Path-space information metrics and variational inference for non-equilibrium coarse-grained systems.

Petr Plecháč

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

Coupled Mathematical Models for Physical and Biological Nanoscale Systems, Banff, Aug 28–Sep 1, 2016

Collaborators:

- M.A. Katsoulakis, (University of Massachusetts, Amherst)
- V. Harmandaris, E. Kalligiannaki (University of Crete)

Funding: Department of Energy, Applied Mathematics

References: http://www.math.udel.edu/~plechac

- E. Kalligiannaki, A. Chazirakis, M.A. Katsoulakis, P. Plecháč, V. Harmandaris:*Parametrizing coarse-grained models for molecular systems*, Eur. Phys. J. (2016)
- V. Harmandaris, E. Kalligiannaki, M.A. Katsoulakis, P. Plecháč: *Path-space variational inference for non-equilibrium coarse-grained systems*, J. Comp. Phys. 314 (2016) 355-383
- P. Dupuis, M.A. Katsoulakis, Y. Pantazis, P. Plecháč *Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics*, SIAM/ASA J. Uncertainty Quantification, 4 (2016), 80-111
- V. Harmandaris, E. Kalligiannaki, M.A. Katsoulakis, P. Plecháč: *The geometry of generalized force matching in coarse-graining and related information metrics* J. Chem. Phys. 143, 084105 (2015)

Outline

- Relative entropy, relative entropy rate, duality and bounds
- Coarse-graining & Parameterization
- Non-equilibrium steady states & Path Space Relative Entropy
- Coarse-graining & Parameterization – NESS
- Data driven coarse-graining
- Examples

Coarse-graining, model error and parametrization

Two probabilistic models P and Q on the common measurable space $(\Omega, \mathcal{B})$

Applications:

- reaction networks
- spatially heterogeneous chemical kinetics,
- molecular systems at equilibrium or
  with non-equilibrium steady states

Goal: extend empirical information theory techniques for path-space application

- Discrimination between the two models – distance

[1] D. Giannakis, A. J. Majda I. Horenko, *Physica D* (2012), A. J. Majda, B. Gershgorin, *Proc. Natl. Acad. Sci.* (2011)

Coarse-graining, model error and parametrization

Two probabilistic models P and Q on the common measurable space $(\Omega, \mathcal{B})$

Applications:

- reaction networks
- spatially heterogeneous chemical kinetics,
- molecular systems at equilibrium or
  with non-equilibrium steady states

Goal: extend empirical information theory techniques for path-space application

- Discrimination between the two models – distance
- Error for observables: $|\mathbb{E}_P[f] - \mathbb{E}_Q[f]|$

[1] D. Giannakis, A. J. Majda I. Horenko, *Physica D* (2012), A. J. Majda, B. Gershgorin, *Proc. Natl. Acad. Sci.* (2011)

Coarse-graining, model error and parametrization

Two probabilistic models P and Q on the common measurable space $(\Omega, \mathcal{B})$

Applications:

- reaction networks
- spatially heterogeneous chemical kinetics,
- molecular systems at equilibrium or
  with non-equilibrium steady states

Goal: extend empirical information theory techniques for path-space application

- Discrimination between the two models – distance
- Error for observables: $|\mathbb{E}_P[f] - \mathbb{E}_Q[f]|$
- Stability under perturbations $P^\theta \to Q \equiv P^{\theta+\epsilon}$

[1] D. Giannakis, A. J. Majda I. Horenko, *Physica D* (2012), A. J. Majda, B. Gershgorin, *Proc. Natl. Acad. Sci.* (2011)

Coarse-graining, model error and parametrization

Two probabilistic models P and Q on the common measurable space $(\Omega, \mathcal{B})$

Applications:

- reaction networks
- spatially heterogeneous chemical kinetics,
- molecular systems at equilibrium or
  with non-equilibrium steady states

Goal: extend empirical information theory techniques for path-space application

- Discrimination between the two models – distance
- Error for observables: $|\mathbb{E}_P[f] - \mathbb{E}_Q[f]|$
- Stability under perturbations $P^\theta \to Q \equiv P^{\theta+\epsilon}$
- Parameter identifiability in parameterized models P^θ

[1] D. Giannakis, A. J. Majda I. Horenko, *Physica D* (2012), A. J. Majda, B. Gershgorin, *Proc. Natl. Acad. Sci.* (2011)

Coarse-graining, model error and parametrization

Two probabilistic models P and Q on the common measurable space $(\Omega, \mathcal{B})$

Applications:

- reaction networks
- spatially heterogeneous chemical kinetics,
- molecular systems at equilibrium or
  with non-equilibrium steady states

Goal: extend empirical information theory techniques for path-space application

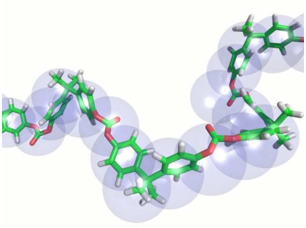
- Discrimination between the two models – distance
- Error for observables: $|\mathbb{E}_P[f] - \mathbb{E}_Q[f]|$
- Stability under perturbations $P^\theta \to Q \equiv P^{\theta+\epsilon}$
- Parameter identifiability in parameterized models P^θ
- "Best-fit" for coarse-grained models.

[1] D. Giannakis, A. J. Majda I. Horenko, *Physica D* (2012), A. J. Majda, B. Gershgorin, *Proc. Natl. Acad. Sci.* (2011)

Coarse-Graining – Reduced molecular models

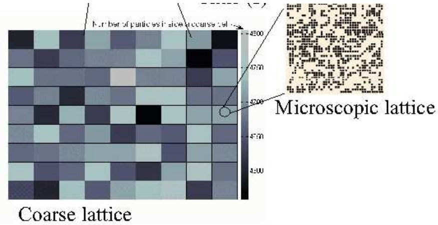
1. Coarse-graining of polymers; DPD

Briels, et. al. *J.Chem.Phys.* '01;
Müller-Plathe *Chem.Phys.Chem* '02;
Laaksonen et. al. *Soft Matter* '03;
Kremer et al. *Macromolecules '06*
Deserno et. al. *Nature* '07;
Espanol *J Chem. Phys. '07, '11*;
Shell *J. Chem. Phys. '12*;
Noid *J Chem. Phys. '13*



2. Stochastic lattice dynamics

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*



Coarse-grained models and variational inference

- coarse-graining map; $\mathbf{T} : \Sigma \to \bar{\Sigma}$, $P \mapsto \bar{P} = \mathbf{T}_* P$
- exactly coarse-grained model $\bar{P}$ intractable
- $\bar{P}$ approximated from a tractable, parametrized family $\bar{Q}^\theta$
- minimize a "distance" between $\bar{P}$ and $\bar{Q}^\theta$

$$\min_{\theta \in \Theta} \mathrm{dist}(\bar{P}, \bar{Q}^\theta)$$

[1] $P = \mu$ equilibrium distrib.; Shell (2008,2011), Noid, (2011,2013), Bilionis, Koustsourelakis (2012), Zabaras et al. (2013)

Coarse-grained models and variational inference

- coarse-graining map; $\mathbf{T} : \Sigma \to \bar{\Sigma},\ P \mapsto \bar{P} = \mathbf{T}_* P$
- exactly coarse-grained model $\bar{P}$ intractable
- $\bar{P}$ approximated from a tractable, parametrized family $\bar{Q}^\theta$
- minimize a "distance" between $\bar{P}$ and $\bar{Q}^\theta$

$$\min_{\theta\in\Theta} \mathrm{dist}(\bar{P}, \bar{Q}^\theta)$$

- Choice of $\mathrm{dist}(\cdot,\cdot)$?
  - Model based (information theoretic): dist = relative entropy

$$\min_{\theta\in\Theta} \mathcal{R}\left(P \,|\, \mathbf{T}_*^\dagger \bar{Q}^\theta\right) \quad \text{or} \quad \min_{\theta\in\Theta} \mathcal{R}\left(\mathbf{T}_* P \,|\, \bar{Q}^\theta\right)$$

  - Observable based (QoI): $\{\phi_1, \phi_2, \ldots, \phi_n\}$

$$\min_{\theta\in\Theta} \sum_i \left|\mathbb{E}_P[\phi_i] - \mathbb{E}_{\bar{Q}^\theta}[\phi_i]\right|^2$$

Example: radial distribution function, forces $f = -\nabla U$

[1] $P = \mu$ equilibrium distrib.; Shell (2008,2011), Noid, (2011,2013), Bilionis, Koustsourelakis (2012), Zabaras et al. (2013)

Why information-based methods ?

Relative Entropy and $\mathcal{R}$-projections

- Pseudo-distance (Kullback-Leibler divergence)

$$\mathcal{R}\left(P \mid Q\right) = \int \log\left(\frac{dP}{dQ}\right)\ dP$$

for $P \ll R,\ Q \ll R$ $\quad \mathcal{R}\left(P \mid Q\right) = \int p_R \log\left(\frac{p_R}{q_R}\right)\ dR$

Why information-based methods ?

Relative Entropy and $\mathcal{R}$-projections

- Pseudo-distance (Kullback-Leibler divergence)

$$\mathcal{R}\left(P \mid Q\right) = \int \log\left(\frac{dP}{dQ}\right)\, dP$$

for $P \ll R$, $Q \ll R$ $\quad \mathcal{R}\left(P \mid Q\right) = \int p_R \log\left(\frac{p_R}{q_R}\right)\, dR$

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

Why information-based methods ?

Relative Entropy and $\mathcal{R}$-projections

- Pseudo-distance (Kullback-Leibler divergence)

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

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

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

- $\mathcal{R}$-geometry of probability distributions $\mathcal{B}(R,\rho) = \{P \,|\, \mathcal{R}\left(P \,|\, R\right) < \rho\}$
  $\mathcal{R}$-projection on $\mathcal{A}$ convex, TV closed, $\mathcal{A} \cap \mathcal{B}(R,\rho) \neq \emptyset$ (Kullback, Csiszár)

$$\mathcal{R}\left(Q \,|\, R\right) = \min_{P \in \mathcal{A}} \mathcal{R}\left(P \,|\, R\right)$$

Why information-based methods ?

Relative Entropy and $\mathcal{R}$-projections

- Pseudo-distance (Kullback-Leibler divergence)

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

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

- 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 $\mathcal{B}(R,\rho) = \{P\,|\,\mathcal{R}(P\,|\,R) < \rho\}$
  $\mathcal{R}$-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, $\rho = \mathcal{R}(Q\,|\,R)$

$$P \text{ s.t. } \int \log\frac{dQ}{dR}\, dP = \rho\,, \quad \mathcal{R}(P\,|\,R) = \mathcal{R}(P\,|\,Q) + \mathcal{R}(Q\,|\,R)$$

Why information-based methods ?

Relative Entropy and $\mathcal{R}$-projections

- Pseudo-distance (Kullback-Leibler divergence)

$$\mathcal{R}(P \mid Q) = \int \log\left(\frac{dP}{dQ}\right) \, dP$$

for $P \ll R$, $Q \ll R$ $\quad \mathcal{R}(P \mid Q) = \int p_R \log\left(\frac{p_R}{q_R}\right) \, dR$

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

- The "best fit" in relative entropy: $\min_{Q \in \mathcal{A}} \mathcal{R}(P \mid Q)$
  modeling error + numerical error + statistical error
  Modelling error$\sim \mathcal{R}(P \mid Q) \sim \epsilon^\alpha$
  Bounds on the weak error:

$$|\mathbb{E}_P[f] - \mathbb{E}_Q[f]| \leq C_f \Phi(\mathcal{R}(P \mid Q))$$

Bounding the error

modeling error

Bounding the error

modeling error + numerical error + statistical error

Bounding the error

modeling error + numerical error + statistical error

- Csiszár ϕ-divergences, convex $\phi : \mathbb{R}^+ \to \mathbb{R}$ with $\phi(1) = 0$

$$\mathcal{R}_\phi(Q \,|\, P) = \begin{cases} \int \phi\left(\frac{dQ}{dP}(\omega)\right)\ P(d\omega)\,, & \text{if } Q \ll P \text{ and } \phi\left(\frac{dQ}{dP}\right) \text{ is } P\text{-integrable} \\ +\infty & \text{otherwise,} \end{cases}$$

  - $\phi(x) = x \log x$ – relative entropy (Kullback-Leibler divergence)
  - $\phi(x) = (x-1)^2$ – χ^2-divergence

Bounding the error

modeling error + numerical error + statistical error

- Csiszár ϕ-divergences, convex $\phi : \mathbb{R}^+ \to \mathbb{R}$ with $\phi(1) = 0$

$$\mathcal{R}_\phi(Q \mid P) = \begin{cases} \int \phi\left(\frac{dQ}{dP}(\omega)\right) \, P(d\omega)\,, & \text{if } Q \ll P \text{ and } \phi\left(\frac{dQ}{dP}\right) \text{ is } P\text{-integrable} \\ +\infty & \text{otherwise,} \end{cases}$$

  - $\phi(x) = x \log x$ – relative entropy (Kullback-Leibler divergence)
  - $\phi(x) = (x-1)^2$ – χ^2-divergence
- Csiszár-Kullback-Pinsker inequality: $\|P - Q\|_{\mathrm{TV}} \le \sqrt{2\mathcal{R}\left(P \mid Q\right)}$

$$|\mathbb{E}_P[f] - \mathbb{E}_Q[f]| \le \|f\|_\infty \sqrt{2\mathcal{R}\left(P \mid Q\right)}$$

Bounding the error

modeling error + numerical error + statistical error

- Csiszár ϕ-divergences, convex $\phi : \mathbb{R}^+ \to \mathbb{R}$ with $\phi(1) = 0$

$$\mathcal{R}_\phi(Q \,|\, P) = \begin{cases} \int \phi\left(\frac{dQ}{dP}(\omega)\right) \, P(d\omega)\,, & \text{if } Q \ll P \text{ and } \phi\left(\frac{dQ}{dP}\right) \text{ is } P\text{-integrable} \\ +\infty & \text{otherwise,} \end{cases}$$

  - $\phi(x) = x \log x$ – relative entropy (Kullback-Leibler divergence)
  - $\phi(x) = (x-1)^2$ – χ^2-divergence
- Csiszár-Kullback-Pinsker inequality: $\|P - Q\|_{\mathrm{TV}} \leq \sqrt{2\mathcal{R}\,(P \,|\, Q)}$

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

- $\mathcal{R}\,(P \,|\, Q) \leq \chi^2\,(P \,|\, Q)$

CG: Error Quantification and Parameterization using RE in molecular simulations:

Katsoulakis, P.P. Sopasakis (2006), M.S. Shell (2008), Katsoulakis, P.P., Rey-Bellet, Tsagkarogiannis (07, 08, 09), M.S. Shell (08,12), Bilionis et al (12), Zabaras et al (13), M. Katsoulakis, P.P. (2013) (dynamics, non-equilibrium), Luskin,Simpson,Srolowitz (2015)

Variational error bounds

Error estimate of the type:

$$|\mathbb{E}_P[f] - \mathbb{E}_Q[f]| \le C_f \Phi(\mathcal{R}(P \mid Q))$$

Variational representation of $\mathcal{R}(P \mid Q)$ and log-moment generating function

$$\Lambda_{P,f}(c) \equiv \frac{1}{c} \log \mathbb{E}_P[e^{cf}] = \sup_{Q \in \mathcal{P}(\Omega)} \left\{ \mathbb{E}_Q[f] - \frac{1}{c}\mathcal{R}(Q \mid P) \right\}$$

For $f - \mathbb{E}_P[f]$ tight variational bounds

$$\sup_{c>0} \left\{ -\frac{1}{c}\tilde{\Lambda}_{P,f}(-c) - \frac{1}{c}\mathcal{R}(Q \mid P) \right\} \le \mathbb{E}_Q[f] - \mathbb{E}_P[f] \le$$

$$\le \inf_{c>0} \left\{ \frac{1}{c}\tilde{\Lambda}_{P,f}(c) + \frac{1}{c}\mathcal{R}(Q \mid P) \right\}$$

[1] P. Dupuis, M. Katsoulakis, Y. Pantazis, P.P. SIAM JUQ (2016)

Variational error bounds

Error estimate of the type:

$$|\mathbb{E}_P[f] - \mathbb{E}_Q[f]| \leq C_f \Phi(\mathcal{R}(P \mid Q))$$

Variational representation of $\mathcal{R}(P \mid Q)$ and log-moment generating function

$$\Lambda_{P,f}(c) \equiv \frac{1}{c} \log \mathbb{E}_P[e^{cf}] = \sup_{Q \in \mathcal{P}(\Omega)} \left\{ \mathbb{E}_Q[f] - \frac{1}{c}\mathcal{R}(Q \mid P) \right\}$$

For $f - \mathbb{E}_P[f]$ tight variational bounds

$$\sup_{c>0} \left\{ -\frac{1}{c}\tilde{\Lambda}_{P,f}(-c) - \frac{1}{c}\mathcal{R}(Q \mid P) \right\} \leq \mathbb{E}_Q[f] - \mathbb{E}_P[f] \leq$$

$$\leq \inf_{c>0} \left\{ \frac{1}{c}\tilde{\Lambda}_{P,f}(c) + \frac{1}{c}\mathcal{R}(Q \mid P) \right\}$$

$$\inf_{c>0} \left\{ \tfrac{1}{c}\tilde{\Lambda}_{P,f}(c) + \tfrac{1}{c}\rho^2 \right\} \equiv (\tilde{\Lambda}^*_{P,f})^{-1}(\rho^2)$$

unique minimizer $\quad c^*(\rho) = c_1^* \rho + \mathcal{O}(\rho^2)\,, \quad c_1^* = \sqrt{\dfrac{2}{\mathrm{Var}_P[f]}}\,.$

[1] P. Dupuis, M. Katsoulakis, Y. Pantazis, P.P. SIAM JUQ (2016)

Variational error bounds

Error estimate:

$$|\mathbb{E}_P[f] - \mathbb{E}_Q[f]| \le (\tilde{\Lambda}^*_{P,f})^{-1}(\mathcal{R}\,(Q \mid P))$$

Asymptotics using $c^*(\rho) = \sqrt{\frac{2}{\mathrm{Var}_P[f]}\rho} + \mathcal{O}(\rho^2)$

$$|\mathbb{E}_Q[f] - \mathbb{E}_P[f]| \le \sqrt{\mathrm{Var}_P[f]}\sqrt{2\mathcal{R}\,(Q \mid P)} + \mathcal{O}(\mathcal{R}\,(Q \mid P))\,,$$

Stability/Sensitivity estimate: $Q = P^{\theta+\epsilon}$, $P = P^{\theta}$

$$|\mathbb{E}_{P^{\theta+\epsilon}}[f] - \mathbb{E}_{P^{\theta}}[f]| \le \sqrt{\mathrm{Var}_{P^{\theta}}[f]}\sqrt{\mathbf{F}(P^{\theta})}\,\epsilon + \mathcal{O}(|\epsilon|^2)$$

Fisher Information Matrix: $\mathbf{F}(P^{\theta})_{ij} = \int \frac{\partial p^{\theta}_R}{\partial \theta_i}\frac{\partial p^{\theta}_R}{\partial \theta_j} p^{\theta}_R \, dR$

[1] P. Dupuis, M. Katsoulakis, Y. Pantazis, P.P. SIAM JUQ (2016)

Path-space error estimates

- Measurable functional $\mathcal{F}$ of the process $\{X_t\}_{t\geq 0}$

$$|\mathbb{E}_{Q_{[0,T]}}[\mathcal{F}] - \mathbb{E}_{P_{[0,T]}}[\mathcal{F}]| \leq \sqrt{\frac{1}{T}\mathrm{Var}_{P_{[0,T]}}[T\mathcal{F}]}\sqrt{\frac{2}{T}\mathcal{R}\left(Q_{[0,T]} \mid P_{[0,T]}\right)} + \mathcal{O}\left(\frac{1}{T}\mathcal{R}\left(Q_{[0,T]} \mid P_{[0,T]}\right)\right)$$

Path-space error estimates

- Measurable functional $\mathcal{F}$ of the process $\{X_t\}_{t\geq 0}$

$$|\mathbb{E}_{Q_{[0,T]}}[\mathcal{F}] - \mathbb{E}_{P_{[0,T]}}[\mathcal{F}]| \leq \sqrt{\frac{1}{T}\mathrm{Var}_{P_{[0,T]}}[T\mathcal{F}]}\sqrt{\frac{2}{T}\mathcal{R}\left(Q_{[0,T]}\,|\,P_{[0,T]}\right)} + \mathcal{O}\left(\frac{1}{T}\mathcal{R}\left(Q_{[0,T]}\,|\,P_{[0,T]}\right)\right)$$

- Stationary process $(T \to \infty)$

$$\frac{1}{T}\mathcal{R}\left(Q_{[0,T]}\,|\,P_{[0,T]}\right) = \mathcal{H}(q\,|\,p) + \frac{1}{T}\mathcal{R}\left(\mu\,|\,\nu\right)$$

$$|\mathbb{E}_{Q_{[0,T]}}[\mathcal{F}] - \mathbb{E}_{P_{[0,T]}}[\mathcal{F}]| \leq \sqrt{\frac{1}{T}\mathrm{Var}_{P_{[0,T]}}[T\mathcal{F}]}\sqrt{2\mathcal{H}(q\,|\,p) + \frac{2}{T}\mathcal{R}\left(\mu\,|\,\nu\right)} + \text{h.o.t.}$$

- Ergodic-type observables: $\mathcal{F}(X) = \frac{1}{T}\int_0^T f(X_t)\,dt$

$$|\mathbb{E}_{Q_{[0,T]}}[\mathcal{F}] - \mathbb{E}_{P_{[0,T]}}[\mathcal{F}]| \le \sqrt{\tau_T(f)}\sqrt{2\mathcal{H}(q\,|\,p) + \frac{2}{T}\mathcal{R}(\mu\,|\,\nu)} + \text{h.o.t.}$$

Integrated Autocorrelation Time (IAT)

$$\tau(f) = \lim_{T\to\infty}\tau_T(f) = \mathrm{Var}_\mu^\theta[f] + 2\sum_{k=1}^{\infty} A_f(k)$$

$$A_f(k) = \mathbb{E}_{P^\theta_{[0,T]}}[(f(X_0) - \mathbb{E}_{\mu^\theta}[f(X_0)])(f(X_k) - \mathbb{E}_{\mu^\theta}[f(X_0)])]$$

- as $T \to \infty$
perturbations $\epsilon = |\epsilon| e$ of the invariant measure μ^θ

$$\frac{1}{|\epsilon|}|\mathbb{E}_{\mu^{\theta+\epsilon e}}[f] - \mathbb{E}_{\mu^\theta}[f]| \le \sqrt{\tau(f)}\sqrt{e^T \mathbf{F}_{\mathcal{H}}(p^\theta) e} + \mathcal{O}(\epsilon)$$

Parametrized CG – Equilibrium

- Invariant distribution $\mu \sim e^{-\beta H(\sigma)}$
- Detailed balance for the coarse-grained process w.r.t. $\bar{\mu} \sim e^{-\beta \bar{H}(\eta)}$
- Coarse-grained Hamiltonian $\bar{H}(\eta)$
- Parametrized approximation

$$\min_\theta \mathcal{R}\left(\mu \,|\, \mu^{\text{app}}(\theta)\right) \quad \text{or} \quad \min_\theta \mathcal{R}\left(\mu^{\text{app}}(\theta) \,|\, \mu\right)$$

Gibbs structure allows explicit calculations of $\mathcal{R}$

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

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

- Solution using typically gradient methods, Newton-Raphson, etc: [1]
- Is the parametric family $\bar{H}^{(0)}(\theta)$ rich enough ? [2]
  Multi-body Hamiltonians, Cluster expansions etc

[1] M.S. Shell (2008, 2012), Bilionis et al (2012), Zabaras et al (2013), Noid (2013)
[2] Katsoulakis, P.P., Rey-Bellet (2008)

Coarse-graining non-equilibrium systems

- focus on systems with steady states
- Non-equilibrium steady state (NESS) $\neq$ const. $e^{-\beta H(\sigma)}$
  - systems driven by external fields, boundary conditions etc.
  - reaction networks
  - polymer flows
  - heterogeneous reaction systems with multiple mechanisms: reaction-diffusion-adsorption-desorption
- Find the "best-fit" coarse-grained Markovian approximation
  possibly non-Markovian – estimate memory kernels in Mori-Zwanzig formalism.

Non-equilibrium steady states

Example: Continuous-time jump process (KMC)

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

Stationary states: $\partial_t P = 0 \implies \sum_{\sigma'} \mathbf{j}_s(\sigma', \sigma) = 0$
Current $\sigma' \to \sigma$: $\mathbf{j}_s(\sigma', \sigma) = c(\sigma', \sigma)\mu(\sigma') - c(\sigma, \sigma')\mu(\sigma)$

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-equlibrium steady states

$$\sum_{\sigma'} \mathbf{j}_s(\sigma', \sigma) = \sum_{\sigma'} (c(\sigma', \sigma)\mu(\sigma') - c(\sigma, \sigma')\mu(\sigma)) = 0$$

irreversible rate loops, i.e., a non-zero current at stationary states.

Relative entropy on the path space

- Markov chains on Σ:
  $\{\sigma_n\}_{n\in\mathbb{Z}^+}$, $P^\theta(\sigma, d\sigma)$, $\mu^\theta(\sigma)$
- Approximating Markov chain $\{\tilde{\sigma}_n\}_{n\in\mathbb{Z}^+}$, $\tilde{P}^\theta(\sigma, d\sigma)$, $\tilde{\mu}^\theta(\sigma)$
- 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

$$\mathcal{R}\left(Q^\theta \mid \tilde{Q}^\theta\right) = \int_{\Sigma^M} \mu^\theta(\sigma_0) \prod_{i=0}^{M-1} p^\theta(\sigma_i, \sigma_{i+1}) \log \frac{\mu^\theta(\sigma_0) \prod_{i=0}^{M-1} p^\theta(\sigma_i, \sigma_{i+1})}{\tilde{\mu}^\theta(\sigma_0) \prod_{i=0}^{M-1} \tilde{p}^\theta(\sigma_i, \sigma_{i+1})} \, d\sigma$$

Relative entropy decomposition and scaling

$$\mathcal{R}\left(Q^\theta \mid \tilde{Q}^\theta\right) = \int_\Sigma \cdots \int_\Sigma \mu^\theta(\sigma_0) \prod_{i=0}^{M-1} p^\theta(\sigma_i, \sigma_{i+1}) \left(\log \frac{\mu^\theta(\sigma_0)}{\tilde{\mu}^\theta(\sigma_0)} + \sum_{i=0}^{i=M-1} \log \frac{p^\theta(\sigma_i, \sigma_{i+1})}{\tilde{p}^\theta(\sigma_i, \sigma_{i+1})}\right) \, d\sigma_0 \dots d\sigma_M$$

Relative entropy decomposition and scaling

$$\mathcal{R}\left(Q^\theta \mid \tilde{Q}^\theta\right) = \int_\Sigma \cdots \int_\Sigma \mu^\theta(\sigma_0) \prod_{i=0}^{M-1} p^\theta(\sigma_i, \sigma_{i+1}) \left(\log \frac{\mu^\theta(\sigma_0)}{\tilde{\mu}^\theta(\sigma_0)} + \sum_{i=0}^{i=M-1} \log \frac{p^\theta(\sigma_i, \sigma_{i+1})}{\tilde{p}^\theta(\sigma_i, \sigma_{i+1})} \right) \, d\sigma_0 \dots d\sigma_M$$

Using

$$\int_\Sigma p(\sigma, \sigma') \, d\sigma' = 1 \,, \quad \int_\Sigma \mu(\sigma) p(\sigma, \sigma') \, d\sigma = \mu(\sigma')$$

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

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

Relative entropy decomposition and scaling

$$\mathcal{R}\left(Q^\theta \mid \tilde{Q}^\theta\right) = \int_\Sigma \cdots \int_\Sigma \mu^\theta(\sigma_0) \prod_{i=0}^{M-1} p^\theta(\sigma_i, \sigma_{i+1}) \left(\log \frac{\mu^\theta(\sigma_0)}{\tilde{\mu}^\theta(\sigma_0)} + \sum_{i=0}^{i=M-1} \log \frac{p^\theta(\sigma_i, \sigma_{i+1})}{\tilde{p}^\theta(\sigma_i, \sigma_{i+1})} \right) d\sigma_0 \dots d\sigma_M$$

Using

$$\int_\Sigma p(\sigma, \sigma')\, d\sigma' = 1\,, \quad \int_\Sigma \mu(\sigma) p(\sigma, \sigma')\, d\sigma = \mu(\sigma')$$

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

$$= M \mathbb{E}^\theta_\mu \left[\int_\Sigma p^\theta(\sigma, \sigma') \log \frac{p^\theta(\sigma, \sigma')}{\tilde{p}^\theta(\sigma, \sigma')}\, d\sigma' \right] + \mathcal{R}\left(\mu^\theta \mid \tilde{\mu}^\theta\right)$$

$$\mathcal{R}\left(Q^\theta \mid \tilde{Q}^\theta\right) = M\mathcal{H}(Q^\theta \mid \tilde{Q}^\theta) + \mathcal{R}\left(\mu^\theta \mid \tilde{\mu}^\theta\right)$$

Continuous time Markov chain

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

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

The initial distribution is the stationary measure μ (resp. $\tilde{\mu}$).
Radon-Nikodym derivative:

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

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

Langevin dynamics

- Microscopic dynamics with forcefield $x \in \mathbb{R}^n \mapsto b(x) \in \mathbb{R}^n$, rank $\sigma(x) \le n$

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t\,, \quad t \ge 0\,, \quad X_0 \sim \mu_0\,.$$

- Approximating dynamics

$$d\tilde{X}_t = \tilde{b}(\tilde{X}_t;\theta)dt + \sigma(\tilde{X}_t)dB_t\,, \quad t \ge 0\,, \quad \tilde{X}_0 \sim \nu_0\,,$$

- Radon-Nikodym derivative

$$\frac{dP_X^{[0,T]}}{dP_{\tilde{X}}^{[0,T]}}(X_t,t) = \frac{d\mu_0}{d\nu_0}(X_0)\exp\left\{-\int_0^t \langle u(X_s;\theta), dB_s\rangle - \frac{1}{2}\int_0^t |u(X_s;\theta)|^2\,ds\right\}.$$

$\sigma(X_s)u(X_s;\theta) = b(X_s) - \tilde{b}(X_s;\theta)$

- Relative entropy

$$\mathcal{R}\left(P_X^T \mid P_{\tilde{X}}^T\right) = \mathbb{E}_{P_X^T}\left[\int_0^T \frac{1}{2}|u(X_s;\theta|^2\,ds\right] + \mathcal{R}\left(\mu_0 \mid \nu_0\right)$$

Relative entropy rate

For stationary process: $\mathcal{R}\left(P_X^T \,|\, P_{\tilde{X}}^T\right) = T\mathcal{H}(P_X^T \,|\, P_{\tilde{X}}^T) + \mathcal{R}\left(\mu_0 \,|\, \nu_0\right)$

$$\mathcal{H}(P_X^T \,|\, P_{\tilde{X}}^T) = \lim_{T\to\infty} \frac{1}{T}\mathcal{R}\left(P_X^T \,|\, P_{\tilde{X}}^T\right)$$

- RER representation for stationary process

$$\mathcal{H}(P_X \,|\, P_{\tilde{X}}) = \mathbb{E}_\mu\left[\frac{1}{2}\|b(X) - \tilde{b}(X;\theta)\|_\Xi^2\right] ,$$

the norm $\|\cdot\|_\Xi$ defined by $\Xi(x) = \left[\sigma^T(x)\sigma(x)\right]^{-1}\sigma^T(x)$.

- RE representation for finite time

$$\mathcal{R}\left(P_X^T \,|\, P_{\tilde{X}}^T\right) = \mathcal{H}^T(P_X^T \,|\, P_{\tilde{X}}^T) + \mathcal{R}\left(\mu_0 \,|\, \nu_0\right) ,$$

where

$$\mathcal{H}^T(P_X^T \,|\, P_{\tilde{X}}^T) = \mathbb{E}_{P_X^T}\left[\frac{1}{2}\int_0^T \|b(X_s) - \tilde{b}(X_s;\theta)\|_\Xi^2 \, ds\right] .$$

Relative Entropy Rate and Dynamics Parametrization

Langevin dynamics

- CG map: $\mathbb{R}^n \to \mathbb{R}^m$, $x \mapsto \bar{x} = \mathbf{T}x$
- CG Markovian dynamics

$$d\bar{X}_t = \bar{b}(\bar{X}_t;\theta)\,dt + \bar{\sigma}(\bar{X}_t;\theta)\,d\bar{B}_t\,, \quad \bar{X}_0 \sim \bar{\mu}_0$$

- Reconstructed process $\mathbf{T}\tilde{X}_t = \bar{X}_t$ in distribution

$$d\tilde{X}_t = \tilde{b}(\tilde{X}_t;\theta)\,dt + \tilde{\sigma}(\tilde{X}_t;\theta)\,dB_t\,, \quad \tilde{X}_0 \sim \nu_0$$

$$\mathbf{T}\tilde{b}(x;\theta) = \bar{b}(\mathbf{T}x;\theta)\,, \quad \mathbf{T}\tilde{\Sigma}(x;\theta)\mathbf{T}^T = \bar{\Sigma}(\mathbf{T}x;\theta) \;\text{ for all } x \in \mathbb{R}^n$$
$$\tilde{\Sigma}(x;\theta) = \tilde{\sigma}(x;\theta)\tilde{\sigma}^T(x;\theta)\,, \quad \bar{\Sigma}(\bar{x};\theta) = \bar{\sigma}(\bar{x};\theta)\bar{\sigma}^T(\bar{x};\theta)$$

- Best fit at stationary regime – minimize $\mathcal{H}$

$$\min_{\theta\in\Theta} \mathbb{E}_\mu \left[\frac{1}{2}\|\mathbf{T}b(X) - \bar{b}(\mathbf{T}X;\theta)\|^2_{\mathbf{T}^\sharp\Xi}\right]$$

new metric $\|\bar{b}\|_{\mathbf{T}^\sharp\Xi} \equiv \bar{b}^T\mathbf{T}^{\sharp,T}\Xi^T\Xi\mathbf{T}^\sharp\bar{b}$, $\quad \mathbf{T}^\sharp \equiv \mathbf{T}^T(\mathbf{T}\mathbf{T}^T)^{-1}$.

[1] Kalligianaki, Harmandaris, Katsoulakis, P.P. (2016)

Relative Entropy Rate and Dynamics Parametrization

Lattice KMC dynamics

- Define parametrized CG transition probabilities $q^{\theta^*}(\sigma,\sigma')$:
  - Parametrized CG transition probabilities $\bar{p}^{\theta}(\eta,\eta')$
  - Reconstruction scheme: $\nu(\sigma'|\mathbf{T}\sigma')$, e.g. uniform: $\frac{1}{|\{\sigma:\mathbf{T}\sigma=\eta'\}|}$
  - $q^{\theta}(\sigma,\sigma') = \nu(\sigma'|\mathbf{T}\sigma')\bar{p}^{\theta}(\mathbf{T}\sigma,\mathbf{T}\sigma')$,
- $\mathcal{R}\left(P \mid Q^{\theta}\right) =$ Loss of Information (in time-series) due to CG
- For long times $M \gg 1$, RER is dominant:

$$\mathcal{R}\left(P \mid Q^{\theta}\right) = M\mathcal{H}(P \mid Q^{\theta}) + \mathcal{R}\left(\mu \mid \mu^{\theta}\right)$$

$$\mathcal{H}(P \mid Q^{\theta}) = \sum_{\sigma\in\Sigma}\mu(\sigma)\sum_{\sigma'\in\Sigma}p(\sigma,\sigma')\log\frac{p(\sigma,\sigma')}{q^{\theta}(\sigma,\sigma')}].$$

- No need for explicit knowledge of NESS: suitable for reaction networks, driven systems, reaction-diffusion, etc.

Path-space Fisher Information Matrix (FIM)

Under a smoothness assumption wrt θ, (checkable, on the rates only!)

$$\mathcal{H}\left(Q_{0,M}^{\theta} \mid Q_{0,M}^{\theta+\epsilon}\right) = \frac{1}{2}\epsilon^T \mathbf{F}_{\mathcal{H}}\left(Q_{0,M}^{\theta}\right)\epsilon + O(|\epsilon|^3)$$

Path-space Fisher Information Matrix.
Example (Markov chain):

$$\mathbf{F}_{\mathcal{H}}\left(Q_{0,M}^{\theta}\right) = \mathbb{E}_{\mu^\theta}\left[\int_E p^\theta(\sigma,\sigma')\nabla_\theta \log p^\theta(\sigma,\sigma')\nabla_\theta \log p^\theta(\sigma,\sigma')^T \, d\,\sigma'\right]$$

- Spectral analysis of FIM gives the most/least sensitive directions.
- Derivative-free sensitivity analysis method.
- Characterizes robustness under parameter perturbations.
- Determines parameter identifiability, [e.g. Cramer-Rao Theorems].
- Sparse structure of the path FIM.

Cramer-Rao inequalities for time-series

- a biased estimator $\hat{\theta} = f(X)$ of the parameter θ with bias function $\psi(\theta)$
  $\mathbb{E}_{P^\theta}[f] = \psi(\theta)$
  Cramer-Rao inequality

$$\mathrm{Var}_{P^\theta}(\hat{\theta}) \geq \frac{[\psi'(\theta)]^2}{\mathbf{F}(P^\theta)}$$

- a new Cramer-Rao type inequality for time series stationary statistics
  $\hat{\theta} = \mathcal{F}_T(X)$ with the path-space observables such as $\mathcal{F}_T(X) = \frac{1}{T}\sum_i f(X_i)$

$$\tau_{P^\theta}(f) \geq \frac{[\psi'(\theta)]^2}{\mathbf{F}_{\mathcal{H}}(P^\theta)}$$

where $\psi(\theta) = \mathbb{E}_{P^\theta_{[0,T]}}[\mathcal{F}_T]$ is the bias of the estimator.

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:

$$G^n = \mathrm{Hess}(\mathcal{H}(P \mid Q^{\theta^n}))^{-1} \nabla_\theta \mathcal{H}(P \mid Q^{\theta^n})\,.$$

$$\mathbf{F}_{\mathcal{H}}(Q^\theta) = \mathrm{Hess}(\mathcal{H}(P \mid Q^\theta)) = -\mathbb{E}_\mu \left[\sum_{\sigma'} p(\sigma,\sigma) \nabla_\theta^2 \log q^\theta(\sigma,\sigma') \right]\,.$$

Relative Entropy Rate (RER) $\mathcal{H}$

$$\mathcal{R}\left(Q^\theta \mid \tilde{Q}^\theta\right) = M\mathcal{H}(Q^\theta \mid \tilde{Q}^\theta) + \mathcal{R}\left(\mu^\theta \mid \tilde{\mu}^\theta\right)$$

$$\mathcal{H}(Q^\theta \mid \tilde{Q}^\theta) = \mathbb{E}_\mu^\theta\left[\int_\Sigma p^\theta(\sigma,\sigma')\log\frac{p^\theta(\sigma,\sigma')}{\tilde{p}^\theta(\sigma,\sigma')}\,d\sigma'\right]$$

- RER is an observable $\Rightarrow$ tractable and statistical estimators are available. [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**

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

Estimator I:

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

Estimator II:

$$\widehat{\mathcal{H}}_2^{(n)} = \frac{1}{n}\sum_{k=0}^{n-1}\log\frac{c(\sigma_k,\sigma_{k+1})}{\widetilde{c}(\sigma_k,\sigma_{k+1})} - \frac{1}{T}\sum_{k=0}^{n-1}\delta\tau_k(\lambda(\sigma_k) - \widetilde{\lambda}(\sigma_k))$$

[1] Pantazis, Katsoulakis J. Chem. Phys. (2013)

Data-based parametrization of CG dynamics

- Unbiased estimator for RER,

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

- Minimization of RER:

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

- *Coarse-grained path space log-likelihood* maximization

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

- No need for microscopic reconstruction: $q^\theta(\sigma, \sigma') = \nu(\sigma' | \mathbf{T}\sigma')\bar{p}^\theta(\mathbf{T}\sigma, \mathbf{T}\sigma')$

Fisher Information Matrix

Parameter identifiability

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

- Asymptotic Gaussianity of the Maximum Likelihood Estimator:

$$\hat{\theta}_N \to \theta^* \text{ a.s. and} N^{-1/2}(\hat{\theta}_N - \theta^*) \rightharpoonup \mathcal{N}(0, \mathbf{F_\mathcal{H}}^{-1}(Q^{\theta^*}))$$

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

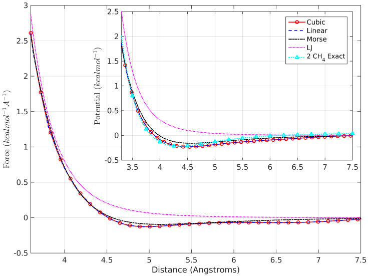
- Estimating the FIM $\mathbf{F}_\mathcal{H}\left(Q^{\hat{\theta}_N}\right)$ provides rigorous error bars on computed optimal parameter values θ^*.

[1] Katsoulakis, P.P., *J.Chem. Phys.* (2013)

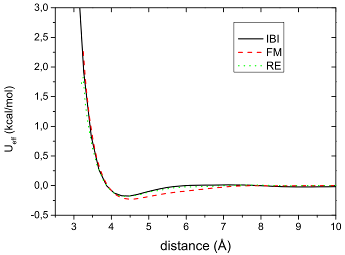
Examples

Bulk methane liquid at equilibrium (NVT)

- CG: CH_4 in one-site representation with a pair potential (non-bonded only)
- micro scale: $T = 100K$, 512 CH_4 molecules, $\rho = 0.38\text{g/cm}^3$



(a) Dynamic force matching scheme the derived potential.

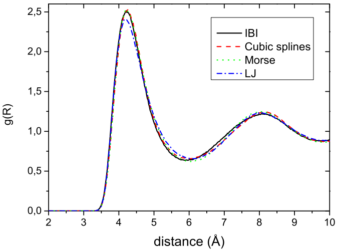


(b) CG effective potentials approximated from different methods

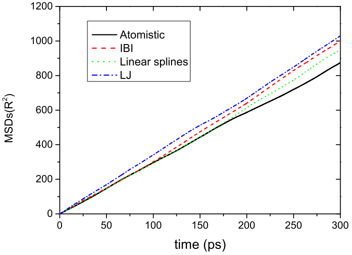
Examples

Bulk methane liquid at equilibrium (NVT)

- CG: CH_4 in one-site representation with a pair potential (non-bonded only)
- micro scale: $T = 100K$, 512 CH_4 molecules, $\rho = 0.38\text{g/cm}^3$



(c) CG pair correlation function.

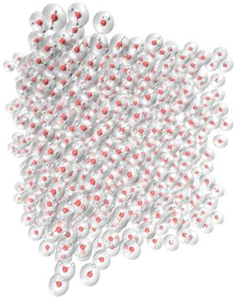


(d) Mean square displacement.

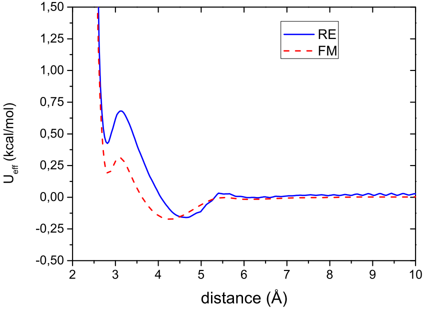
Examples

Water

- CG: H_2O in one-site representation with a pair potential (non-bonded only)
- micro scale: $T = 300K$, $p = 1$atm, 1192 H_2O molecules, SPC/E force-field



(e) Snap shot of micro scale simulation

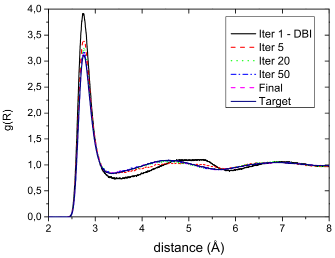


(f) CG effective interactions

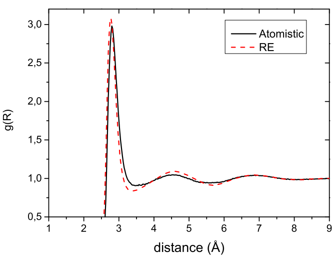
Examples

Water

- CG: H_2O in one-site representation with a pair potential (non-bonded only)
- micro scale: $T = 300K$, $p = 1$atm, 1192 H_2O molecules, SPC/E force-field



(g) CG pair distribution function – inverse Monte Carlo

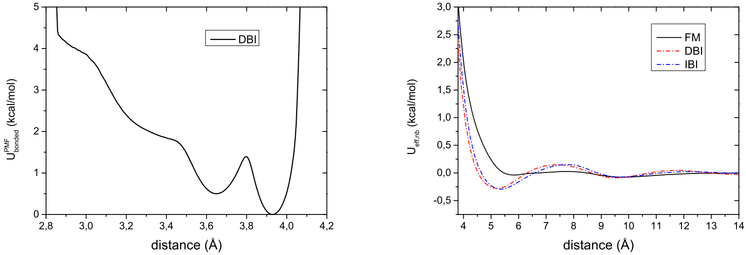


(h) CG pair distribution function – RE based method

Examples

Hexane liquid

- CG: $CH_3(CH_2)_4CH_3$ 3:1 mapping (3 monomers to 1 CG particle) with a pair potential
- micro scale: $T = 300K$, $p = 1$atm, 512 $CH_3(CH_2)_4CH_3$ molecules, OPLS force-field



(i) bonded potential (j) non-bonded potential

Figure: CG effective interactions for hexane molecules.

Examples

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.
- Parametrized coarse-grained potential:

$$\bar{U}(k,\eta) = \sum_{l} \bar{J}(k,l)\eta(k) + \bar{J}(0,0)(\eta(k)-1) - \bar{h}$$

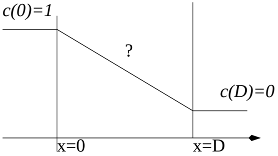
- Coarse-grained rates: assume *local equilibrium*, $\sigma(x) \approx q^{-1}\eta(k)$

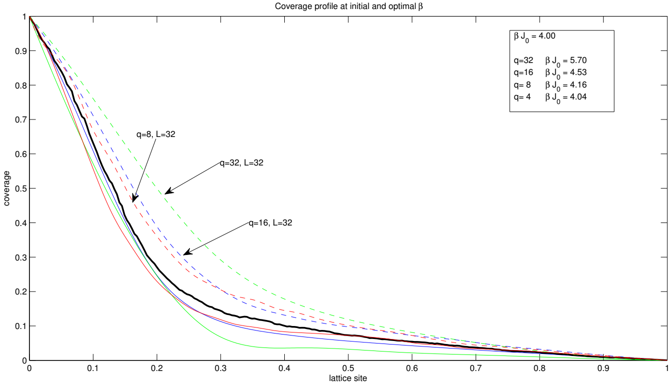
$$\bar{c}(k,l,\eta) = \frac{1}{q}\eta(k)(q-\eta(l))\,d\, e^{-\beta \bar{U}(k,\eta)}$$

- $\theta = (\beta \bar{J}_0, \bar{J}(k,l), \bar{J}(k,l,m)\dots)$

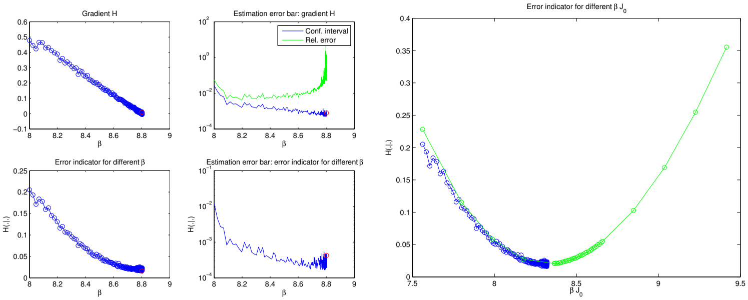
Non-equilibrium stationary states

Bounded domain with a gradient in concentrations





Stationary concentration profile from CG dynamics with fitted rates using $\theta = \beta \bar{J}_0$


Gradient H
β
Estimation error bar: gradient H
Conf. interval
Rel. error
β
Error indicator for different β
H(.|.)
β
Estimation error bar: error indicator for different β
H(.|.)
β
Error indicator for different β J0
H(.|.)
β J0

CG of systems with NESS and RER

- 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]