = 1$ third order term on (nearest correlation)
Model Comparison

- Linear term necessary to describe data
- Second order interaction necessary
- In correlation: errors decreasing for including 2nd order interactions with higher interaction lengths $L_c$
- No further improvement for longer ranged third order interactions
Outline

Numerical issues / efficiency
  ▶ Use advanced sampling methods / optimization methods, using curvature information (in progress)
  ▶ Variational approximations

Coarse-Graining
  ▶ Hierarchical coarse-graining
  ▶ How to select mappings from coarse to fine?
  ▶ Probit- or Logit-classification model, O model for coarse variables

Algorithmic
  ▶ Approach with sparsity priors for model selection (in progress).

I. Bilionis and N. Zabaras.
A stochastic optimization approach to coarse-graining using a relative-entropy framework.  

C. M. Bishop.
*Pattern Recognition and Machine Learning*.  

A. Chaimovich and M. S. Shell.
Coarse-graining errors and numerical optimization using a relative entropy framework.  

F. Ercolessi and J. B. Adams.
Interatomic potentials from first-principles calculations: The force-matching method.  

Multiscale coarse graining of liquid-state systems.  

Numerical and statistical methods for the coarse-graining of many-particle stochastic systems.  

A. P. Lyubartsev and A. Laaksonen.
Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach.  
D. Ming and M. E. Wall.
Allostery in a coarse-grained model of protein dynamics.

M. S. Shell.
The relative entropy is fundamental to multiscale and inverse thermodynamic problems.