Variational methods for rare event sampling

Ben Harland

10th November, 2022

Sampling problem & timescales in conformational dynamics

Conformational space

$$x \in \Omega \subset \mathbb{R}^{3N}$$

Vastness of conformational space

- $3N \gg 1$: "curse of dimensionality"
- Levinthal's paradox: "combinatorial hell"

Metastability

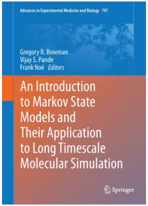
- states & barriers, transitions are rare events
- Boltmann distribution: $Z^{-1}\mathrm{e}^{-\beta u(x)}$

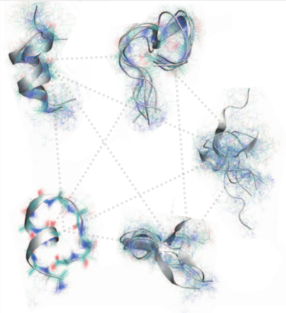
MSM picture of conformational space

Markov state models

$$\Omega \approx \cup_{i=1}^{M} s_i \quad s_i \cap s_j = \emptyset$$

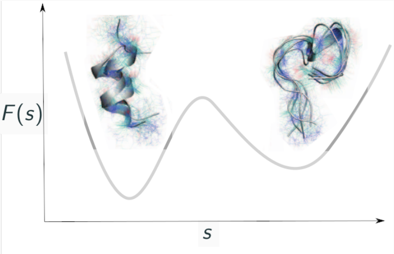
- Zwanzig: Markov approximation, kinetic master equation (1983)
- Schutte, Deuflhard: theory with transfer operator (1999)
- Swope, Pande, Noe, others (∼2005-)
    1. adaptive sampling: "divide and conquer"
    2. focus on eigendecomposition
- Noe: VAC (2016)
- Noe: VAMPnets (2017)





adapted from Noe *et al.*, *J. Chem. Phys.* **126**, 155102 (2007)

Slow manifold picture



$F(s)$ = free energy surface

$s \in \Omega_s \subset \mathbb{R}^d, \quad d \ll 3N$

- reaction coordinate
- order parameter
- collective variables
- latent variables
- Ω_s = "slow manifold"

What makes a good s, Ω_s?

Parrinello, Zuckerman: "don't marginalize barriers"

Barducci, Bonomi, Parrinello, *WIREs Comp. Mol. Sci.*, **1**, 826 (2011)
`https://statisticalbiophysicsblog.org/?p=160#more-160` (2017)

Pande: "natural reaction coordinate"

McGibbon, Husic, Pande *J. Chem. Phys.* **146**, 044109 (2017)

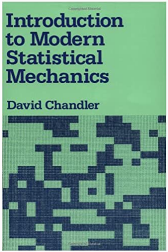
"uniquely satisfied by a dominant eigenfunction of an integral operator associated with the ensemble dynamics."

Noe & Clementi: "slow collective variables"

Noe, Clementi, *Curr. Op. Struct. Biol.* **43**, 141 (2017)

"naturally leads to the concept of eigenfunctions of the dynamical operator underlying MD simulation"

VAC I: Time correlation functions



Onsager's principle states that in the linear regime, the relaxation obeys

$$\frac{\Delta\bar{A}(t)}{\Delta\bar{A}(0)} = \frac{C(t)}{C(0)},$$

where

$$\Delta\bar{A}(t) = \bar{A}(t) - \langle A \rangle = \overline{\delta A}(t)$$

and

$$C(t) = \langle \delta A(0)\, \delta A(t) \rangle.$$

$$\begin{aligned}
C_{ij}(\tau) &= \langle \zeta_i(0)\zeta_j(\tau) \rangle && \text{zero mean} \\
&= \mathbb{E}\left[\zeta_i(x_t)\zeta_j(x_{t+\tau})\right] && \text{stationary, ergotic} \\
&= \int \underbrace{\mathrm{d}x\ Z^{-1}\mathrm{e}^{-\beta u(x)}}_{\mu(\mathrm{d}x)} \zeta_i(x) \underbrace{\mathbb{E}\left[\zeta_j(x_{t+\tau}) | x_t = x\right]}_{(\hat{K}_\tau \circ \zeta_j)(x)} \\
&= \langle \zeta_i | \hat{K}_\tau \circ \zeta_j \rangle_\mu
\end{aligned}$$

VAC II: Detailed balance

Reversible dynamics: transfer operator is self-adjoint

$$\langle \zeta_i | \hat{K}_\tau \circ \zeta_j \rangle \equiv \langle \hat{T}_\tau \circ \zeta_i | \zeta_j \rangle = \langle \zeta_i | \hat{T}_\tau \circ \zeta_j \rangle$$

Self-adjoint operators

- real eigenvalues

$$\hat{T}_\tau \circ \psi_i = \lambda_i(\tau)\, \psi_i \qquad \lambda_i(\tau) \in \mathbb{R}$$

- orthogonal eigenfunctions

$$\langle \psi_i | \psi_j \rangle = \delta_{ij} \qquad \|\psi_i\| = \sqrt{\langle \psi_i | \psi_i \rangle} = 1$$

- completeness, spectral decomposition

$$\mathbb{1} = \sum_{i=0}^{\infty} |\psi_i\rangle\langle\psi_i| \qquad \hat{T}_\tau = \sum_{i=0}^{\infty} \lambda_i(\tau)\, |\psi_i\rangle\langle\psi_i|$$

VAC III: Variational principle

Timescales

$$\lambda_i(\tau) = \lambda_i^\tau = \mathrm{e}^{-\kappa_i \tau} = \mathrm{e}^{-\tau/\tau_i}$$

Perron spectrum

$$1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq 0 \qquad \Big(|\psi_0\rangle = |1\rangle \Big)$$

Variational theorem

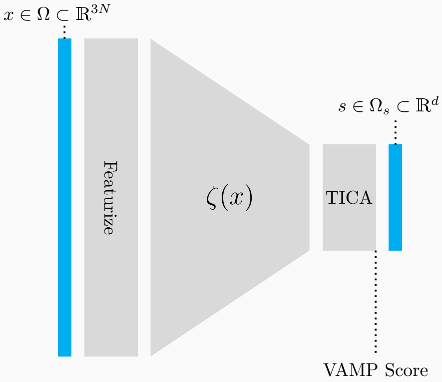
$$\lambda_i^\tau, \psi_i = \max_u, \arg\max_u \ \langle u | \hat{T}_\tau \circ u \rangle \ \Big| \ \langle u|u\rangle = 1, \ \langle u|\psi_{j<i}\rangle = 0$$

"Deep TICA": neural network + TICA

Chen, Sidky, Ferguson, *J. Chem. Phys.* **150**, 212112 (2019)

$$\mathrm{SGD}: \ \big\{ |\zeta_i(\alpha)\rangle \big\}_{i=1}^{d} \qquad \mathrm{TICA}: \ |\tilde{\psi}_i\rangle = \sum_{j=1}^{d} c_{ij} |\zeta_j(\alpha)\rangle$$

VAMP scores



Covariance matrices

$$C_0 = \mathbb{E}\left[\zeta(x_t)\zeta(x_t)^\top\right]$$

$$C_\tau = \mathbb{E}\left[\zeta(x_t)\zeta(x_{t+\tau})^\top\right]$$

Rayleigh trace

$$\begin{aligned}\text{VAMP-1} &= 1 + \sum_{i=1}^{d} \tilde{\lambda}_i \\ &= 1 + \operatorname{tr}\ C_0^{-1/2} C_\tau C_0^{-1/2}\end{aligned}$$

Kinetic variance

$$\begin{aligned}\text{VAMP-2} &= 1 + \sum_{i=1}^{d} \tilde{\lambda}_i^2 \\ &= 1 + \left\| C_0^{-1/2} C_\tau C_0^{-1/2} \right\|_F\end{aligned}$$

Sampling the ensemble

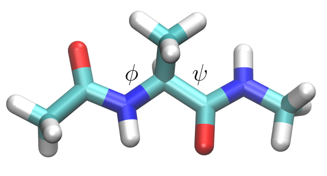
Equilibrium distribution + transitions

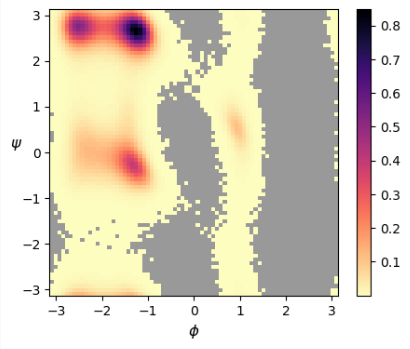
$$\{x_t, x_{t+\tau}\} \;\Rightarrow\; \mathbb{E}\left[\zeta(x_t)\zeta(x_{t+\tau})^\top\right]$$

Methods to consider (remaining presentation)

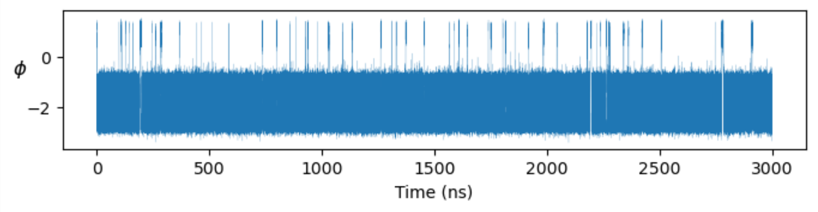
1. Brute force — simulation results
2. Weighted ensemble — brief
3. Boltzmann generators — brief
4. Enhanced sampling — theory & some results

Brute force: 3 μs simulation of "alanine dipeptide"





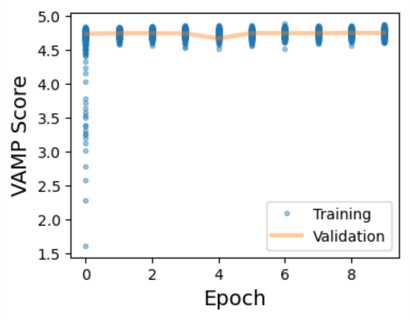
amber99sb
538 tip3p waters
(2.6 nm)^3 cell
$\tau = 1$ ps



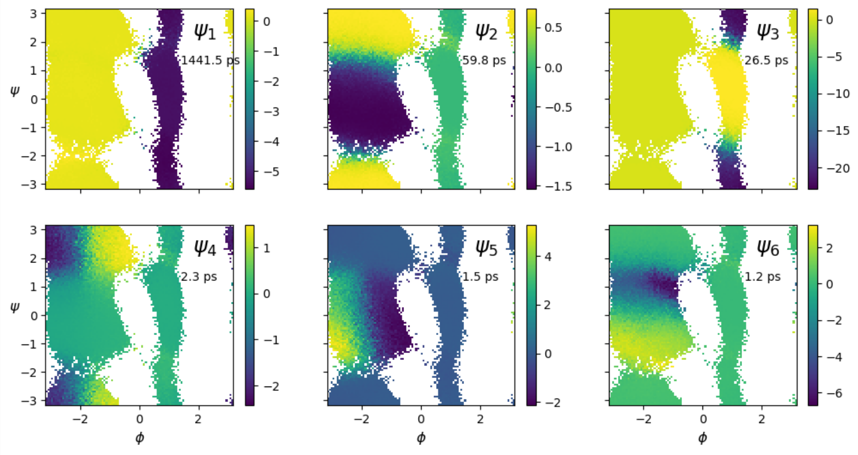
Training

Network

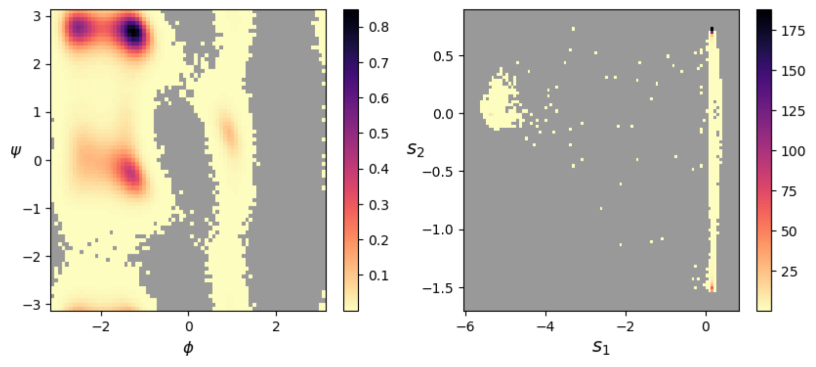
1. featurization: 45 heavy atom distances
2. batchnorm
3. 100 nodes, elu
4. 100 nodes, elu
5. 30 nodes, elu
6. 6 output nodes, tanh



Eigenfunctions

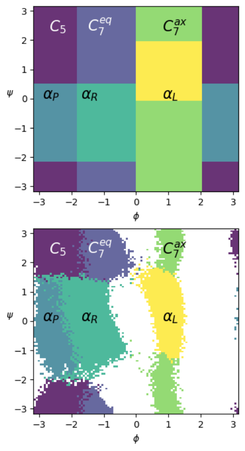


Equilibrium distribution in Ω_s

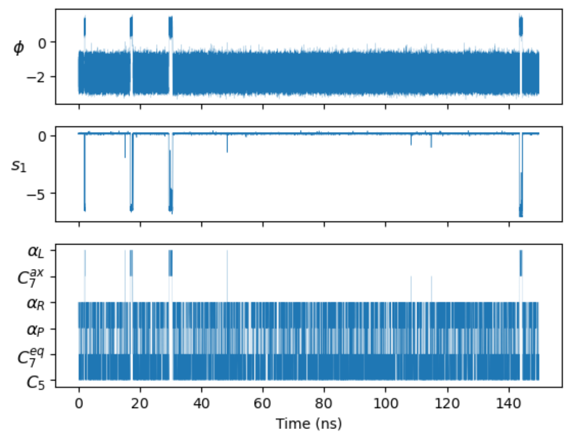


Compare to MSM's

States

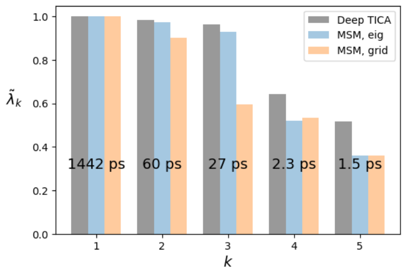


Trajectory (150 ns)

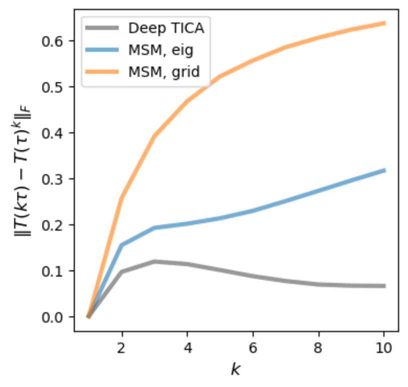


Analysis I

Eigenvalues

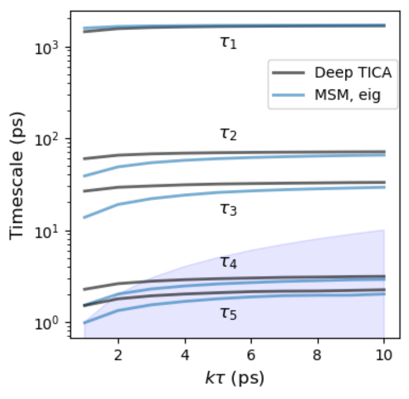


Chapman-Kolmogorov test

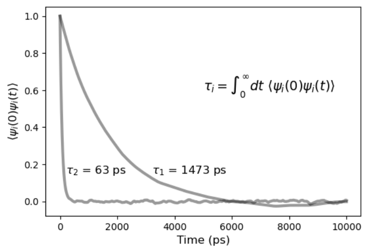


Analysis II

Timescales



Sanity check: autocorrelation functions

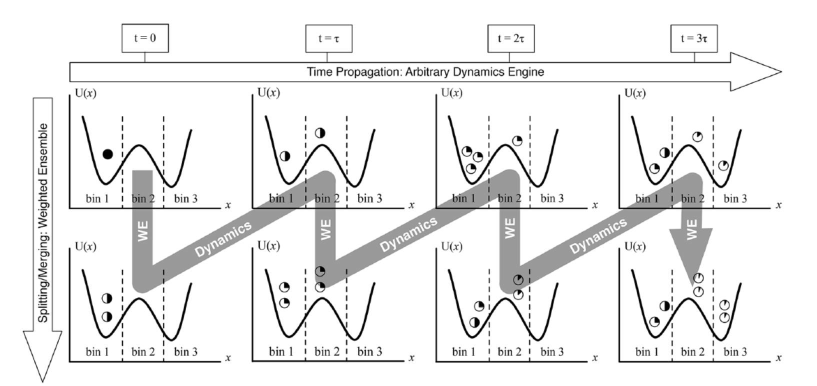


Remaining presentation

1.	Brute force	simulation results
2.	Weighted ensemble	brief
3.	Boltzmann generators	brief
4.	Enhanced sampling	theory & some results

Weighted ensemble

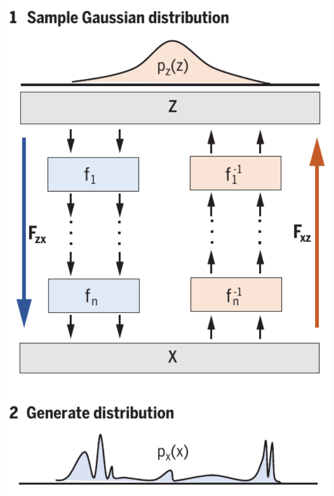
Zuckerman & Chong, *Annu. Rev. Biophys*, **46**, 43 (2017)

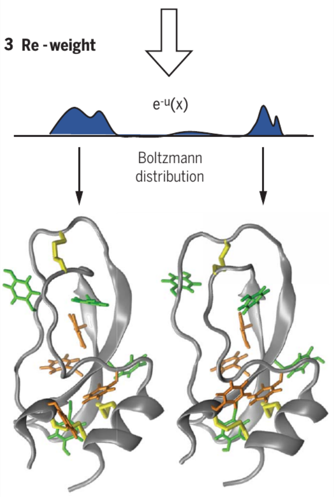


Ensemble dynamics
Voter, *Phys. Rev. B*, **57**, R13-985 (1998)

Boltmann generators

Noé, Olsson, Köhler & Wu, *Science*, **365**, 1147 (2019)





Remaining presentation

1.	Brute force	simulation results
2.	Weighted ensemble	brief
3.	Boltzmann generators	brief
4.	Enhanced sampling	theory & some results

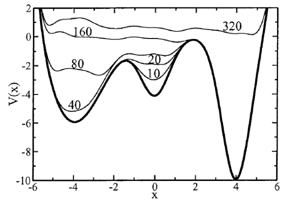
MetaD
Laio & Parrinello,
PNAS, **99**, 12562 (2002)

WTMetaD
Barducci, Bussi & Parrinello,
Phys. Rev. Lett., **100**, 020603 (2008)

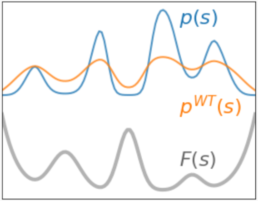
VES
Valsson & Parrinello,
Phys. Rev. Lett., **113**, 090601 (2014)

OPES
Invernizzi & Parrinello,
Phys. Chem. Lett, **11**, 2731, (2020)

Bias potential: $w(s(x))$



$$p^{\mathrm{WT}}(s) \propto p(s)^{1/\gamma}$$



WTMetaD

Sampling histogram

$$N_t(s): \quad \dot{N}_t(s) = \delta(s - s_t), \quad N_0(s) = 0$$

Bias potential ansatz

$$w_t(s) = (\beta')^{-1} \log\left(1 + \beta' \omega N_t(s)\right)$$

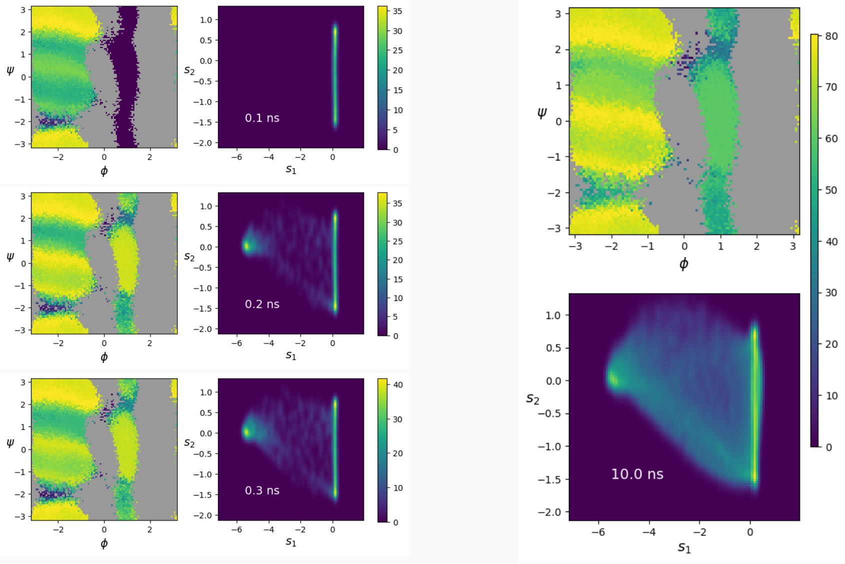
Idealized dynamical equation

$$\dot{w}_t(s) = \omega \mathrm{e}^{-\beta' w_t(s)} \delta(s - s_t)$$

Asymptotic limit

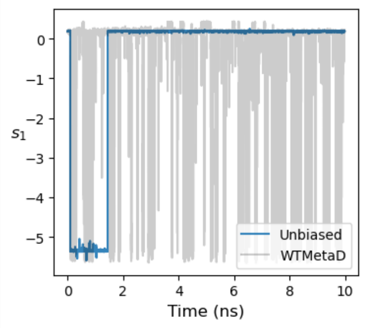
$$\lim_{t\to\infty} w_t(s) = -(1 - \gamma^{-1}) F(s) \qquad \gamma = 1 + \frac{\beta}{\beta'}$$

Evolution of $w_t(s)$

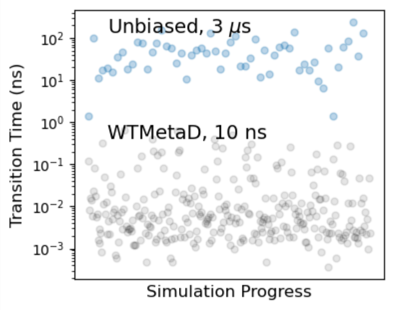


Enhanced sampling

Sampling s_1



Transition times



https://github.com/bbharland/dialanine-metad

Bonati, Piccini, Parrinello, *PNAS*, **118**, e2113533118 (2021)

Probability enhanced sampling

Ensemble induced by $w(s)$

$$p_w(s) \propto \mathrm{e}^{-\beta[F(s)+w(s)]}$$

Bias-driven

	$w_{t\to\infty}(s)$	Ensemble
MetaD	$-F(s)$	$p^{\mathrm{Uniform}}(s) = \lvert\Omega_s\rvert^{-1}$
WTMetaD	$-(1-\gamma^{-1})F(s)$	$p^{\mathrm{WT}}(s) \propto p(s)^{1/\gamma}$

Ensemble-driven

OPES, VES: "probability enhanced sampling"
target distribution, $p^*(s)$

VES variational principle

Convex functional

$$\Phi\left[w; p^*\right] = k_{\mathrm{B}} T \log \frac{\displaystyle\int \mathrm{d}s\ \mathrm{e}^{-\beta[F(s)+w(s)]}}{\displaystyle\int \mathrm{d}s\ \mathrm{e}^{-\beta F(s)}} + \int \mathrm{d}s\ p^*(s) w(s)$$

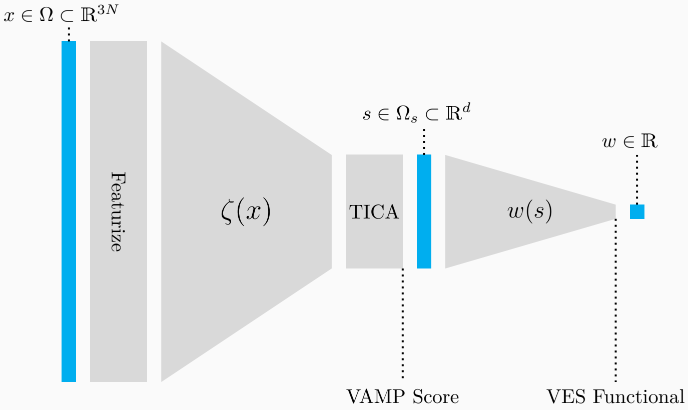
$$\arg\min_{w} \Phi\left[w; p^*\right] = w^*$$

Estimating Φ from data

$$\Phi\left[w; p^*\right] = -k_{\mathrm{B}} T \log \mathbb{E}_{p_w}\left[\mathrm{e}^{+\beta w}\right] + \mathbb{E}_{p^*}\left[w\right]$$

Minimizing the functional

$$\Phi\left[w; p^*\right] = -k_{\mathrm{B}} T \log \mathbb{E}_{p_w}\left[\mathrm{e}^{+\beta w}\right] + \mathbb{E}_{p^*}\left[w\right]$$



Algorithm overview

Take well-tempered target

Valsson & Parrinello, *JCTC*, **11**, 1996 (2015)

$$p^*(s) = p^{\mathrm{WT}}(s) \equiv p_\gamma(s) \propto \mathrm{e}^{-\beta F(s)/\gamma}$$

Situation at outset

- $s(x)$, $w(s)$: untrained networks
- $p_w(s)$: empty dataset
- no knowledge of $p(s)$, $F(s)$
- $p_\gamma(s)$ defined by $F(s)$

Goal: iteratate until self-consistent

$$\underbrace{F(s) + w(s)}_{p_w} = \underbrace{\gamma^{-1} F(s)}_{p_\gamma}$$

Algorithm setup

1. Initial target distribution (grid)

$$F^{(0)}(s) = 0 \qquad p_\gamma^{(0)}(s) \propto \mathrm{e}^{-\beta F^{(0)}(s)/\gamma}$$

2. Unbiased seed simulation

$$\{x_t, x_{t+\tau}\}$$

3. Train

$$\begin{aligned} s^{(0)}(x) &\leftarrow \{x_t, x_{t+\tau}\},\ \mathrm{VAMP\text{-}2} \\ w^{(0)}(s) &\leftarrow \mathbb{E}_{p_w^{(0)}}\left[\mathrm{e}^{+\beta w}\right],\ \mathbb{E}_{p_\gamma^{(0)}}\left[w\right],\ \Phi \end{aligned}$$

Algorithm, k^{th} iteration

1. Update target distribution (grid)

$$F^{(k)}(s) = -w^{(k-1)}(s) + \gamma^{-1} F^{(k-1)}(s)$$
$$p_\gamma^{(k)}(s) \propto \mathrm{e}^{-\beta F^{(k)}(s)/\gamma}$$

2. Biased simulation, $\{x_t\}_{w^{(k-1)}}$
3. Reweight and accumulate in $\{x_t, x_{t+\tau}\}$
4. Train

$$s^{(k)}(x) \ \leftarrow \ \{x_t, x_{t+\tau}\},\ \text{VAMP-2}$$
$$w^{(k)}(s) \ \leftarrow \ \mathbb{E}_{p_w^{(k)}}\left[\mathrm{e}^{+\beta w}\right],\ \mathbb{E}_{p_\gamma^{(k)}}\left[w\right],\ \Phi$$

Main ideas

Leading eigenfunctions of the Koopman/transfer operator provide the optimal slow manifold.

1. s as molecular df
2. s as observables
3. s as projections of dynamics. Ekhart-Young:

$$\underset{\hat{O}^{(d+1)}}{\arg\min} \left\| \hat{K}_\tau - \hat{O}^{(d+1)} \right\|_F = |1\rangle\langle 1| + \sum_{i=1}^{d} \lambda_i^\tau |\psi_i\rangle\langle\psi_i|$$

Adiabatically enhancing sampling in this subspace provides a compelling approach to the sampling problem.

Deep-TICA/VES approach solves chicken/egg problem? Network resembles an end-to-end learning framework (fashionable).