Continuously tempered Hamiltonian Monte Carlo

MATT GRAHAM <matt-graham.github.io>   AMOS STORKEY <amos.storkey.org>

APPROXIMATE INFERENCE

Given a (unnormalised) target density on $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^D$

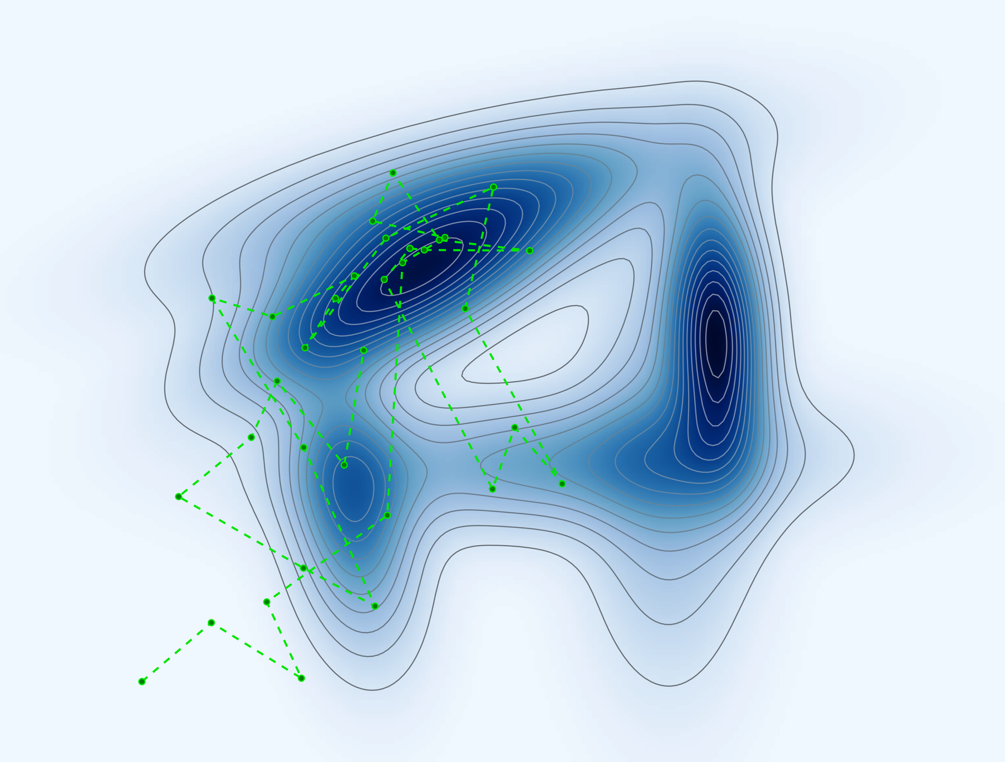
$$\pi(\boldsymbol{x}) \propto \exp\left[-\phi(\boldsymbol{x})\right],$$

how can we compute expectations with respect to π

$$\mathbb{E}_\pi[f] = \int_{\mathcal{X}} f(\boldsymbol{x})\, \pi(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$$

and the unknown normalising constant of the density

$$Z = \int_{\mathcal{X}} \exp\left[-\phi(\boldsymbol{x})\right] \mathrm{d}\boldsymbol{x}?$$



MARKOV CHAIN MONTE CARLO (MCMC)

Define a transition operator T which leaves π invariant

$$\pi(\boldsymbol{x}') = \int_{\mathcal{X}} T(\boldsymbol{x}' \mid \boldsymbol{x})\, \pi(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}.$$

Sample arbitrary initial state then iteratively apply transition operator

$$\boldsymbol{x}^{(0)} \leftarrow p_0(\cdot) \qquad \boldsymbol{x}^{(s)} \leftarrow T(\cdot \mid \boldsymbol{x}^{(s-1)}) \quad \forall t \in \{1 \dots S\}.$$

Providing chain is ergodic, time averages converge to space averages

$$\mathbb{E}_\pi[f] \approx \frac{1}{S}\sum_{s=1}^{S}\left[f(\boldsymbol{x}^{(s)})\right].$$

$$T(\boldsymbol{x}' \mid \boldsymbol{x}) = ? \qquad Z = ?$$

HAMILTONIAN MONTE CARLO (HMC) [1, 2]

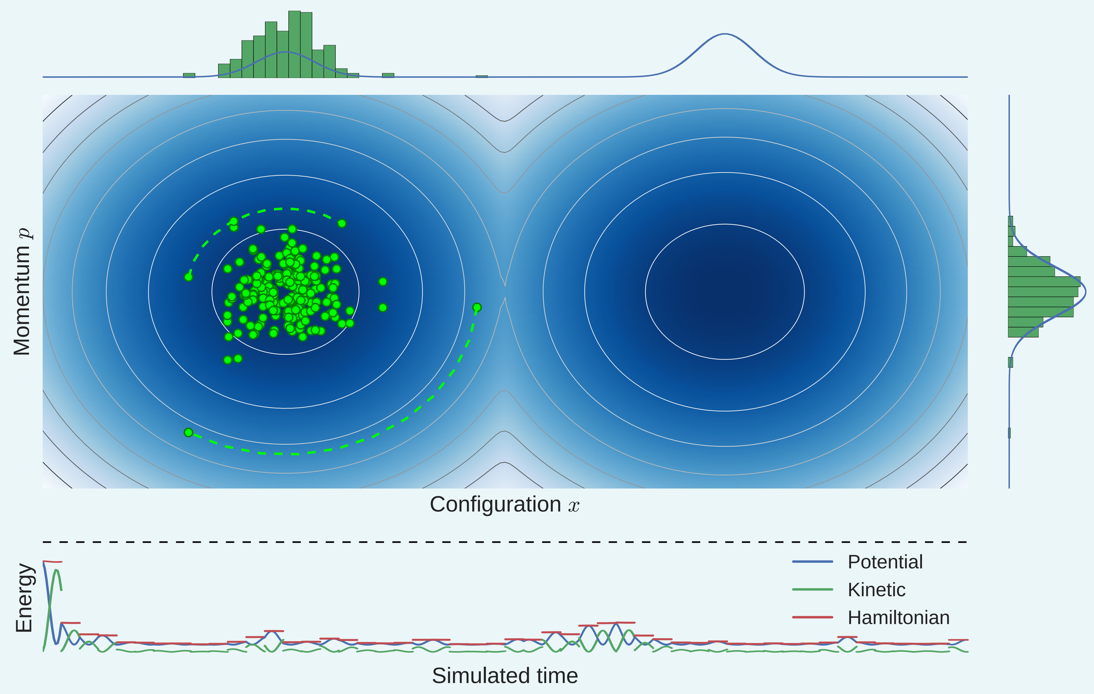
$$\boldsymbol{x} \in \mathbb{R}^D \to (\boldsymbol{x},\, \boldsymbol{p}) \in \mathbb{R}^D \times \mathbb{R}^D \qquad \pi[\boldsymbol{x},\, \boldsymbol{p}] \propto \exp\underbrace{\left[-\phi(\boldsymbol{x}) - \frac{1}{2}\boldsymbol{p}^{\mathrm{T}}\mathbf{M}^{-1}\boldsymbol{p}\right]}_{-H(\boldsymbol{x},\boldsymbol{p})}$$

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \frac{\partial H}{\partial \boldsymbol{p}} = \mathbf{M}^{-1}\boldsymbol{p}, \qquad \frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t} = -\frac{\partial H}{\partial \boldsymbol{x}} = -\frac{\partial \phi}{\partial \boldsymbol{x}}$$

$$\left\{\begin{array}{lll} \text{for } s \in \{0,\, 1 \dots\, L-1\}: & & \\ \boldsymbol{p}^{(s+0.5)} = & \boldsymbol{p}^{(s)} & - \frac{\delta t}{2}\left.\frac{\partial \phi}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}^{(s)}} \\ \boldsymbol{x}^{(s+1)} = & \boldsymbol{x}^{(s)} & + \delta t\, \mathbf{M}^{-1}\boldsymbol{p}^{(s+0.5)} \\ \boldsymbol{p}^{(s+1)} = & \boldsymbol{p}^{(s+0.5)} & - \frac{\delta t}{2}\left.\frac{\partial \phi}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}^{(s+1)}} \end{array}\right.$$

$$a\left[\boldsymbol{x}'\, \boldsymbol{p}' \mid \boldsymbol{x}, \boldsymbol{p}\right] = \min\left\{1,\, \exp\left[H(\boldsymbol{x},\, \boldsymbol{p}) - H(\boldsymbol{x}',\, \boldsymbol{p})\right]\right\}$$

HMC IN 1D GAUSSIAN MIXTURE



BLACK-BOX INFERENCE WITH HMC

mc-stan.org   PyMC3 pymc-devs.github.io/pymc3

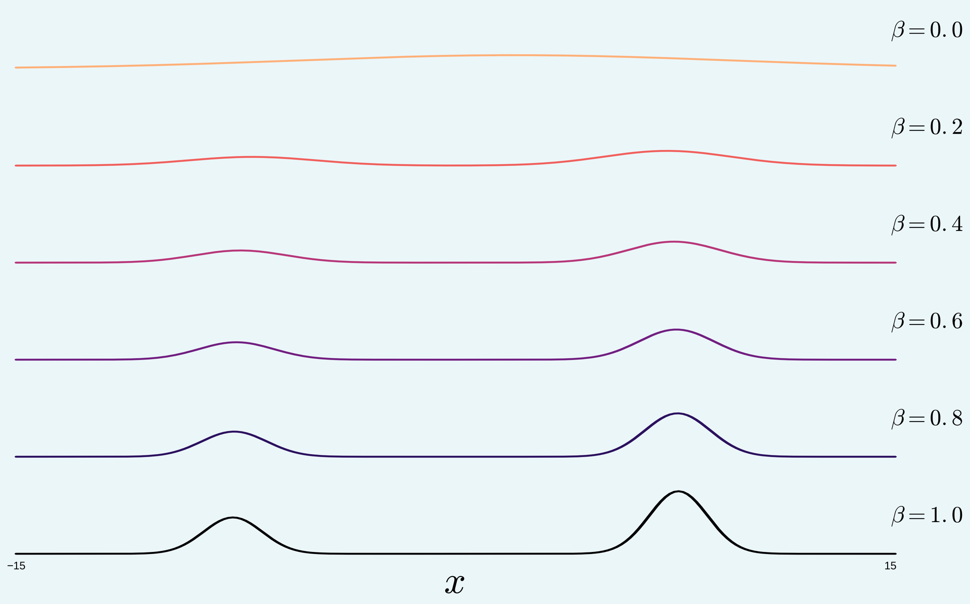
HMC allows long-range moves in high-dimensional $\mathcal{X}$: scalable to large models.

Requires model gradients $\frac{\partial \phi}{\partial \boldsymbol{x}}$: reverse-mode automatic differentiation.

Need to tune step-size δt and number of steps L: No U-Turns Sampler (NUTS)[3].

However copes poorly with multimodal targets and does allow direct estimation of Z.

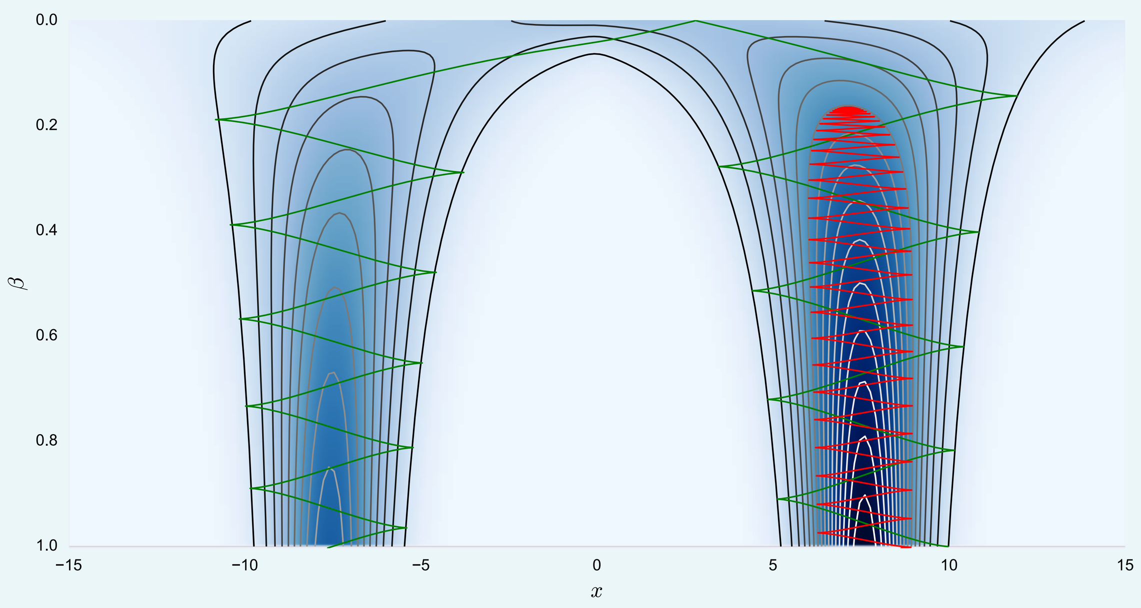
THERMODYNAMIC ENSEMBLES



$$\pi(\boldsymbol{x} \mid \beta) = \frac{1}{\mathcal{Z}(\beta)} \exp\left[-\beta\phi(\boldsymbol{x}) - (1-\beta)\psi(\boldsymbol{x})\right]$$

$$\mathcal{Z}(\beta) = \int_{\mathcal{X}} \exp\left[-\beta\phi(\boldsymbol{x}) - (1-\beta)\psi(\boldsymbol{x})\right] \mathrm{d}\boldsymbol{x}$$

ADIABATIC MONTE CARLO [4]



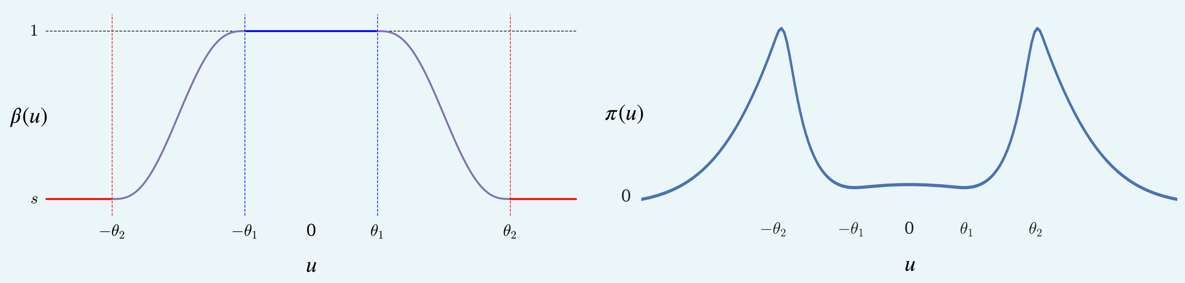
$$\pi(\boldsymbol{x},\, \boldsymbol{p},\, \beta) = \exp\underbrace{\left[-\beta\phi(\boldsymbol{x}) - (1-\beta)\,\psi(\boldsymbol{x}) - \frac{1}{2}\boldsymbol{p}^{\mathrm{T}}\mathbf{M}^{-1}\boldsymbol{p} - \log \mathcal{Z}(\beta)\right]}_{H_c(\boldsymbol{x},\boldsymbol{p},\beta)}$$

EXTENDED HAMILTONIAN APPROACH TO CONTINUOUS TEMPERING [5]

$$\pi(\boldsymbol{x},\, u,\, \boldsymbol{p},\, v) \propto \exp\underbrace{\left\{-\beta(u)\phi(\boldsymbol{x}) - \frac{u^2}{2\sigma^2} - \frac{1}{2}\boldsymbol{p}^{\mathrm{T}}\mathbf{M}^{-1}\boldsymbol{p} - \frac{v^2}{2m}\right\}}_{-H(\boldsymbol{x},u,\boldsymbol{p},v)}$$

$$\pi\left(\boldsymbol{x} \mid -\theta_1 \le |u| \le \theta_1\right) \propto \exp\left[-\phi(\boldsymbol{x})\right]$$

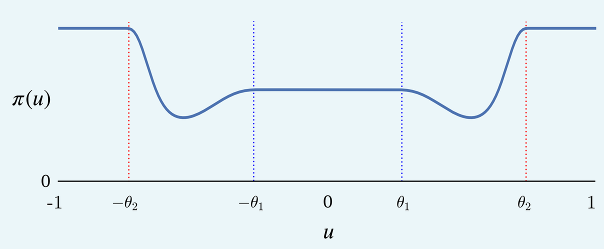
$$\pi(u) \propto \exp\left[-\frac{u^2}{2\sigma^2}\right] \int_{\mathcal{X}} \exp\left[-\beta(u)\phi(\boldsymbol{x})\right] \mathrm{d}\boldsymbol{x}$$



OUR APPROACH

$$\pi(\boldsymbol{x},\, u,\, \boldsymbol{p},\, v) = \exp\underbrace{\left\{-\beta(u)\left[\phi(\boldsymbol{x}) + \log\zeta\right] - \left[1 - \beta(u)\right]\psi(\boldsymbol{x}) - \frac{1}{2}\boldsymbol{p}^{\mathrm{T}}\mathbf{M}^{-1}\boldsymbol{p} - \frac{v^2}{2m}\right\}}_{-H(\boldsymbol{x},u,\boldsymbol{p},v)}$$

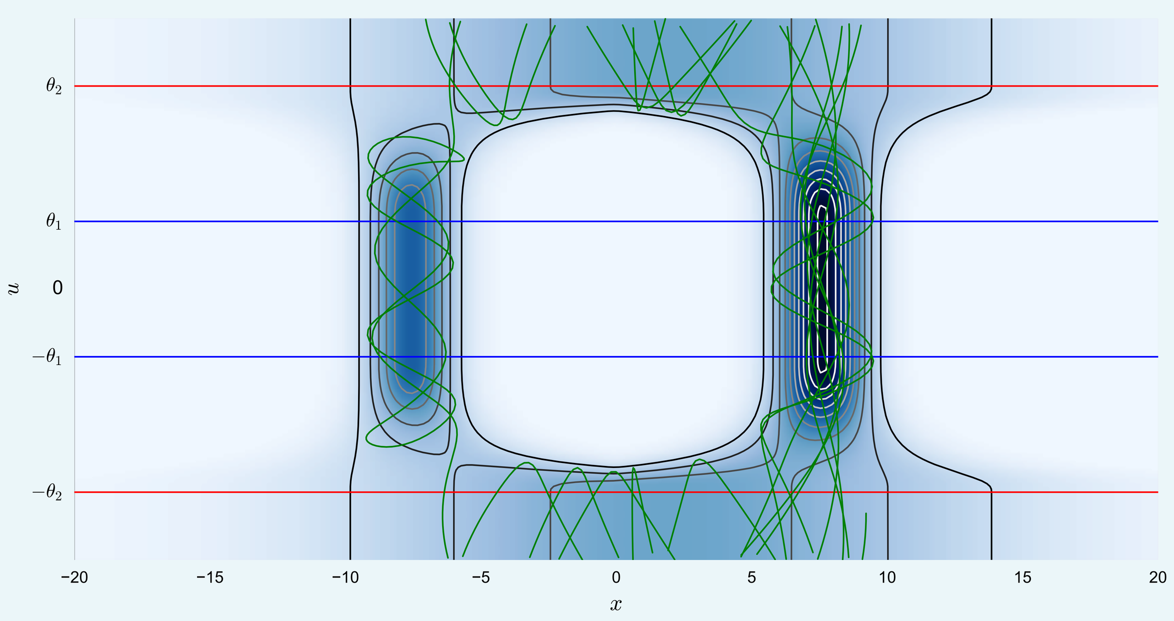
$$\log\zeta \approx \log Z \qquad \exp\left[-\psi(\boldsymbol{x})\right] \overset{\text{moments}}{\approx} \frac{1}{Z}\exp\left[-\phi(\boldsymbol{x})\right]$$



$$\pi(u) \propto \mathscr{Z}\left[\beta(u)\right] = \frac{\mathcal{Z}\left[\beta(u)\right]}{\zeta^{\beta(u)}} \qquad \beta(u)\left[\log\frac{Z}{\zeta} - \mathbb{D}_{\mathrm{KL}}^{b\|t}\right] \le \log\mathscr{Z}\left[\beta(u)\right] \le \beta(u)\log\frac{Z}{\zeta}$$

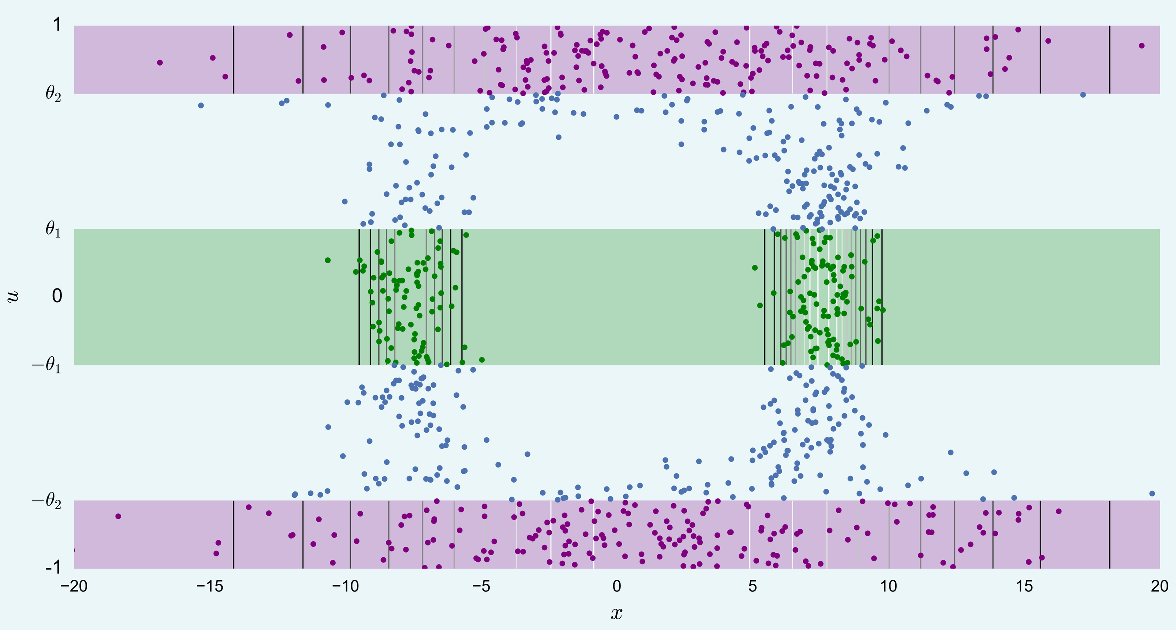
$$\mathbb{D}_{\mathrm{KL}}^{b\|t} = \int_{\mathcal{X}} \exp\left[-\psi(\boldsymbol{x})\right] \log\frac{\exp\left[-\psi(\boldsymbol{x})\right]}{\frac{1}{Z}\exp\left[-\phi(\boldsymbol{x})\right]}\, \mathrm{d}\boldsymbol{x}$$

DYNAMIC IN EXTENDED SYSTEM



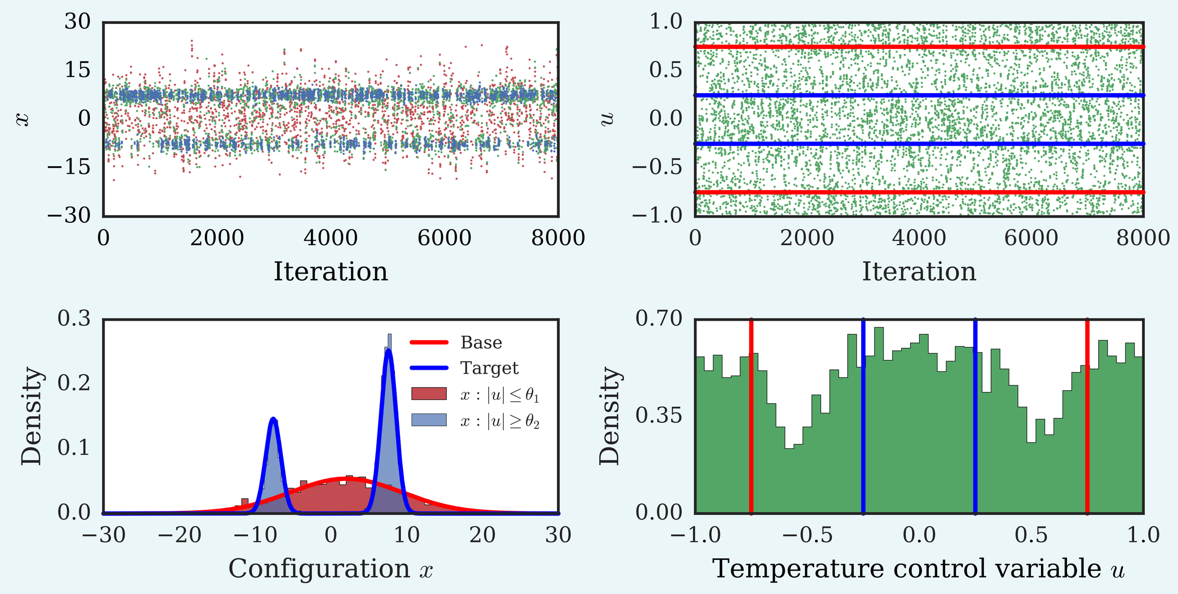
$$H(\underbrace{\boldsymbol{x}, u}_{\tilde{\boldsymbol{x}}},\, \underbrace{\boldsymbol{p}, v}_{\tilde{\boldsymbol{p}}}) = \overbrace{\beta(u)\left[\phi(\boldsymbol{x}) + \log\zeta\right] + \left[1-\beta(u)\right]\psi(\boldsymbol{x})}^{\tilde{\phi}(\tilde{\boldsymbol{x}})} + \underbrace{\frac{1}{2}\boldsymbol{p}^{\mathrm{T}}\mathbf{M}^{-1}\boldsymbol{p} + \frac{v^2}{2m}}_{\frac{1}{2}\tilde{\boldsymbol{p}}^{\mathrm{T}}\tilde{\mathbf{M}}^{-1}\tilde{\boldsymbol{p}}}$$

ESTIMATING Z



$$Z = \frac{1-\theta_2}{\theta_1}\frac{\mathbb{E}_\pi\left[\mathbb{1}[0 \le |u| \le \theta_1]\right]}{\mathbb{E}_\pi\left[\mathbb{1}[\theta_2 \le |u| \le 1]\right]}\zeta \approx \frac{1-\theta_2}{\theta_1}\frac{\sum_{s=1}^{S}\left\{\mathbb{1}\left[0 \le |u^{(s)}| \le \theta_1\right]\right\}}{\sum_{s=1}^{S}\left\{\mathbb{1}\left[\theta_2 \le |u^{(s)}| \le 1\right]\right\}}\zeta$$

1D GAUSSIAN MIXTURE RESULTS



BOLTZMANN MACHINE RELAXATIONS

$$\pi(\boldsymbol{s}) = \frac{1}{Z_B}\exp\left[\frac{1}{2}\boldsymbol{s}^{\mathrm{T}}\mathbf{W}\boldsymbol{s} + \boldsymbol{s}^{\mathrm{T}}\boldsymbol{b}\right]$$

$$\pi(\boldsymbol{x}) = \frac{1}{Z}\exp\left[-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} + \sum_{i=1}^{D_B}\log\cosh\left(\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{x} + b_i\right)\right]$$

$$\mathbf{Q} = \left[\boldsymbol{q}_1\ \boldsymbol{q}_2\ \cdots\ \boldsymbol{q}_{D_B}\right] \qquad \mathbf{W} + \mathbf{D} = \mathbf{Q}^{\mathrm{T}}\mathbf{Q}$$

$$\log Z = \log Z_B + \frac{1}{2}\mathrm{Tr}\left[\mathbf{D}\right] + \frac{D}{2}\log(2\pi) - D_B\log(2)$$

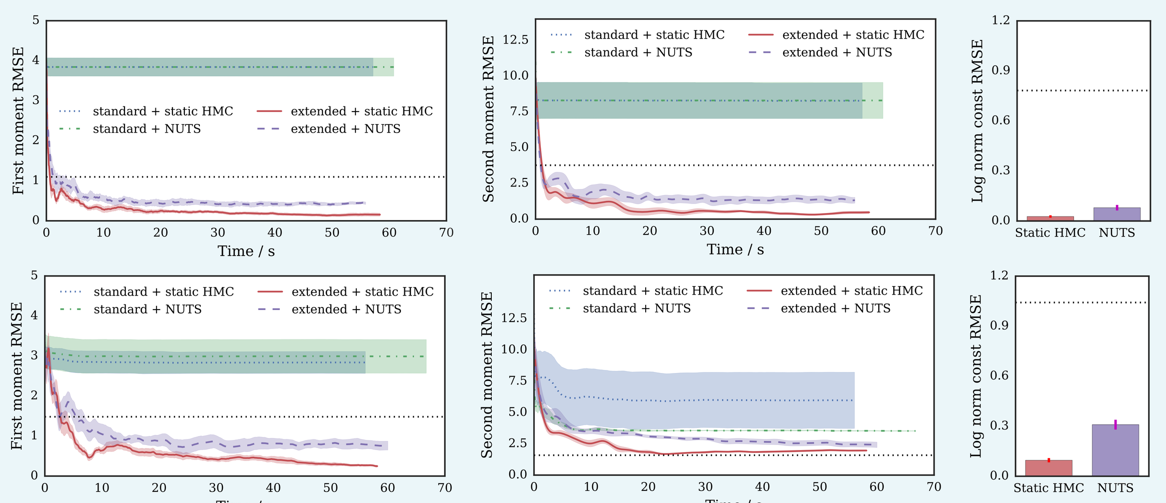
$$\mathbb{E}_\pi\left[\boldsymbol{x}\right] = \mathbf{Q}\mathbb{E}_\pi\left[\boldsymbol{s}\right] \qquad \mathbb{E}_\pi\left[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\right] = \mathbf{Q}\mathbb{E}_\pi\left[\boldsymbol{s}\boldsymbol{s}^{\mathrm{T}}\right]\mathbf{Q}^{\mathrm{T}} + \mathbf{I}$$

19-DIMENSIONAL RELAXATION RESULTS

Random relaxtions generated by sampling $\mathbf{W}$, $\boldsymbol{b}$ to encourage multimodality.

Standard and extended Hamiltonian approaches run in Stan with static HMC and NUTS.

For each parameter and method combination 8 chains were run.



CONCLUSIONS

- Thermodynamic HMC augmentation which improves mode-hopping and allows estimation of Z.
- Once a base density $\exp\left[-\psi(\boldsymbol{x})\right]$ and ζ are chosen can be easily used with existing HMC code.
- Exploits cheap deterministic approximations to $\pi(\boldsymbol{x})$ while still allowing asymptotic exactness.

REFERENCES

1. Hybrid Monte Carlo. *Physics Letters B*, Duane, Kennedy, Pendleton & Roweth (1987).
2. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, Neal (2011).
3. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, Hoffman & Gelman (2014).
4. Adiabatic Monte Carlo. *arXiv preprint arXiv:1405.3489*, Betancourt (2014).
5. Extended Hamiltonian approach to continuous tempering. *Physical Review E*, Gobbo & Leimkuhler (2015).

This work was supported in part by grants EP/F500385/1 and BB/F529254/1 for the University of Edinburgh School of Informatics Doctoral Training Centre in Neuroinformatics and Computational Neuroscience (www.anc.ac.uk/dtc) from the UK Engineering and Physical Sciences Research Council (EPSRC), UK Biotechnology and Biological Sciences Research Council (BBSRC), and the UK Medical Research Council (MRC).