Information-theoretic tools for parametrized coarse-graining of non-equilibrium extended systems

Petr Plecháč

Dept. Mathematical Sciences, University of Delaware
plechac@math.udel.edu
http://www.math.udel.edu/~plechac

ACMAC, Crete, Sep 16–20, 2013
Collaborators:

- M.A. Katsoulakis (University of Massachusetts, Amherst)
- Y. Pantazis (University of Massachusetts, Amherst)

Funding:

- NSF-CMMI, DOE

References:

http://www.math.udel.edu/~plechac/CDIsite-php

Outline

- Coarse-graining & Parameterization – Equilibrium
- Non-equilibrium processes & Path Space Relative Entropy
- Coarse-graining & Parameterization–Non-Equilibrium
- Benchmarks
- Relative Entropy Rate & Statistical Estimation
Relative Entropy and $\mathcal{R}$-projections

- **Pseudo-distance** (Leibler-Kullback divergence)

\[
\mathcal{R}(P|Q) = \int \log \left( \frac{dP}{dQ} \right) dP
\]

for $P \ll R$, $Q \ll R$ \hspace{1cm} $\mathcal{R}(P|Q) = \int p_R \log \left( \frac{p_R}{q_R} \right) dR$

- **Properties:**
  - (i) $\mathcal{R}(P|Q) \geq 0$ and
  - (ii) $\mathcal{R}(P|Q) = 0$ iff $P = Q$ a.e.

- **$\mathcal{R}$-geometry of probability distributions**

\[
\mathcal{B}(R, \rho) = \{ P \mid \mathcal{R}(P|R) < \rho \}
\]

$\mathcal{R}$-projection on $\mathcal{A}$ convex, TV closed, $\mathcal{A} \cap \mathcal{B}(R, \rho) \neq \emptyset$ (Kullback, Csiszár)

\[
\mathcal{R}(Q|R) = \min_{P \in \mathcal{A}} \mathcal{R}(P|R)
\]
“Geometry”: tangent hyperplane to $\mathcal{B}(R, \rho)$ at $Q$, $\rho = \mathcal{R}(Q|R)$

$$P : \quad \int \log \frac{dQ}{dR} \, dP = \rho$$

and

$$\mathcal{R}(P|R) = \mathcal{R}(P|Q) + \mathcal{R}(Q|R)$$
Bounding the approximation error

modeling error
Bounding the approximation error

modeling error + numerical error
Bounding the approximation error

modeling error + numerical error + statistical error

Csiszar-Kullback-Pinsker inequality:

$$|\mathbb{E}_P[\phi] - \mathbb{E}_Q[\phi]| \leq \|\phi\|_\infty \sqrt{2\mathcal{R}(P|Q)}$$

CG: Error Quantification and Parameterization using RE:
Bounding the approximation error

modeling error

In parametrized setting: \( Q \equiv P^{\theta^*} \) and \( P \equiv P^{\theta^*+\epsilon} \)

- Csiszar-Kullback-Pinsker inequality:

\[
|E_{P^{\theta}}[\phi] - E_{P^{\theta+\epsilon}}[\phi]| \leq \|\phi\|_\infty \sqrt{2\mathcal{R}(P^{\theta} \mid P^{\theta+\epsilon})}
\]
Bounding the approximation error

modeling error

In parametrized setting: $Q \equiv P^{\theta^*}$ and $P \equiv P^{\theta^*+\epsilon}$

- Csiszar-Kullback-Pinsker inequality:

$$|\mathbb{E}_{P^{\theta}}[\phi] - \mathbb{E}_{P^{\theta+\epsilon}}[\phi]| \leq \|\phi\|_\infty \sqrt{2\mathcal{R}(P^\theta | P^{\theta+\epsilon})}$$

- Variational bounds: Donsker-Varadhan variational characterization

$$\mathcal{R}(P | Q) = \sup_{\phi} \{\mathbb{E}_P[\phi] - \log \mathbb{E}_Q[e^{\phi}]\}$$

Legendre transform

$$\frac{1}{c} \log \mathbb{E}_P[e^{c\phi}] = \sup_Q \left\{-\frac{1}{c} \mathcal{R}(P | Q) + \mathbb{E}_Q[\phi]\right\}$$
Bounding the approximation error

modeling error

In parametrized setting: \( Q \equiv P^{\theta^*} \) and \( P \equiv P^{\theta^*+\epsilon} \)

- Csiszar-Kullback-Pinsker inequality:
  \[
  |E_{P^\theta} [\phi] - E_{P^{\theta+\epsilon}} [\phi]| \leq ||\phi||_\infty \sqrt{2R(P^\theta | P^{\theta+\epsilon})}
  \]

- Variational bounds: Donsker-Varadhan variational characterization
  \[
  R(P | Q) = \sup_{\phi} \{E_P[\phi] - \log E_Q[e^\phi]\}
  \]

Legendre transform

\[
\frac{1}{c} \log E_P[e^{c\phi}] = \sup_Q \left\{ -\frac{1}{c} R(P | Q) + E_Q[\phi] \right\}
\]

Variational Bounds [Chowdhary & Dupuis 2009]

\[
\sup_{c>0} \left\{ -\frac{1}{c} \log E_P[e^{-c\phi}] - \frac{1}{c} R(P | Q) \right\} \leq E_Q[\phi]
\]

\[
E_Q[\phi] \leq \inf_{c>0} \left\{ \frac{1}{c} \log E_P[e^{c\phi}] + \frac{1}{c} R(P | Q) \right\}
\]
Variational bounds

Dupuis, Katsoulakis, Pantazis, PP 2013

\[ Q \equiv P^{\theta^*} \text{ and } P \equiv P^{\theta^* + \epsilon} \]

\[ |\mathbb{E}_{P^{\theta}}[\phi] - \mathbb{E}_{P^{\theta + \epsilon}}[\phi]| \leq \sqrt{\text{Var}_{P^{\theta}}[\phi]} \sqrt{\mathbb{F}(P^{\theta})} \epsilon + \mathcal{O}(\epsilon^2) \]

Fisher Information \( \mathbb{F}(P^{\theta}) = \int \nabla_\theta p^{\theta} \nabla_\theta p^{\theta} \, dP^{\theta} \)

\[
|\mathbb{E}_P[\phi] - \mathbb{E}_Q[\phi]| = \left| \int \phi (1 - \frac{dP}{dQ}) \, dQ \right| \\
= \left| \int \phi (1 - \frac{dP}{dQ}) \, dQ - \mathbb{E}_Q[\phi] \int (1 - \frac{dP}{dQ}) \, dQ \right| \\
= \left| \int (\phi - \mathbb{E}_Q[\phi]) (1 - \frac{dP}{dQ}) \, dQ \right| \\
\leq \left( \int (\phi - \mathbb{E}_Q[\phi])^2 \, dQ \right)^{1/2} \left( \int (1 - \frac{dP}{dQ})^2 \, dQ \right)^{1/2}
\]

Petr Plecháč (UDEL) Approximating
Coarse-Graining – Equilibrium

1. Coarse-graining of polymers; DPD methods

Briels, et. al. J.Chem.Phys. '01;
Doi et. al. J.Chem.Phys. '02;
Kremer et. al. Macromolecules '06;
Müller-Plathe Chem.Phys.Chem '02;
Laaksonen et. al. Soft Matter '03, etc;
Deserno et. al. Nature '07;
Espanol J Chem. Phys. '07, '11

Harmandaris Macromolecules

2. Stochastic lattice dynamics/KMC

Katsoulakis, Majda, Vlachos, PNAS'03;
Katsoulakis, P.P., Sopasakis, SIAM Num. Anal. '06;
Are, Katsoulakis, P.P., Rey-Bellet SIAM J.Sci.Comp. '08;
Sinno et al. J.Chem.Phys.'08, '13, PRE '12

Microscopic lattice

Coarse lattice
Coarse-graining in lattice systems

**Example:** Block spins

Coarse map: \( T : \Sigma_N \rightarrow \tilde{\Sigma}_M \)

\[
\{\eta(k)\} = T\sigma = \sum_{x \in C_k} \sigma(x)
\]

Microscopic process: \( \{\sigma_t\}_{t \geq 0}, \mathcal{L} \), \( c(x, \sigma) \)

Coarse process: \( \{\eta_t\}_{t \geq 0}, \tilde{\mathcal{L}} \), \( \tilde{c}(k, \eta) \)

Reconstructed process: \( \{\tilde{\sigma}_t\}_{t \geq 0}, \tilde{\mathcal{L}} \), \( \tilde{c}(x, \sigma) \)

Errors: at finite \( t \) and as \( t \to \infty \)

modeling error
Coarse-graining in lattice systems

**Example:** Block spins

Coarse map: \( T : \Sigma_N \to \tilde{\Sigma}_M \)

\[ \{ \eta(k) \} = T \sigma = \sum_{x \in C_k} \sigma(x) \]

Microscopic process: \( \{ \sigma_t \}_{t \geq 0}, \mathcal{L} \), \( c(x, \sigma) \)

Coarse process: \( \{ \eta_t \}_{t \geq 0}, \tilde{\mathcal{L}} \), \( \tilde{c}(k, \eta) \)

Reconstructed process: \( \{ \tilde{\sigma}_t \}_{t \geq 0}, \tilde{\mathcal{L}} \), \( \tilde{c}(x, \sigma) \)

Errors: at finite \( t \) and as \( t \to \infty \)

modeling error + numerical error
Coarse-graining in lattice systems

**Example:** Block spins

Coarse map: \( T : \Sigma_N \rightarrow \tilde{\Sigma}_M \)

\[ \{ \eta(k) \} = T\sigma = \sum_{x \in C_k} \sigma(x) \]

Microscopic process: \( \{ \sigma_t \}_{t \geq 0}, \mathcal{L} \), \( c(x, \sigma) \)

Coarse process: \( \{ \tilde{\eta}_t \}_{t \geq 0}, \tilde{\mathcal{L}} \), \( \tilde{c}(k, \eta) \)

Reconstructed process: \( \{ \tilde{\sigma}_t \}_{t \geq 0}, \tilde{\mathcal{L}} \), \( \tilde{c}(x, \sigma) \)

**Errors:** at finite \( t \) and as \( t \rightarrow \infty \)

- modeling error
- numerical error
- statistical error
Coarse-graining in lattice systems

**Example:** Block spins

Coarse map: \( \mathbf{T} : \Sigma_N \rightarrow \bar{\Sigma}_M \)

\[ \{ \eta(k) \} = \mathbf{T} \sigma = \sum_{x \in C_k} \sigma(x) \]

Microscopic process: \( \{ \sigma_t \}_{t \geq 0}, \mathcal{L} \), \( c(x, \sigma) \)

Coarse process: \( \{ \eta_t \}_{t \geq 0}, \bar{\mathcal{L}} \), \( \bar{c}(k, \eta) \)

Reconstructed process: \( \{ \bar{\sigma}_t \}_{t \geq 0}, \bar{\mathcal{L}} \), \( \bar{c}(x, \sigma) \)

Errors: at finite \( t \) and as \( t \rightarrow \infty \)

modeling error + numerical error + statistical error

Renormalization Group map:

\[ \bar{H}(\eta) = -\frac{1}{\beta} \log \int \exp\{-\beta H_N(\sigma)\} P(d\sigma|\eta) \approx \bar{H}^0(\eta) \text{ "computable"} \]
**Equilibrium**

Invariant measure: \( \mu \sim e^{-\beta H(\sigma)} \)

Detailed balance for the coarse-grained process w.r.t. \( \bar{\mu} \sim e^{-\beta \bar{H}(\eta)} \)

Coarse-grained Hamiltonian \( \bar{H}(\eta) \)

\[
e^{-\beta \bar{H}(\eta)} = \mathbb{E}[e^{-\beta H_N} | \eta] \equiv \int_\chi e^{-\beta H_N(\sigma)} P_N(d\sigma | \eta)
\]

Approximation of RG map

Approximate \( \bar{H}(\eta) \approx \bar{H}^{(0)}(\eta) \), i.e., \( \bar{\mu}(d\eta) \approx \bar{\mu}^{(0)}(d\eta) \)
Equilibrium

Invariant measure: $\mu \sim e^{-\beta H(\sigma)}$

Detailed balance for the coarse-grained process w.r.t. $\bar{\mu} \sim e^{-\beta \bar{H}(\eta)}$

Coarse-grained Hamiltonian $\bar{H}(\eta)$

$$e^{-\beta \bar{H}(\eta)} = \mathbb{E}[e^{-\beta H_N} \mid \eta] \equiv \int_{\chi} e^{-\beta H_N(\sigma)} P_N(d\sigma \mid \eta)$$

Approximation of RG map

Approximate $\bar{H}(\eta) \approx \bar{H}^{(0)}(\eta)$, i.e., $\bar{\mu}(d\eta) \approx \bar{\mu}^{(0)}(d\eta)$

Task: estimate & control the error in the relative entropy $\mathcal{R}(\bar{\mu} \mid \bar{\mu}^{(0)})$

$$\mathcal{R}(\pi_1 \mid \pi_2) = \int_{\chi} \log \frac{d\pi_1}{d\pi_2} \pi_1(d\sigma) \equiv \mathbb{E}_{\pi_1} \left[ \log \frac{d\pi_1}{d\pi_2} \right]$$
Example:

- $\mathcal{N}(\eta) = |\{\sigma | T\sigma = \eta\}|$ and $\nu(d\sigma|\eta)$ is uniform

Approximation at the fine level: $\mu^{\text{app}}(\sigma) = \tilde{\mu}^{(0)}(T\sigma) \frac{1}{\mathcal{N}(\eta)}$
Example:

- $\mathcal{N}(\eta) = |\{\sigma | T\sigma = \eta\}|$ and $\nu(\text{d}\sigma | \eta)$ is uniform
  
  Approximation at the fine level: $\mu^{\text{app}}(\sigma) = \bar{\mu}^{(0)}(T\sigma) \frac{1}{\mathcal{N}(\eta)}$

- The relative entropy

\[
\mathcal{R}(\mu^{\text{app}} | \mu) = \sum_{\sigma} \mu(\sigma) \log \frac{\mu(\sigma)}{\bar{\mu}^{(0)}(T\sigma)} + \sum_{\sigma} \mu(\sigma) \log \mathcal{N}(T\sigma)
\]
Example:

- $\mathcal{N}(\eta) = |\{\sigma | T\sigma = \eta\}|$ and $\nu(d\sigma|\eta)$ is uniform
  
  Approximation at the fine level: $\mu^{\text{app}}(\sigma) = \bar{\mu}^{(0)}(T\sigma) \frac{1}{\mathcal{N}(\eta)}$

- The relative entropy

  $$\mathcal{R}(\mu^{\text{app}}|\mu) = \sum_{\sigma} \mu(\sigma) \log \frac{\mu(\sigma)}{\bar{\mu}^{(0)}(T\sigma)} + \sum_{\sigma} \mu(\sigma) \log \mathcal{N}(T\sigma)$$

- Insert back $\mu(\sigma) = \frac{1}{Z} e^{-\beta H(\sigma)}$ etc.

  $$\sum_{\sigma} \beta (\bar{H}^{(0)}(T\sigma) - H(\sigma)) \frac{1}{Z} e^{-\beta H(\sigma)} + \log \frac{\bar{Z}}{Z} + \mathbb{E}_\mu[\log \mathcal{N}(T\sigma)] =$$

  $$\mathbb{E}_\mu[\beta (\bar{H}^{(0)} - H)] - \beta (\bar{A}^{(0)} - A) + \mathbb{E}_\mu[\log \mathcal{N}(T\sigma)]$$

  Helmholtz free energy $A \equiv U - TS = -\frac{1}{\beta} \log Z$
Reconstruction measures

Coarse-grained equilibrium measure:

$$\int f(\eta)\bar{\mu}(d\eta) = \int f(\mathbf{T}\sigma)\mu(d\sigma)$$

$$\mu(d\sigma) = \frac{1}{Z} e^{-\beta \bar{H}(\sigma)} P(d\sigma)$$ and $$\bar{\mu}(d\eta) = \frac{1}{Z} e^{-\beta \bar{H}(\eta)} \bar{P}(d\eta)$$

Perfect reconstruction:

$$\mu(d\sigma) = e^{-\beta (H(\sigma) - \bar{H}(\eta))} P(d\sigma|\eta)\bar{\mu}(d\eta) \equiv \mu(d\sigma|\eta)\bar{\mu}(d\eta)$$

Approximate reconstruction:

$$\mu^{\text{app}}(d\sigma) = \nu(d\sigma|\eta)\bar{\mu}^{(0)}(d\eta)$$

error = coarse-level error + reconstruction error

$$\mathcal{R}(\mu^{\text{app}}|\mu) = \mathcal{R}(\bar{\mu}^{\text{app}}|\bar{\mu}) + \int \mathcal{R}(\nu(\cdot|\eta)|\mu(\cdot|\eta))\bar{\mu}^{(0)}(d\eta)$$

Inverse Monte Carlo

“Inverse thermodynamic problems”

- Build parametrized approximations $\tilde{H}^{(0)}(\eta; \theta)$
Inverse Monte Carlo

“Inverse thermodynamic problems”

- Build parametrized approximations $\tilde{H}^{(0)}(\eta; \theta)$

- Find optimal values of parameters $\theta^*$, s.t., for selected $\phi_i$

\[
\min_\theta \sum_i |\mathbb{E}_\mu[\phi_i] - \mathbb{E}_{\tilde{\mu}^{(0)}}[\phi_i]|^2
\]

Inverse Monte Carlo

“Inverse thermodynamic problems”

- Build parametrized approximations $\tilde{H}^{(0)}(\eta; \theta)$
- Find optimal values of parameters $\theta^*$, s.t., for selected $\phi_i$

$$\min_{\theta} \sum_i |E_\mu[\phi_i] - E_{\tilde{\mu}}^{(0)}[\phi_i]|^2$$


- Parametrization depends on specific observable(s) $\phi_i$. 
Inverse Monte Carlo

“Inverse thermodynamic problems”

- Build parametrized approximations $\tilde{H}^{(0)}(\eta; \theta)$
- Find optimal values of parameters $\theta^*$, s.t., for selected $\phi_i$

$$\min_{\theta} \sum_i \left| \mathbb{E}_\mu[\phi_i] - \mathbb{E}_{\tilde{\mu}^{(0)}}[\phi_i] \right|^2$$


- Parametrization depends on specific observable(s) $\phi_i$.
- Can we improve the “transferability” of the method?
Inverse Monte Carlo

“Inverse thermodynamic problems”

- Build parametrized approximations $\bar{H}^{(0)}(\eta; \theta)$

- $\phi$ can be relative entropy

\[
\mathbb{E}_\mu [\beta (\bar{H}^{(0)}(\theta) - H)] - \beta (\bar{A}^{(0)}(\theta) - A) + \mathbb{E}_\mu [\log \mathcal{N}(T\sigma)]
\]

Optimality condition: $\nabla_\theta \mathcal{R} = \mathbb{E}_\mu [\nabla_\theta \bar{H}^{(0)}] - \mathbb{E}_{\bar{\mu}^{(0)}} [\nabla_\theta \bar{H}^{(0)}] = 0$

Shell (2008), Katsoulakis, P.P., Rey-Bellet (2008), Zabaras (2012) ...
Coarse Graining & Parameterization

Gibbs measure: \( \mu \sim e^{-\beta H(\sigma)} \)

Coarse-grained Hamiltonian \( \tilde{H}(\eta) \)

\[
\tilde{\mu} \sim e^{-\beta \tilde{H}(\eta)} = \mathbb{E}[e^{-\beta H_N} \mid \eta] \equiv \int_{\chi} e^{-\beta H_N(\sigma)} P_N(d\sigma \mid \eta)
\]

Approximate \( \tilde{H}(\eta) \approx \tilde{H}^{(0)}(\eta) \), i.e., \( \tilde{\mu}(d\eta) \approx \tilde{\mu}^{(0)}(d\eta) \)

Task: estimate & control the error in the relative entropy \( R(\tilde{\mu} \mid \tilde{\mu}^{(0)}) \)
Coarse Graining & Parameterization

Gibbs measure: \( \mu \sim e^{-\beta H(\sigma)} \)

Coarse-grained Hamiltonian \( \tilde{H}(\eta) \)

\[
\tilde{\mu} \sim e^{-\beta \tilde{H}(\eta)} = \mathbb{E}[e^{-\beta H_N} | \eta] \equiv \int_\mathcal{X} e^{-\beta H_N(\sigma)} P_N(d\sigma | \eta)
\]

Approximate \( \tilde{H}(\eta) \approx \tilde{H}^{(0)}(\eta) \), i.e., \( \tilde{\mu}(d\eta) \approx \mu^{(0)}(d\eta) \)

Task: estimate & control the error in the relative entropy \( \mathcal{R}(\mu | \mu^{(0)}) \)

Optimal parameters for parametrized potentials: \( \tilde{H}^{(0)}(\sigma; \theta) \)

\[
\min_\theta \mathcal{R}(\mu | \mu^{\text{app}}(\theta)) \quad \text{or} \quad \min_\theta \mathcal{R}(\mu^{\text{app}}(\theta) | \mu)
\]
Coarse Graining & Parameterization

Gibbs measure: \( \mu \sim e^{-\beta H(\sigma)} \)

Coarse-grained Hamiltonian \( \bar{H}(\eta) \)

\[
\bar{\mu} \sim e^{-\beta \bar{H}(\eta)} = \mathbb{E}[e^{-\beta H_N} | \eta] \equiv \int_\chi e^{-\beta H_N(\sigma)} P_N(d\sigma | \eta)
\]

Approximate \( \bar{H}(\eta) \approx \bar{H}^{(0)}(\eta) \), i.e., \( \bar{\mu}(d\eta) \approx \bar{\mu}^{(0)}(d\eta) \)

Task: estimate & control the error in the relative entropy \( \mathcal{R}(\bar{\mu}|\bar{\mu}^{(0)}) \)

Optimal parameters for parametrized potentials: \( \bar{H}^{(0)}(\sigma; \theta) \)

\[
\min_\theta \mathcal{R}(\mu|\mu^{\text{app}}(\theta)) \quad \text{or} \quad \min_\theta \mathcal{R}(\mu^{\text{app}}(\theta)|\mu)
\]

Gibbs structure allows explicit calculations of \( \mathcal{R} \)
Parametrized CG – Approximations heuristics

\[ \mathcal{R} \left( \bar{\mu} \mid \bar{\mu}^{(0)} \right) \sim \mathbb{E}_\mu [\beta (\bar{H}^{(0)}(\theta) - H)] + \log \frac{\bar{Z}^0(\theta)}{Z} \]

Optimality condition: \( \nabla_\theta \mathcal{R} = 0 \)


- Is the parametric family \( \bar{H}^{(0)}(\theta) \) rich enough?.
Cluster Expansion CG Hamiltonians


\[ \bar{H}_m(\eta) = \bar{H}^{(0)}_m(\eta) + \bar{H}^{(1)}_m(\eta) + \ldots \]

Multi-body terms:

\[ \bar{H}^{(1)}(\eta) = \beta \sum_{k_1} \sum_{k_2 > k_1} \sum_{k_3 > k_2} j^{(2)}_{k_1 k_2 k_3} \left( -E_1(k_1) E_2(k_2) E_1(k_3) \right) + \ldots \]
Cluster Expansion CG Hamiltonians


\[ \bar{H}_m(\eta) = \bar{H}_m^{(0)}(\eta) + \bar{H}_m^{(1)}(\eta) + \ldots \]

Multi-body terms:

\[ \bar{H}^{(1)}(\eta) = \beta \sum_{k_1} \sum_{k_2 > k_1} \sum_{k_3 > k_2} j_{k_1 k_2 k_3}^{(2)} (-E_1(k_1)E_2(k_2)E_1(k_3)) + \ldots \]

Typically omitted, but essential to capture phase transitions
Cluster Expansion CG Hamiltonians


\[ \tilde{H}_m(\eta) = \tilde{H}_m^{(0)}(\eta) + \tilde{H}_m^{(1)}(\eta) + ... \]

Multi-body terms:

\[ \tilde{H}^{(1)}(\eta) = \beta \sum_{k_1} \sum_{k_2 > k_1} \sum_{k_3 > k_2} [j^{(2)}_{k_1 k_2 k_3} (-E_1(k_1) E_2(k_2) E_1(k_3)) + ... \]
Cluster Expansion CG Hamiltonians


\[ \tilde{H}_m(\eta) = \tilde{H}_m^{(0)}(\eta) + \tilde{H}_m^{(1)}(\eta) + ... \]

Multi-body terms:

\[ \tilde{H}^{(1)}(\eta) = \beta \sum_{k_1} \sum_{k_2 > k_1} \sum_{k_3 > k_2} [j_{k_1 k_2 k_3}^{(2)} (-E_1(k_1)E_2(k_2)E_1(k_3)) + ... ] \]

Typically omitted, but essential to capture switching times etc.
Dynamics: Lattice Models

- Continuous Time Markov Chain \( \{\sigma_t\}_{t \geq 0}, \mathcal{L} \)
  \( \sigma \in \Sigma \equiv \{0, 1\}^{\Lambda_N}, \Lambda_N \subset \mathbb{Z}^d \)

\[
\mathbb{P}(\sigma_{t+\delta t} = \sigma' | \sigma_t = \sigma) = c(\sigma, \sigma') \delta t + o(\delta t)
\]

Rates: \( c(\sigma, \sigma') \equiv c(x, \omega; \sigma) \)
Dynamics: Lattice Models

- Continuous Time Markov Chain \( \{\sigma_t\}_{t \geq 0}, \mathcal{L} \)
  \( \sigma \in \Sigma \equiv \{0, 1\}^{\Lambda_N}, \Lambda_N \subset \mathbb{Z}^d \)

\[
\mathbb{P}(\sigma_{t+\delta t} = \sigma' | \sigma_t = \sigma) = c(\sigma, \sigma') \delta t + o(\delta t)
\]

Rates: \( c(\sigma, \sigma') \equiv c(x, \omega; \sigma) \)

- Forward Kolmogorov Equation (aka Master Equation)

\[
\partial_t P(\sigma, t; \zeta) = \sum_{\sigma', \sigma' \neq \sigma} c(\sigma', \sigma) P(\sigma', t; \zeta) - \lambda(\sigma) P(\sigma, t; \zeta),
\]

\[
P(\sigma, 0; \zeta) = \delta(\sigma - \zeta)
\]
Dynamics : Lattice Models

- Continuous Time Markov Chain \( \{\sigma_t\}_{t \geq 0}, \mathcal{L} \)
  \( \sigma \in \Sigma \equiv \{0, 1\}^{\Lambda_N}, \Lambda_N \subset \mathbb{Z}^d \)

  \[ \mathbb{P}(\sigma_{t+\delta t} = \sigma' | \sigma_t = \sigma) = c(\sigma, \sigma') \delta t + o(\delta t) \]

  Rates: \( c(\sigma, \sigma') \equiv c(x, \omega; \sigma) \)

- Forward Kolmogorov Equation (aka Master Equation)

  \[ \partial_t P(\sigma, t; \zeta) = \sum_{\sigma', \sigma' \neq \sigma} c(\sigma', \sigma) P(\sigma', t; \zeta) - \lambda(\sigma) P(\sigma, t; \zeta), \]

  \[ P(\sigma, 0; \zeta) = \delta(\sigma - \zeta) \]

- Simulation: Embedded Markov Chain \( \{X_n\}_{n \geq 0} = \{\sigma_n\delta t\}, \sigma \rightarrow \sigma^{x,\omega} \)

  \[ p(\sigma, \sigma^{x,\omega}) = \frac{c(x, \omega; \sigma)}{\lambda(\sigma)}, \quad \lambda(\sigma) = \sum_x \sum_\omega c(x, \omega; \sigma) \]

  Exponential clock: \( \delta t \sim \text{Exp}(\lambda(\sigma)) \)
Generator of the process

- Evolution of observables (Backward Kolomogorov Equation):

\[
    u(\zeta, t) = \mathbb{E}_\zeta[f(\sigma_t)] \equiv \sum_\sigma f(\sigma) P(\sigma, t; \zeta)
\]

\[
    \partial_t u(\zeta, t) = \mathcal{L} u(\zeta, t), \quad u(\zeta, 0) = f(\zeta)
\]
Generator of the process

- Evolution of observables (Backward Kolomogorov Equation):

\[
    u(\zeta, t) = \mathbb{E}_\zeta[f(\sigma_t)] = \sum_{\sigma} f(\sigma) P(\sigma, t; \zeta)
\]

\[
    \partial_t u(\zeta, t) = \mathcal{L} u(\zeta, t), \quad u(\zeta, 0) = f(\zeta)
\]

- Generator:

\[
    \mathcal{L} f(\sigma) = \sum_{\sigma'} c(\sigma, \sigma')[f(\sigma') - f(\sigma)] = \sum_{x} \sum_{\omega} c(x, \omega; \sigma)[f(\sigma^{x,\omega}) - f(\sigma)]
\]

- Markov semigroup \( P_t = e^{t\mathcal{L}} \)

\[
    \delta_x f(\sigma) = f(\sigma^{x,\omega}) - f(\sigma)
\]

\[
    C_\infty(\Sigma) = \{ f \in C_b(\Sigma) | \sum_{x} ||\delta_x(f)||_\infty < \infty \} 
\]
Events

- Adsorption/desorption:
  \[
  \sigma^x = \begin{cases} 
  1 - \sigma(x) & \text{if } z = x \\
  \sigma(z) & \text{if } z \neq x 
  \end{cases}
  \]

- Diffusion (spin exchange, Kawasaki dynamics):
  \[
  \sigma^{x,y}(z) = \begin{cases} 
  \sigma(y) & \text{if } z = x \\
  \sigma(x) & \text{if } z = y \\
  \sigma(z) & \text{if } x \neq y 
  \end{cases}
  \]

- Multicomponent reactions.
  \[
  \sigma(x) \in \{0, 1, \ldots, K\}
  \]
  \[
  \sigma^{(x,k)}(z) = \begin{cases} 
  \sigma(z) & \text{if } z \neq x, y, \\
  k & \text{if } z = x. 
  \end{cases}
  \]

- Reactions involving particles with internal degrees of freedom.
  \[
  \sigma^{(x,y,k,l)}(z) = \begin{cases} 
  \sigma(z) & \text{if } z \neq x, y, \\
  k & \text{if } z = x, \\
  l & \text{if } z = y, 
  \end{cases}
  \]
Non-equilibrium steady states

Behavior as $t \to \infty$

$$\partial_t P(\sigma, t; \zeta) = \sum_{\sigma'} [c(\sigma', \sigma) P(\sigma', t; \zeta) - c(\sigma, \sigma') P(\sigma, t; \zeta)] ,$$

Stationary states: $\partial_t P = 0 \implies \sum_{\sigma'} j_s(\sigma', \sigma) = 0$

Current $\sigma' \to \sigma$: $j_s(\sigma', \sigma) = c(\sigma', \sigma) \mu(\sigma') - c(\sigma, \sigma') \mu(\sigma)$

Reversible dynamics with the equilibrium $\mu(\sigma)$

Detailed Balance condition with respect to $\mu(\sigma)$ (e.g., $\mu \sim e^{-\beta H(\sigma)}$)

$$c(\sigma', \sigma) \mu(\sigma') = c(\sigma, \sigma') \mu(\sigma)$$

Irreversible dynamics $\implies$ Non-equilibrium steady states

$$\sum_{\sigma'} j_s(\sigma', \sigma) = \sum_{\sigma'} (c(\sigma', \sigma) \mu(\sigma') - c(\sigma, \sigma') \mu(\sigma)) = 0$$

irreversible rate loops, i.e., a non-zero current at stationary states.
**Example:** Arrhenius diffusion of interacting particles

- **Rates:** Exchange dynamics with the migration rate to n.n. site.

\[ |x - y| = 1 \quad c(x, y, \sigma) = de^{-\beta(U_0 + U(x, \sigma))} \sum_{|y-x|=1} \sigma(x)(1 - \sigma(y)) \]

- **Energy barrier:**

\[ U(x, \sigma) = \sum_{z \neq x} J(x - z)\sigma(z) - h \]

- **Events**

\[ \sigma^{x,y}(z) = \sigma(y), \quad z = x \]

\[ \sigma^{x,y}(z) = \sigma(x), \quad z = y \]

- **Periodic lattice \implies detailed balance w.r.t.**

\[ \mu \sim e^{-\beta H(\sigma)} \]

\[ c(x, y, \sigma)e^{-\beta H(\sigma)} = c(x, y, \sigma^{x,y})e^{-\beta H(\sigma^{x,y})} \]

**Bounded domain with a gradient in concentrations**

![Diagram of bounded domain with gradient in concentrations](chart.png)
Coverage profile

High temperature regime $\beta = 1.0$

Coverage profile at initial and optimal $\beta$

$\beta J_0 = 4.00$
$q=32 \quad \beta J_0 = 5.70$
$q=16 \quad \beta J_0 = 4.53$
$q=8 \quad \beta J_0 = 4.16$
$q=4 \quad \beta J_0 = 4.04$
Inverse Monte Carlo

“Inverse thermodynamic problems”

- Build parametrised approximations $\bar{H}^{(0)}(\eta; \theta)$
- Find optimal values of parameters $\theta^*$, s.t., for selected $\phi_i$

$$\min_{\theta} \sum_i \left| \mathbb{E}_\mu [\phi_i] - \mathbb{E}_{\bar{\mu}^{(0)}} [\phi_i] \right|^2$$


- $\phi$ can be relative entropy

$$\mathbb{E}_\mu [\beta (\bar{H}^{(0)}(\theta) - H)] - \beta (\bar{A}^{(0)}(\theta) - A) + \mathbb{E}_\mu [\log \mathcal{N}(T\sigma)]$$

optimality condition: $\nabla_\theta \mathcal{R} = \mathbb{E}_\mu [\nabla_\theta \bar{H}^{(0)}] - \mathbb{E}_{\bar{\mu}^{(0)}} [\nabla_\theta \bar{H}^{(0)}] = 0$

Shell (2008), Katsoulakis, P.P., Rey-Bellet (2008), Zabaras (2012) ...
Parametric estimation

\[ X_i \sim p(\sigma; \hat{\theta}) \text{ and } X \sim p(\sigma; \theta) \]

Log-likelihood function: \( \log L(\theta) = \sum_{i=1}^{N} \log p(X_i; \theta) \)

Max-likelihood \( \iff \min_\theta \mathcal{R} \left( p(\cdot; \theta) \middle| p(\cdot; \hat{\theta}) \right) \)

\[
\frac{1}{N} \log L(\hat{\theta}) - \text{const.} \\
= \frac{1}{N} \sum_{i=1}^{N} p(X_i; \hat{\theta}) - \int p(\sigma; \theta) \log p(\sigma; \theta) \, d\sigma \\
\rightarrow \int p(\sigma; \theta) \log p(\sigma; \hat{\theta}) \, d\sigma - \int p(\sigma; \theta) \log p(\sigma; \theta) \, d\sigma \\
= \int p(\sigma; \theta) \log \frac{p(\sigma; \hat{\theta})}{p(\sigma; \theta)} \, d\sigma \\
= -\mathcal{R} \left( p(\sigma; \theta) \middle| p(\sigma; \hat{\theta}) \right)
\]
Equilibrium: CG measure $\bar{\mu}_\theta \sim e^{-\bar{H}(\sigma; \theta)}$
Equilibrium: CG measure $\bar{\mu}_\theta \sim e^{-\bar{H}(\sigma; \theta)}$

Parametric estimation in the family of $\bar{H}(\cdot; \theta)$ yields effective potentials
- **Equilibrium:** CG measure $\bar{\mu}_\theta \sim e^{-\bar{H}(\sigma;\theta)}$

- Parametric estimation in the family of $\bar{H}(\cdot;\theta)$ yields effective potentials minimize the “distance” in $\mathcal{R}(\bar{\mu}|\bar{\mu}_\theta)$
Equilibrium: CG measure $\tilde{\mu}_\theta \sim e^{-\tilde{H}(\sigma;\theta)}$

Parametric estimation in the family of $\tilde{H}(\cdot;\theta)$ yields effective potentials minimize the “distance” in $\mathcal{R}(\bar{\mu}|\bar{\mu}_\theta)$

Non-equilibrium steady states & stationary processes?
Relative entropy on the path space

Markov chains on $\mathcal{X}$:

$\{\sigma_n\}_{n \in \mathbb{Z}^+}$, $P^\theta(\sigma, d\sigma)$, $\mu^\theta(\sigma)$

$\{\tilde{\sigma}_n\}_{n \in \mathbb{Z}^+}$, $\tilde{P}^\theta(\sigma, d\sigma)$, $\tilde{\mu}^\theta(\sigma)$

Path measures:

$$Q^\theta(\sigma_0, \ldots, \sigma_M) = \mu^\theta(\sigma_0) p^\theta(\sigma_0, \sigma_1) \ldots p^\theta(\sigma_{M-1}, \sigma_M)$$

Radon-Nikodym derivative

$$\frac{dQ^\theta}{d\tilde{Q}^\theta}(\{\sigma_n\}) = \frac{\mu^\theta(\sigma_0) \prod_{i=0}^{M-1} p^\theta(\sigma_i, \sigma_{i+1})}{\tilde{\mu}^\theta(\sigma_0) \prod_{i=0}^{M-1} \tilde{p}^\theta(\sigma_i, \sigma_{i+1})}$$

Relative entropy

$$\mathcal{R}(Q^\theta \mid \tilde{Q}^\theta) = \int_{\mathcal{X}} \ldots \int_{\mathcal{X}} \mu^\theta(\sigma_0) \prod_{i=0}^{M-1} p^\theta(\sigma_i, \sigma_{i+1}) \log \frac{\mu^\theta(\sigma_0) \prod_{i=0}^{M-1} p^\theta(\sigma_i, \sigma_{i+1})}{\tilde{\mu}^\theta(\sigma_0) \prod_{i=0}^{M-1} \tilde{p}^\theta(\sigma_i, \sigma_{i+1})} d\sigma$$
\[
\mathcal{R}(\mathcal{Q}^\theta | \tilde{\mathcal{Q}}^\theta) = \int_\mathcal{X} \ldots \int_\mathcal{X} \mu^\theta(\sigma_0) \prod_{i=0}^{M-1} p^\theta(\sigma_i, \sigma_{i+1}) \left( \log \frac{\mu^\theta(\sigma_0)}{\tilde{\mu}^\theta(\sigma_0)} \right) + \sum_{i=0}^{i=M-1} \log \frac{p^\theta(\sigma_i, \sigma_{i+1})}{\tilde{p}^\theta(\sigma_i, \sigma_{i+1})} \right) d\sigma_0 \ldots d\sigma_M
\]
\[ R \left( Q^\theta \mid \tilde{Q}^\theta \right) = \int_x \ldots \int_x \mu^\theta(\sigma_0) \prod_{i=0}^{M-1} p^\theta(\sigma_i, \sigma_{i+1}) \left( \log \frac{\mu^\theta(\sigma_0)}{\tilde{\mu}^\theta(\sigma_0)} \right. \\
\left. + \sum_{i=0}^{i=M-1} \log \frac{p^\theta(\sigma_i, \sigma_{i+1})}{\tilde{p}^\theta(\sigma_i, \sigma_{i+1})} \right) \, d\sigma_0 \ldots d\sigma_M \\
\int_x p(\sigma, \sigma') \, d\sigma' = 1, \quad \int_x \mu(\sigma) p(\sigma, \sigma') \, d\sigma = \mu(\sigma') \]
\[
R \left( Q^\theta | \tilde{Q}^\theta \right) = \int_\mathcal{X} \ldots \int_\mathcal{X} \mu^\theta(\sigma_0) \prod_{i=0}^{M-1} p^\theta(\sigma_i, \sigma_{i+1}) \left( \log \frac{\mu^\theta(\sigma_0)}{\tilde{\mu}^\theta(\sigma_0)} + \sum_{i=0}^{i=M-1} \log \frac{p^\theta(\sigma_i, \sigma_{i+1})}{\tilde{p}^\theta(\sigma_i, \sigma_{i+1})} \right) \, d\sigma_0 \ldots d\sigma_M
\]

\[
\int_\mathcal{X} p(\sigma, \sigma') \, d\sigma' = 1, \quad \int_\mathcal{X} \mu(\sigma)p(\sigma, \sigma') \, d\sigma = \mu(\sigma')
\]

\[
\int_\mathcal{X} \mu^\theta(\sigma_0) \log \frac{\mu^\theta(\sigma_0)}{\tilde{\mu}^\theta(\sigma_0)} \, d\sigma_0 + \sum_{i=0}^{M-1} \int_\mathcal{X} \int_\mathcal{X} \mu^\theta(\sigma_i)p^\theta(\sigma_i) \log \frac{p^\theta(\sigma_i, \sigma_{i+1})}{\tilde{p}^\theta(\sigma_i, \sigma_{i+1})} \, d\sigma_i
\]

\[
= M \mathbb{E}_{\mu} \left[ \int_\mathcal{X} p^\theta(\sigma, \sigma') \log \frac{p^\theta(\sigma, \sigma')}{\tilde{p}^\theta(\sigma, \sigma')} \, d\sigma' \right] + R \left( \mu^\theta | \tilde{\mu}^\theta \right)
\]
\[
\mathcal{R}(Q^{\theta} | \tilde{Q}^{\theta}) = \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} \mu^{\theta}(\sigma_0) \prod_{i=0}^{M-1} p^{\theta}(\sigma_i, \sigma_{i+1}) \left( \log \frac{\mu^{\theta}(\sigma_0)}{\tilde{\mu}^{\theta}(\sigma_0)} \right. \\
+ \sum_{i=0}^{i=M-1} \log \frac{p^{\theta}(\sigma_i, \sigma_{i+1})}{\bar{p}^{\theta}(\sigma_i, \sigma_{i+1})} \left. \right) d\sigma_0 \cdots d\sigma_M
\]

\[
\int_{\mathcal{X}} p(\sigma, \sigma') d\sigma' = 1, \quad \int_{\mathcal{X}} \mu(\sigma) p(\sigma, \sigma') d\sigma = \mu(\sigma')
\]

\[
\int_{\mathcal{X}} \mu^{\theta}(\sigma_0) \log \frac{\mu^{\theta}(\sigma_0)}{\tilde{\mu}^{\theta}(\sigma_0)} d\sigma_0 + \sum_{i=0}^{M-1} \int_{\mathcal{X}} \int_{\mathcal{X}} \mu^{\theta}(\sigma_i) p^{\theta}(\sigma_i) \log \frac{p^{\theta}(\sigma_i, \sigma_{i+1})}{\bar{p}^{\theta}(\sigma_i, \sigma_{i+1})}
\]

\[
= M \mathbb{E}_{\mu}^{\theta} \left[ \int_{\mathcal{X}} p^{\theta}(\sigma, \sigma') \log \frac{p^{\theta}(\sigma, \sigma')}{\bar{p}^{\theta}(\sigma, \sigma')} d\sigma' \right] + \mathcal{R}(\mu^{\theta} | \tilde{\mu}^{\theta})
\]

\[
\mathcal{R}(Q^{\theta} | \tilde{Q}^{\theta}) = M \mathcal{H}(Q^{\theta} | \tilde{Q}^{\theta}) + \mathcal{R}(\mu^{\theta} | \tilde{\mu}^{\theta})
\]
Continuous time Markov chain

$\mathcal{D}_{[0,T]} \ (\text{resp. } \tilde{\mathcal{D}}_{[0,T]})$ is the distribution of the process $\{\sigma_t\}_{t \in [0,T]} \ (\text{resp. } \{\tilde{\sigma}_t\}_{t \in [0,T]})$ on the path space $Q([0, T], \Sigma_N)$

$$R \left( \mathcal{D}_{[0,T]} | \tilde{\mathcal{D}}_{[0,T]} \right) = \int \log \left( \frac{d\mathcal{D}_{[0,T]}}{d\tilde{\mathcal{D}}_{[0,T]}} \right) d\mathcal{D}_{[0,T]},$$

The initial distribution is the stationary measure $\mu$ (resp. $\tilde{\mu}$).

Radon-Nikodym derivative:

$$\frac{d\mathcal{D}_{[0,T]}}{d\tilde{\mathcal{D}}_{[0,T]}} = \frac{\mu(\sigma_0)}{\tilde{\mu}(\sigma_0)} \exp \left\{ - \int_0^T \left[ \lambda(\sigma_s) - \tilde{\lambda}(\sigma_s) \right] ds + \int_0^T \log \frac{c(\sigma_{s-}, \sigma_s)}{\tilde{c}(\sigma_{s-}, \sigma_s)} dN_s \right\}$$

$N_s(\rho)$ — the number of jumps of the path $\sigma_s$ up to time $s$. 
(a) $N_t - \int_0^t \lambda(\rho_s) \, ds$ is a (zero mean) martingale

(b) exchanging $\int_0^T$ and $\mathbb{E}[\cdot]$

(c) stationarity

$$\mathbb{E}_D \left[ \int_0^T \phi(\rho_s) \, dN_s(\rho) \right] = \mathbb{E}_D \left[ \int_0^T \phi(\rho_s) \lambda(\rho_s) \, ds \right] = T \mathbb{E}_\mu[\phi \lambda],$$

Hence:

$$\mathcal{R} \left( \mathcal{D}_{[0, T]} | \tilde{\mathcal{D}}_{[0, T]} \right) = \mathbb{E}_D \left[ \log \frac{d\mathcal{D}_{[0, T]}}{d\tilde{\mathcal{D}}_{[0, T]}} \right] = \mathbb{E}_D \left[ \log \frac{\mu}{\tilde{\mu}} \right]
+ \mathbb{E}_D \left[ - \int_0^T \left[ \lambda(\sigma_s) - \tilde{\lambda}(\sigma_s) \right] \, ds + \int_0^T \lambda(\sigma_{s-}) \log \frac{c(\sigma_{s-}, \sigma_s)}{c(\sigma_{s-}, \sigma_s)} \, ds \right]
= \mathbb{E}_\mu \left[ \lambda(\sigma) - \tilde{\lambda}(\sigma) - \sum_{\sigma'} \lambda(\sigma) p(\sigma, \sigma') \log \frac{\lambda(\sigma) p(\sigma, \sigma')}{\tilde{\lambda}(\sigma) \tilde{p}(\sigma, \sigma')} \right] + \mathcal{R}(\mu | \tilde{\mu})
= T \mathcal{H}(\mathcal{D}_{[0, T]} | \tilde{\mathcal{D}}_{[0, T]}) + \mathcal{R}(\mu | \tilde{\mu})
Parametrization of rates

“Best-fit” Markovian dynamics

\[
\mathcal{R} \left( \mathcal{D}_{[0,T]} | \tilde{\mathcal{D}}_{[0,T]} \right) = T \mathcal{H}(\mathcal{D}_{[0,T]} | \tilde{\mathcal{D}}_{[0,T]}) + \mathcal{R}(\mu | \tilde{\mu})
\]

Find \( \theta \) in \( \bar{c}(\eta; \theta) \), s.t.,

\[
\min_{\theta} \{ \mathcal{H}(\mathcal{D}_{[0,T]} | \tilde{\mathcal{D}}_{[0,T]}^{\theta}) \}
\]

\[
\mathcal{H}(\mathcal{D}_{[0,T]} | \tilde{\mathcal{D}}_{[0,T]}^{\theta}) = \mathbb{E}_\mu \left[ \sum_{\sigma'} c(\sigma, \sigma') \log \frac{c(\sigma, \sigma')}{\bar{c}(\sigma, \sigma'; \theta)} - (\lambda(\sigma) - \tilde{\lambda}(\sigma; \theta)) \right]
\]

Optimality condition: \( \nabla_{\theta} \mathcal{H}(\mathcal{D}_{[0,T]} | \tilde{\mathcal{D}}_{[0,T]}^{\theta}) = 0 \)

\[
\nabla_{\theta} \mathcal{H}(\mathcal{D}_{[0,T]} | \tilde{\mathcal{D}}_{[0,T]}^{\theta}) = \mathbb{E}_\mu \left[ \sum_{\sigma'} \left[ -c(\sigma, \sigma') \nabla_{\theta} \bar{c}(\sigma, \sigma'; \theta) + \nabla_{\theta} \bar{c}(\sigma, \sigma'; \theta) \right] \right]
\]
Inverse Dynamic Monte Carlo

- Best-fit obtained by minimizing RER

\[ \theta^* = \arg \min_{\theta} \mathcal{H}(P \mid Q^{\theta}), \]

- Optimality condition \( \nabla_{\theta} \mathcal{H}(P \mid Q^{\theta}) = 0 \); minimization scheme:

\[ \theta^{(n+1)} = \theta^{(n)} - \frac{\alpha}{n} G^{(n+1)}, \]

\( \alpha > 0 \) and \( G^{(n+1)} \) being a suitable approximation of the gradient \( \nabla_{\theta} \mathcal{H}(P \mid Q^{\theta}) \)
Inverse Dynamic Monte Carlo

- Best-fit obtained by minimizing RER

\[ \theta^* = \arg \min_{\theta} \mathcal{H}(P \mid Q^\theta), \]

- Optimality condition \( \nabla_{\theta} \mathcal{H}(P \mid Q^\theta) = 0 \); minimization scheme:

\[ \theta^{(n+1)} = \theta^{(n)} - \frac{\alpha}{n} G^{(n+1)}, \]

\( \alpha > 0 \) and \( G^{(n+1)} \) being a suitable approximation of the gradient \( \nabla_{\theta} \mathcal{H}(P \mid Q^\theta) \)

- FIM revisited-Newton-Raphson:

\[ G^n = \text{Hess}(\mathcal{H}(P \mid Q^\theta^n))^{-1} \nabla_{\theta} \mathcal{H}(P \mid Q^\theta^n). \]

\[ \mathbf{F}_{\mathcal{H}}(Q^\theta) = \text{Hess}(\mathcal{H}(P \mid Q^\theta)) = -\mathbb{E}_{\mu} \left[ \sum_{\sigma'} p(\sigma, \sigma') \nabla^2_{\theta} \log q^{\theta}(\sigma, \sigma') \right]. \]
Data-based parametrization of CG dynamics

- Unbiased estimator for RER,

\[
\hat{H}_N(P|Q^\theta) := \frac{1}{N} \sum_{i=1}^{N} \log \frac{p(\sigma_i, \sigma_{i+1})}{q^\theta(\sigma_i, \sigma_{i+1})},
\]

- Minimization of RER:

\[
\min_\theta \hat{H}_N(P|Q^\theta) = \max_\theta \frac{1}{N} \sum_{i=1}^{N} \log q^\theta(\sigma_i, \sigma_{i+1}) - \frac{1}{N} \sum_{i=1}^{N} \log p(\sigma_i, \sigma_{i+1}),
\]

- Coarse-grained path space Log-Likelihood maximization

\[
\max_\theta L(\theta; \{\sigma_i\}_{i=0}^{N}) := \max_\theta \frac{1}{N} \sum_{i=1}^{N} \log \bar{p}^\theta(T\sigma_i, T\sigma_{i+1}).
\]

- No need for microscopic reconstruction:

\[
q^\theta(\sigma, \sigma') = \nu(\sigma'|T\sigma')\bar{p}^\theta(T\sigma, T\sigma').
\]
Fisher Information Matrix and Parameter Identifiability

Since RER is a relative entropy, \( \mathcal{H}(P \mid Q) = \mathcal{R}(\mu \otimes p \mid \mu \otimes q) \):

- Asymptotic Gaussianity of the **Maximum Likelihood Estimator**:

\[
\hat{\theta}_N \to \theta^* \text{ a.s. and } N^{1/2}(\hat{\theta}_N - \theta^*) \to N(0, \mathbf{F}_\mathcal{H}^{-1}(Q^{\theta^*})),
\]

- Variance determined by the path-space FIM \( \mathbf{F}_\mathcal{H}(Q^{\theta^*}) \), or asymptotically by \( \mathbf{F}_\mathcal{H}(Q^{\hat{\theta}_N}) \).

- Estimating the FIM \( \mathbf{F}_\mathcal{H}(Q^{\hat{\theta}_N}) \) provides rigorous error bars on computed optimal parameter values \( \theta^* \).

Stochastic optimization

Effective rates – Markovian CG dynamics

\[ \theta^* = \arg \min_{\theta} \mathcal{H}(D | \tilde{D}^\theta) \]
Stochastic optimization
Effective rates – Markovian CG dynamics

\[ \theta^* = \arg \min_{\theta} \mathcal{H}(D|\tilde{D}^\theta) \]

\[ \theta^{(n)} \rightarrow \theta^* \text{ a.s. and } n^{1/2}(\theta^{(n)} - \theta^*) \rightarrow N(0, \mathbb{F}^{-1}) \]

Fisher information matrix (FIM):

\[ F_{ij} = \mathbb{E}_\mu[\partial_{\theta_i} \log p^\theta \partial_{\theta_j} \log p^\theta] \]
Stochastic optimization
Effective rates – Markovian CG dynamics

Fisher Information Matrix per unit time

\[ F_{\mathcal{H}} = \lim_{T \to \infty} \frac{1}{T} F_{\mathcal{R}} \]

\[ \mathcal{H} (Q^\theta | Q^{\theta+\epsilon}) = \frac{1}{2} \epsilon^T F_{\mathcal{H}} (Q^\theta) \epsilon + \mathcal{O}(|\epsilon|^3) \]

\[ F_{\mathcal{H}} = \mathbb{E}_{\mu^\theta} \left[ \sum_{\sigma'} c^\theta (\sigma, \sigma') \nabla_\theta \log c^\theta (\sigma, \sigma') \nabla_\theta \log c^\theta (\sigma, \sigma')^T \right] \]

Cramer-Rao bound – Efficiency (if \( \hat{\theta}^* \) unbiased)

\[ \text{Cov} \hat{\theta}^* \geq F_{\mathcal{H}}^{-1} \]
Relative Entropy Rate (RER) $\mathcal{H}$

\[
\mathcal{R} \left( Q^\theta | \tilde{Q}^\theta \right) = M \mathcal{H} (Q^\theta | \tilde{Q}^\theta) + \mathcal{R} (\mu^\theta | \tilde{\mu}^\theta)
\]

\[
\mathcal{H} (Q^\theta | \tilde{Q}^\theta) = \mathbb{E}_\mu^\theta \left[ \int_{\chi} p^\theta (\sigma, \sigma') \log \frac{p^\theta (\sigma, \sigma')}{\tilde{p}^\theta (\sigma, \sigma')} d\sigma' \right]
\]
Relative Entropy Rate (RER) \( \mathcal{H} \)

\[
\mathcal{R} \left( Q^\theta \| \tilde{Q}^\theta \right) = M \mathcal{H}(Q^\theta \| \tilde{Q}^\theta) + \mathcal{R}(\mu^\theta \| \tilde{\mu}^\theta)
\]

\[
\mathcal{H}(Q^\theta \| \tilde{Q}^\theta) = \mathbb{E}^\theta_{\mu} \left[ \int_{\chi} p^\theta(\sigma, \sigma') \log \frac{p^\theta(\sigma, \sigma')}{\tilde{p}^\theta(\sigma, \sigma')} d\sigma' \right]
\]

- \textbf{RER} is an observable \( \Rightarrow \) tractable and statistical estimators are available. 
  \cite{Pantazis, Katsoulakis, J. Chem. Phys. (2013)}
- Contains information not only for the invariant measure but also for the dynamics.
- No need for explicit knowledge of NESS (stationary measure): suitable for reaction networks, driven and/or reaction-diffusion systems, etc.
Examples: Statistical estimators

\[
\mathcal{H}(\mathcal{D}_{[0,T]}|\tilde{\mathcal{D}}_{[0,T]}) = \mathbb{E}_\mu \left[ \sum_{\sigma'} c(\sigma, \sigma') \log \frac{c(\sigma, \sigma')}{\tilde{c}(\sigma, \sigma'; \theta)} - (\lambda(\sigma) - \tilde{\lambda}(\sigma; \theta)) \right]
\]

Estimator I:

\[
\hat{\mathcal{H}}_1^{(n)} = \frac{1}{T} \sum_{k=0}^{n-1} \delta \tau_i \left[ \sum_{\sigma'} c(\sigma_k, \sigma') \log \frac{c(\sigma_k, \sigma')}{\tilde{c}(\sigma_k, \sigma')} - (\lambda(\sigma_k) - \tilde{\lambda}(\sigma_k)) \right]
\]

Estimator II: (using Girsanov formula)

\[
\hat{\mathcal{H}}_2^{(n)} = \frac{1}{n} \sum_{k=0}^{n-1} \log \frac{c(\sigma_k, \sigma_{k+1})}{\tilde{c}(\sigma_k, \sigma_{k+1})} - \frac{1}{T} \sum_{k=0}^{n-1} \delta \tau_k (\lambda(\sigma_k) - \tilde{\lambda}(\sigma_k))
\]

Multi-scale Diffusions and Stochastic Averaging

- Coarse-graining for diffusion processes on $\mathbb{R}^n \times \mathbb{R}^m$

$$dX_t = a(X_t) \, dt + \sqrt{2\beta^{-1}} \, dW_t, \quad X_0 = x.$$
Multi-scale Diffusions and Stochastic Averaging

- Coarse-graining for diffusion processes on $\mathbb{R}^n \times \mathbb{R}^m$

$$dX_t = a(X_t) \, dt + \sqrt{2\beta^{-1}} \, dW_t, \quad X_0 = x.$$ 

- Approximating Markov Chain:

$$X^{n+1} = X^n + a(X^n)h + \sqrt{2\beta^{-1}}Z\sqrt{h},$$
Multi-scale Diffusions and Stochastic Averaging

- Coarse-graining for diffusion processes on $\mathbb{R}^n \times \mathbb{R}^m$

$$dX_t = a(X_t) \, dt + \sqrt{2\beta^{-1}} \, dW_t, \quad X_0 = x.$$  

- Approximating Markov Chain:

$$X^{n+1} = X^n + a(X^n)h + \sqrt{2\beta^{-1}}Z\sqrt{h},$$

- CG (reduced) dynamics: $\bar{x} \equiv Px \in \mathbb{R}^n$, $\bar{x} \equiv P^\perp x \in \mathbb{R}^m$

$$PX^{n+1} = PX^n + Pa(X^n)h + \sqrt{2\beta^{-1}}PZ\sqrt{h},$$
Multi-scale Diffusions and Stochastic Averaging

- Coarse-graining for diffusion processes on $\mathbb{R}^n \times \mathbb{R}^m$

\[
dX_t = a(X_t) \, dt + \sqrt{2\beta^{-1}} \, dW_t, \quad X_0 = x.
\]

- Approximating Markov Chain:

\[
X^{n+1} = X^n + a(X^n)h + \sqrt{2\beta^{-1}}Z\sqrt{h},
\]

- CG (reduced) dynamics: $\bar{x} \equiv Px \in \mathbb{R}^n$, $\bar{x} \equiv P^\perp x \in \mathbb{R}^m$

\[
P X^{n+1} = PX^n + P a(X^n)h + \sqrt{2\beta^{-1}}PZ\sqrt{h},
\]

- Markovian approximation:

\[
\bar{X}^{n+1} = \bar{X}^n + \bar{a}(\bar{X}^n; \theta)h + \sqrt{2\beta^{-1}}\bar{Z}\sqrt{h},
\]
Relative entropy rate: Approximating Markov chain

\[ p_h(x, x') dx' \sim e^{-\frac{\beta}{h} |x' - x - ha(x)|^2} dx', \quad \bar{p}_h(\bar{x}, \bar{x}'; \theta) \sim e^{-\frac{\beta}{h} |\bar{x}' - \bar{x} - \bar{a}(\bar{x};\theta)|^2} d\bar{x}' \]

\[ q_h(x, x' ; \theta) = \bar{p}_h( P x, P x' ; \theta) \nu(x'| P x') \]

\[ \mathcal{H}(P \mid P^\theta) = \int \int \mu(x) p_h(x, x') \log \frac{p_h(x, x')}{q_h(x, x'; \theta)} dx' dx . \]

\[ \min_\theta \mathcal{H}(P \mid P^\theta) \iff \min_\theta \int |\bar{a}(P x; \theta) - P a(x)|^2 \mu(x) dx \]

“Force-matching”
Example: Two-scale systems - Stochastic Averaging

\[ dX_t^\varepsilon = a(X^\varepsilon, Y^\varepsilon)dt + dW_t^1 \]
\[ dY_t^\varepsilon = \varepsilon^{-1} b(X^\varepsilon, Y^\varepsilon)dt + \varepsilon^{-1/2} dW_t^2 , \]

Theory: asymptotics \( \varepsilon \to 0 \) – averaging principle (Khasminskii, etc)

\[ d\bar{X}_t = \bar{a}(\bar{X}_t) + dW_t, \quad \bar{a}(x) = \lim_{\varepsilon \to 0} \int a(x, y) \mu_x^\varepsilon(dy) \]

Minimization of RER: \( \mu^\varepsilon(dx \, dy) = \bar{\mu}^\varepsilon(dx)\mu(dy|x) \) and for \( \varepsilon \ll 1 \)
\[ \mu^\varepsilon(dx \, dy) \approx \bar{\mu}(dx)\mu_x(dy) \]
Example: Two-scale systems - Stochastic Averaging

\[ dX_t^\varepsilon = a(X^\varepsilon, Y^\varepsilon) \, dt + dW_t^1 \]
\[ dY_t^\varepsilon = \varepsilon^{-1} b(X^\varepsilon, Y^\varepsilon) \, dt + \varepsilon^{-1/2} dW_t^2, \]

Theory: asymptotics \( \varepsilon \to 0 \) – averaging principle (Khasminskii, etc)

\[ d\bar{X}_t = \bar{a}(\bar{X}_t) + dW_t, \quad \bar{a}(x) = \lim_{\varepsilon \to 0} \int a(x, y) \mu_x^\varepsilon(dy) \]

Minimization of RER: \( \mu^\varepsilon(dx \, dy) = \bar{\mu}^\varepsilon(dx)\mu(dy|x) \) and for \( \varepsilon \ll 1 \)
\[ \mu^\varepsilon(dx \, dy) \approx \bar{\mu}(dx)\mu_x(dy) \]

\[ \min_{\bar{a}} \int |a(x, y) - \bar{a}(x)|^2 \mu_x(dy)\bar{\mu}(x) \, dx, \]

Unique minimizer as \( \varepsilon \to 0 \)

\[ \bar{a}(x) = \int a(x, y)\mu_x(dy). \]
Example:

\[ a(x, y) = -y - y^3 \]
\[ b(x, y) = x - y - y^3 \]
\[ \mu_x(dy) \sim e^{-\frac{1}{2}(y-x)^2 - \frac{1}{4}y^4} \]

\[ \epsilon \to 0 \]

\[ \bar{a}(x) = -x \]
Example: \( a(x, y) = -y - y^3, \ b(x, y) = x - y - y^3 \)

\[ \mu_x(dy) \sim e^{-\frac{1}{2}(y-x)^2 - \frac{1}{4}y^4} \]

\( \epsilon \to 0 \ a(x) = -x \)

**Figure:** Autocorrelation function of the CG stationary process \( \bar{X}_t^\epsilon \)
Driven Arrhenius diffusion

- Rates: Exchange dynamics with the migration rate to n.n. site $|x - y| = 1$

  \[ c(x, y, \sigma) = d e^{-\beta(U(x, \sigma))} [\sigma(x)(1 - \sigma(x + 1)) + \sigma(x)(1 - \sigma(x - 1))] \]

- Energy barrier: $U(x, \sigma) = \sum_{z \neq x} J(x - z)\sigma(z) - h$

  $J(z) = J_0$, for $|z| \leq L$ and $J = 0$ otherwise.

- Coarse-grained potential:

  \[ \bar{U}(k, \eta) = \sum_l \bar{J}(k, l)\eta(k) + \bar{J}(0, 0)(\eta(k) - 1) - \bar{h} \]

- Coarse-grained rates: assume *local equilibrium*, $\sigma(x) \approx q^{-1}\eta(k)$

  \[ \bar{c}(k, l, \eta) = \frac{1}{q} \eta(k)(q - \eta(l))d e^{-\beta \bar{U}(k, \eta)} \]

The generator $\tilde{\mathcal{L}}$:

\[ \tilde{\mathcal{L}}g(\eta) = \sum_{k, l} \bar{c}(k, l, \eta) [g(\eta + \delta_l - \delta_k) - g(\eta)] \]
Non-equilibrium stationary states
Bounded domain with a gradient in concentrations

\[ c(0) = 1 \]
\[ c(D) = 0 \]
Coverage profile $L=16, \beta=4$

Coverage

Error indicator for different levels of CG

$q = 2 \quad H = 0.099$
$q = 8 \quad H = -11.707$
$q = 16 \quad H = -42.236$
$q = 32 \quad H = -113.703$
Coverage profile at initial and optimal $\bar{J}_0$

- Optimized $\bar{J}_0 = 8.64 \, \text{q/L} = 0.25$
- Initial $\bar{J}_0 = 8.00 \, \text{q/L} = 0.25$
Examples    Multi-scale Diffusions and Stochastic Averaging

Coverage profile at initial and optimal $\beta$
lattice site

$q=32, L=32$
$q=8, L=32$
$q=16, L=32$
$q=4, L=32$

$\beta_J^0 = 4.00$
$q=32$ $\beta_J = 5.70$
$q=16$ $\beta_J = 4.53$
$q=8$ $\beta_J = 4.16$
$q=4$ $\beta_J = 4.04$
Path Space Relative Entropy

- Infers information about the path distribution: it contains information not only for the invariant measure but also for the dynamics.
Path Space Relative Entropy

- Infers information about the path distribution: it contains information not only for the invariant measure but also for the dynamics.

- No need for explicit knowledge of invariant measure. Thus, it is suitable for reaction networks and non-equilibrium steady state systems.
Path Space Relative Entropy

- Infers information about the path distribution: it contains information not only for the invariant measure but also for the dynamics.

- No need for explicit knowledge of invariant measure. Thus, it is suitable for reaction networks and non-equilibrium steady state systems.

- Relative entropy rate $\mathcal{H}$ is an observable $\Rightarrow$ tractable and statistical estimators can provide easily and efficiently its value using KMC solvers.
Path Space Relative Entropy

- Infers information about the path distribution: it contains information not only for the invariant measure but also for the dynamics.

- No need for explicit knowledge of invariant measure. Thus, it is suitable for reaction networks and non-equilibrium steady state systems.

- Relative entropy rate $\mathcal{H}$ is an observable $\Rightarrow$ tractable and statistical estimators can provide easily and efficiently its value using KMC solvers.

- Minimizing the error in $\mathcal{H}$ gives optimal parametrization similar to max-likelihood parameter estimation.
Path Space Relative Entropy

- Infers information about the path distribution: it contains information not only for the invariant measure but also for the dynamics.

- No need for explicit knowledge of invariant measure. Thus, it is suitable for reaction networks and non-equilibrium steady state systems.

- Relative entropy rate $\mathcal{H}$ is an observable $\Rightarrow$ tractable and statistical estimators can provide easily and efficiently its value using KMC solvers.

- Minimizing the error in $\mathcal{H}$ gives optimal parametrization similar to max-likelihood parameter estimation.

- Fisher information matrix allows for parameter identifiability in parameterization of dynamics [analogue to Cramer-Rao Theorems]