Simulated Tempering Method in the Infinite Switch Limit with Adaptive Weight Learning

Anton Martinsson*, Jianfeng Lu, Benedict Leimkuhler, Eric Vanden-Eijnden

University of Edinburgh,
Maxwell Institute for Graduate Studies in Analysis and its Applications

15th November
Molecular Dynamics Sampling

Canonical distribution for $q$ (positions) and $p$ (momentum):

$$\rho_\beta(q, p) = Z^{-1}(\beta)e^{-\beta \frac{1}{2}p^Tm^{-1}p - \beta V(q)}$$

where $V(q)$ is potential and $Z(\beta)$ is the normalisation constant, $\beta > 0$ reciprocal temperature.
From Standard to Accelerated Sampling

- Ergodic Average of observable $A(q)$,

$$\mathbb{E}_\beta [A] = \int A(q) \rho_\beta(q, p) \, dp \, dq = \lim_{T \to \infty} \frac{1}{T} \int_0^T A(q(t)) \, dt$$

- Standard sampling methods for $\rho_\beta(q, p)$ such as e.g. Monte Carlo or Langevin Dynamics,

$$dq = m^{-1} p \, dt$$
$$dp = -\nabla V(q) \, dt - \gamma p \, dt + \sqrt{2\gamma \beta^{-1} m} \, dW$$

struggle with energetic and entropic barriers

- **Accelerated sampling**: Simulated Annealing, Replica Exchange Molecular Dynamics, Simulated Tempering, Wang-Landau, Adaptive Force Biasing, Temperature-accelerated Molecular Dynamics, Hamiltonian Replica Exchange, ...
Simulated Tempering

- Temperature “ladder” with $M$ steps, with spacing $\Delta T$
  \[
  T_{\text{min}} = T_1 < \ldots < T_M = T_{\text{max}}
  \]

- Let $\beta_i = (k_B T_i)^{-1}$ with assigned weight $\omega(\beta_i)$
- Switch every $\tau$ steps from $\beta_i \rightarrow \beta_j$ with probability $\alpha_{ij}$,
  \[
  \alpha_{ij} = \min\{1, \frac{\omega(\beta_i)}{\omega(\beta_j)} e^{-(\beta_i - \beta_j) V(q)}\}
  \]

- Invariant measure with density,
  \[
  \rho(q, p, \beta_i) = C^{-1}(\beta) \omega(\beta_i) e^{-\frac{1}{2} \beta p^T m^{-1} p - \beta_i V(q)}
  \]
  where $C(\beta) = \sum_{i=1}^{M} \int_{\mathcal{D} \times \mathbb{R}^d} \omega(\beta_i) e^{-\frac{1}{2} \beta p^T m^{-1} p - \beta_i V(q)} \, dq \, dp$
Simulated Tempering Parameters

1. Choice of switch period, $\tau$?

2. How do we justify $\omega(\beta_i) = Z_q^{-1}(\beta_i)$?

   $$\mathbb{P}(\beta_i) = \frac{\omega(\beta_i) Z_q(\beta_i)}{\sum_{j=1}^{M} \omega(\beta_j) Z_q(\beta_j)}.$$

3. Can we learn $\omega(\beta_i)$ weights on-the-fly?

Remarks:

- Grid Spacing, $\Delta T$. Is there a way to motivate the choice?

- How do we calculate an average from $\rho_{\beta_i}$ i.e $\mathbb{E}_{\beta_i}[A]$?
Infinite Switch Limit of Simulated Tempering, $\tau \to 0$. 
Summary of $\tau \to 0$ Results

- Follows same large deviation approach as for REMD by Dupuis et. al. \(^1\) \(^2\)

- ST in $\tau \to 0$ limit for some continuous $\beta_c \in [\beta_{\text{min}}, \beta_{\text{max}}]$ is equivalent to sampling the averaged potential \(^3\),

$$\bar{V}(q) = -\beta^{-1} \log \int_{\beta_{\text{min}}}^{\beta_{\text{max}}} \omega(\beta_c) e^{-\beta_c V(q)} d\beta_c$$

for some prior known weights $\omega(\beta_c)$.

---

\(^1\) Dupuis et. al. 2012.  
\(^2\) Plattner et. al 2011.  
\(^3\) A.M. et. al. 2018.
Outline of $\tau \to 0$ Results

- The rate functional $\mathbb{P}(\nu_T \approx \mu) \propto \exp \left[ -T^{-1} I_\tau(\mu) \right]$ of the ergodic dynamics is a monotonically decreasing function of $\tau$, i.e if

  $$\tau < \tau' \implies I_\tau(\mu) \geq I_{\tau'}(\mu)$$

  For faster convergence to $\mu$, let $\tau \to 0$.

- The dynamical equations in $\tau \to 0$ limit,

  $$dq = m^{-1} p \, dt,$$

  $$dp = -\beta^{-1} \bar{\beta}(V(q)) \nabla V \, dt - \gamma p \, dt + \sqrt{2\gamma \beta^{-1} m^{-1}} \, dW_p$$

  which implies that we have averaged potential,

  $$\bar{V}(q) = -\beta^{-1} \log \int_{\beta_{\text{min}}}^{\beta_{\text{max}}} \omega(\beta_c) e^{-\beta_c V(q)} \, d\beta_c$$
Simulated Tempering Parameters

1. Choice of switch period, $\tau$?

2. How do we justify $\omega(\beta_i) = Z_q^{-1}(\beta_i)$?

$$
\mathbb{P}(\beta_i) = \frac{\omega(\beta_i) Z_q(\beta_i)}{\sum_{j=1}^{M} \omega(\beta_j) Z_q(\beta_j)}.
$$

3. Can we learn $\omega(\beta_i)$ weights on-the-fly?

Remarks:

- Grid Spacing, $\Delta T$. Is there a way to motivate the choice?

- How do we calculate an average from $\rho_{\beta_i}$ i.e $\mathbb{E}_{\beta_i}[A]$?
Justification of $\omega(\beta) \propto Z_q^{-1}(\beta)$. 
Overview of $\omega(\beta) \propto Z_q^{-1}(\beta)$

- The density function of $V(q)$

$$\bar{\rho}(E) = \frac{\int_{\beta_{\min}}^{\beta_{\max}} e^{-\beta_c E} \omega(\beta_c) \, d\beta_c}{\int_{\beta_{\min}}^{\beta_{\max}} Z_q(\beta_c) \omega(\beta_c) \, d\beta_c} \Omega(E),$$

where $\Omega(E) = \int_D \delta(V(q) - E) \, dq$.

- In that large system size limit,

$$\bar{\rho}(E) \asymp 1$$

when $\omega(\beta) \propto Z_q^{-1}(\beta)$. 
Simulated Tempering Parameters

1. Choice of switch period, $\tau$?

2. How do we justify $\omega(\beta_i) = Z_q^{-1}(\beta_i)$?

$$P(\beta_i) = \frac{\omega(\beta_i) Z_q(\beta_i)}{\sum_{j=1}^{M} \omega(\beta_j) Z_q(\beta_j)}.$$  

3. Can we learn $\omega(\beta_i)$ weights on-the-fly?

Remarks:

- Grid Spacing, $\Delta T$. Is there a way to motivate the choice?

- How do we calculate an average from $\rho_{\beta_i}$ i.e $\mathbb{E}_{\beta_i}[A]$?
Adaptive $\omega(\beta_c)$ adjustment as we learn $Z_q(\beta_c)$. 
For $\beta_c \in [\beta_{\text{min}}, \beta_{\text{max}}]$ we can construct

$$z(t, \beta_c) = \frac{1}{t} \int_0^t \frac{e^{-\beta_c V(q(s))}}{\int_{\beta_{\text{min}}}^{\beta_{\text{max}}} \omega(\beta'_c) e^{-\beta'_c V(q(s))} \, d\beta'_c} \, ds$$

which can be calculated for all $t > 0$.

In the limit as $t \to \infty$,

$$\lim_{t \to \infty} z(t, \beta_c) = \frac{Z_q(\beta_c)}{\int_{\beta_{\text{min}}}^{\beta_{\text{max}}} Z_q(\beta'_c) \omega(\beta'_c) \, d\beta'_c},$$

This implies that we learn ratios of $Z_q(\beta_c)$ regardless of knowledge of $\omega(\beta_c)$,

$$\lim_{t \to \infty} \frac{z(t, \beta_c)}{z(t, \beta'_c)} = \frac{Z_q(\beta_c)}{Z_q(\beta'_c)}$$
Define $z(t, \beta_c)$ as,

$$z(t, \beta_c) = \frac{1}{t} \int_0^t \frac{e^{-\beta_c V(q(s))}}{\int_{\beta_c}^{\beta_{\text{max}}} \omega(\beta'_c) e^{-\beta'_c V(q(s))} d\beta'_c} ds$$

Adjusting the weights according to,

$$\kappa \omega(t, \beta_c) = z^{-1}(t, \beta_c) - \lambda(t) \omega(t, \beta), \text{ with } \lambda(t) = \int_{\beta_{\text{min}}}^{\beta_{\text{max}}} z^{-1}(t, \beta_c) d\beta_c$$

will ensure that $\omega(t, \beta_c) \propto Z_q^{-1}(\beta_c)$ as $t \to \infty$. 
Simulated Tempering Parameters

1. Choice of switch period, $\tau$?

2. How do we justify $\omega(\beta_i) = Z_q^{-1}(\beta_i)$?

$$
\mathbb{P}(\beta_i) = \frac{\omega(\beta_i) Z_q(\beta_i)}{\sum_{j=1}^{M} \omega(\beta_j) Z_q(\beta_j)}.
$$

3. Can we learn $\omega(\beta_i)$ weights on-the-fly?

Remarks:

- Grid Spacing, $\Delta T$. Is there a way to motivate the choice?

- How do we calculate an average from $\rho_{\beta_i}$ i.e. $\mathbb{E}_{\beta_i}[A]$?
The equations of motion are,

\[ dq = m^{-1} p \, dt, \]
\[ dp = -\beta^{-1} \hat{\beta}(t, V(q)) \nabla V \, dt - \gamma p \, dt + \sqrt{2\gamma\beta^{-1}} m \, dW_p \]

We need to calculate the following force re-scaling

\[ \hat{\beta}(t, V(q)) = \frac{\int_{\beta_{\text{min}}}^{\beta_{\text{max}}} \omega(t, \beta_c) e^{-\beta_c V(q)} \, d\beta_c}{\int_{\beta_{\text{min}}}^{\beta_{\text{max}}} \omega(t, \beta_c) e^{-\beta_c V(q)} \, d\beta_c} \]

We therefore discretise \([\beta_{\text{min}}, \beta_{\text{max}}]\) with the aim of calculating \(\hat{\beta}(t, V(q))\)
Calculating Averages

- In standard Simulated Tempering

\[ \mathbb{E}_{\beta_c}[A] = \lim_{T \to \infty} \frac{1}{T} \int_0^T A(q(t)) \mathbb{1}_{\beta_c} \, dt \]

- We instead use a reweighting of the entire trajectory (importance sampling),

\[ \mathbb{E}_{\beta_c}[A] = \lim_{T \to \infty} \frac{1}{T} \int_0^T A(q(t)) W_{\beta_c}(t, q(t)) \, dt \]

where,

\[ W_{\beta_c}(t, q) = \frac{z^{-1}(t, \beta_c)}{\int_{\beta_{\min}}^{\beta_{\max}} \omega(t, \beta'_c) e^{-(\beta'_c - \beta_c)V(q)} d\beta'_c} \]

where \( z(t, \beta_c) \) is the estimate of \( Z_q(\beta_c) \) at time \( t \).
Numerical Experiments : Part 1

Curie Weiss Magnet – mean field Isings model
Curie Weiss Magnet

Simulated Tempering

Infinite Switch ST

\[ \beta = 1 \, \beta = 2 \, \beta = 3 \]
Molecular Integration Simulation Toolkit (MIST) integrates with Amber 14, GROMACS 5.0.2, NAMD-Lite 2.0.3, LAMMPS.

In vacuum, using Gromacs with MIST.

Measure RMSD and radius of gyration, $r_g$, from initial helix state.

---


5. [https://bitbucket.org/extasy-project/mist/wiki/Home](https://bitbucket.org/extasy-project/mist/wiki/Home)
Alanine-12: LD vs ST vs ISST

6 Bethune et. al., 2018
Alanine-12: ISST

Free energy at 498K

Rg (nm)

RMSD (nm)

Free Energy (kcal/mol units)
Alanine-12: ISST

Free energy at 413K

Rg (nm)

RMSD (nm)

Free Energy (kcal/mol units)
Alanine-12: ISST

Free energy at 354K

Free Energy (kcal/mol units)

Rg (nm)

RMSD (nm)
Alanine-12: ISST

Free energy at 300K

Rg (nm)

RMSD (nm)

Free Energy (kcal/mol units)
Alanine-12: LD vs ISST

Free energy at 300K

Rg (nm)

RMSD (nm)

Bethune et. al., 2018
Conclusions

- Operate ST in the Infinite Switch limit, with adaptive weight learning – without temperature being a dynamical variable.

- We have presented work on MD, future work is looking into applying the method in data science and ML.

- We have implemented the method in MIST and it is available to download at https://bitbucket.org/extasy-project/mist/wiki/Home

---

Paul Dupuis, Yufei Liu, Nuria Plattner, and J. D. Doll.
On the Infinite Swapping Limit for Parallel Tempering.

Nuria Plattner, J D Doll, Paul Dupuis, Hui Wang, Yufei Liu, and J E Gubernatis.
An infinite swapping approach to the rare-event sampling problem.

Iain Bethune, Ralf Banisch, Elena Breitmoser, et. al.
MIST: A Simple and Efficient Molecular Dynamics Abstraction Library for Integrator Development
*Computer Physics Communications*, 2018

Anton Martinsson, Jianfeng Lu, Ben Leimkuhler, Eric Vanden-Eijnden,
Simulated Tempering Method in the Infinite Switch Limit with Adaptive Weight Learning,
to appear in JSTAT, 2018