CHALMERS UNIVERSITY OF TECHNOLOGY | UNIVERSITY OF GOTHENBURG

Transport Accelerated Simulated Annealing

Vincent Molin, Axel Ringh, Moritz Schauer, Akash Sharma
Department of Mathematical Sciences

Simulated annealing

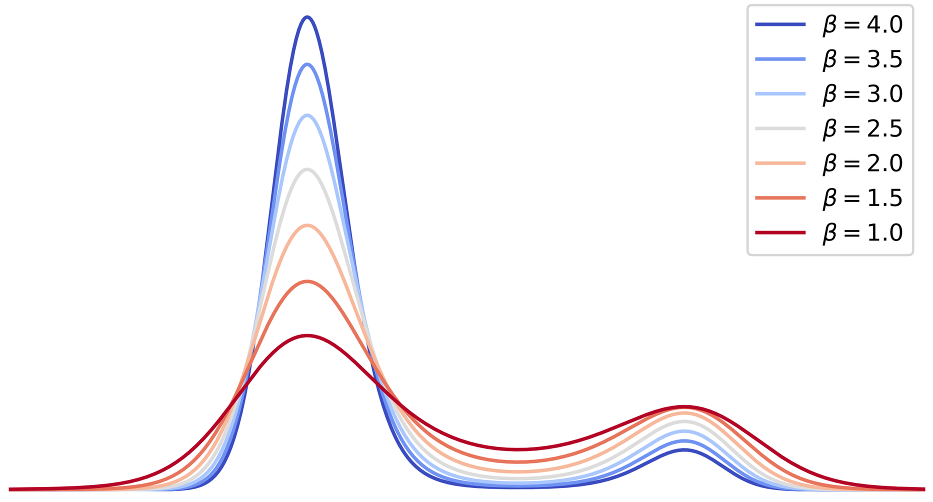
Sampling and optimization are deeply connected, effective algorithms trade-off exploration and exploitation.

Exploration *Sampling* ◁ ⟵ β ⟶ ▷ **Exploitation** *Mode seeking*

For a potential $U : \mathbb{R}^d \to \mathbb{R}$ we consider the optimization problem

$$x^* = \arg\min_{x \in \mathbb{R}^n} U(x)$$

and the *Gibbs measure* $\mu_\beta(\mathrm{d}x) \propto \exp(-\beta U)\,\mathrm{d}x$, where $\mu_\beta \underset{w}{\to} \delta_{x^*}$ as $\beta \to \infty$.



As β increases, probability mass concentrates on the global minima of U

Simulated annealing: Search for x^* by increasing β_t over time and successively sampling from $\mu_t := \mu_{\beta_t}$.

The classical diffusion based simulated annealing process is given by

$$\mathrm{d}X_t = -\nabla U(X_t)\,\mathrm{d}t + \sqrt{\frac{2}{\beta_t}}\,\mathrm{d}W_t$$

Converges to μ_∞ if β_t grows *at most logarithmically* in t.

A greedy Bouncy Particle Sampler

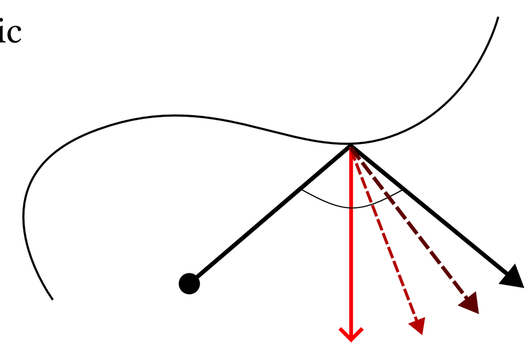
A Markov process $(X_t, W_t) \in \mathbb{R}^d \times \mathbb{R}^d$ with piecewise deterministic trajectories governed by

$$\mathrm{d}(X_t, W_t) = (W_t, 0)\,\mathrm{d}t,$$

until a random bounce event occurs, with

$$P(\text{Bounce} \in [\tau, \tau + h]) = \min(0, \langle W_\tau, \nabla U(X_\tau)\rangle)h + o(h)$$

after which the velocity is reflected on a contour line of the potential.



BPS bouncing on a contour line, greedier in red

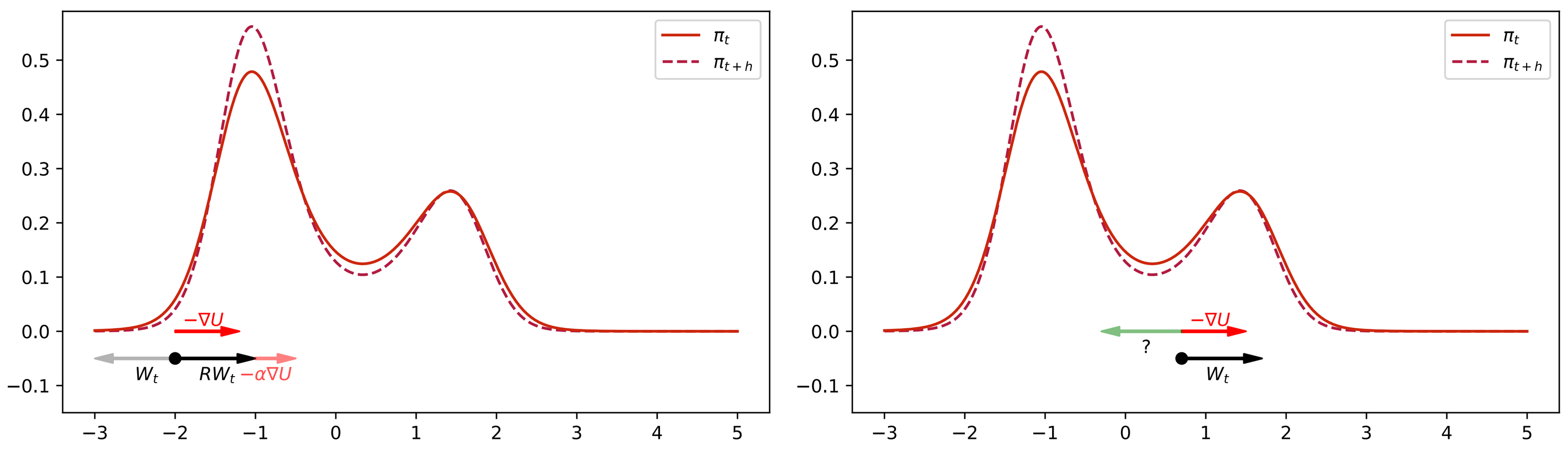
The velocity after reflection is given by

$$\begin{aligned} RW_t &= W_t - \frac{2\langle W_t, \nabla U(X_t)\rangle}{\langle \nabla U(X_t), \nabla U(X_t)\rangle}\nabla U(X_t) \\ &= W_t - 2\mathrm{Proj}_{\nabla U(X_t)} W_t \end{aligned}$$

Previously we have considered a *greedier* reflection kernel

$$R_\alpha W_t = RW_t - \alpha \nabla U(X_t)$$

Ideally α should capture the *flow* of probability mass as the temperature changes.



The continuity equation

For any Markov process targeting μ_β, with generator $\mathcal{A}_\beta$ and initial distribution ρ_0, the continuity equation for the evolution of the distribution ρ_t is given by the Kolmogorov forward equation

$$\mathcal{A}^*_\beta \rho_t = \partial_t \rho_t$$

For the simulated annealing process, think of $\mathcal{A}_t := \mathcal{A}_{\beta(t)}$ as a time inhomogeneous generator for this process. By stationarity of $\mathcal{A}$, we have that

$$\mathcal{A}^*_t \mu_t = 0 \neq \partial_t \mu_t$$

Following the Gibbs curve

Can we modify a Markov process to follow the mass flow prescribed by the Gibbs curve?

Specifically, can we find a deterministic velocity field v_t such that

$$\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0?$$

Denote by $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ the metric space of probability measures on $\mathbb{R}^d$ equipped with the Wasserstein-2 metric

View $t \mapsto \mu_t$ as a *curve* in this metric space. If this curve is sufficiently well-behaved the we can define a *tangent-like* object v_t, where we think of "$\mu'_t = v_t$" as the flow of probability mass.

Theorem

Under suitable conditions on U, the *curve* $t \mapsto \mu_t$ is an *absolutely continuous curve* in the metric space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$. Hence there exists a time dependent velocity field v_t such that

$$\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0.$$

Moreover, the minimal v_t is the derivative of the optimal transport maps from μ_t to $\mu_t + h$, in the sense that

$$v_t(x) = \lim_{h \to 0^+} h^{-1}\left(T_{\mu_t \to \mu_{t+h}}(x) - \mathrm{Id}(x)\right).$$

...with superimposed stochastic dynamics

Now, we combine

- The velocity field v_t drives the distribution along the curve μ_t
- The stochastic dynamics $\mathcal{A}$ mixes any distribution towards μ_t

Illustratively, we think of a process with generator

$$\mathcal{A}^o f = \langle v, \nabla f \rangle + \mathcal{A} f.$$

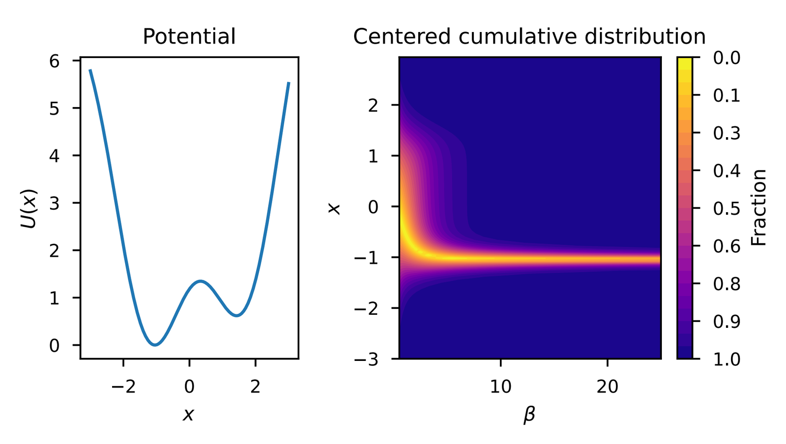
A particle approximation scheme

Problem: The velocity field v_t is in general not available in closed form.

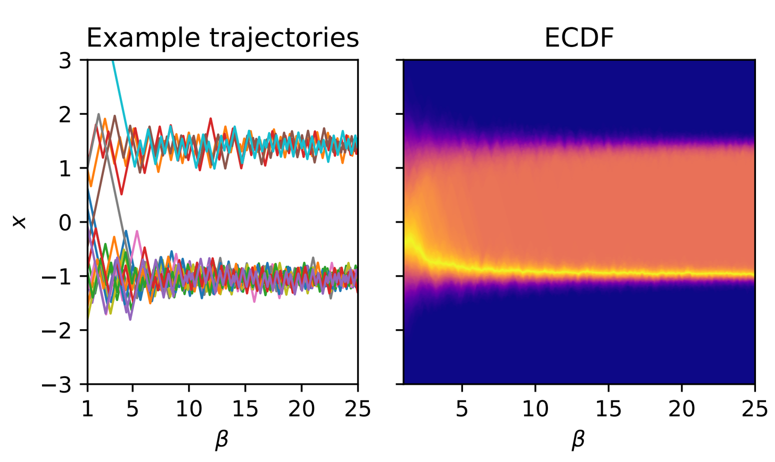
We sketch out a simple **quasi-independent particle scheme** inspired by the true dynamics. In short:

- Simulate a population of particles according to *independent* stochastic dynamics $\mathcal{A}_t$ and an approximative velocity field $\hat{V}$.
- At finite intervals $h > 0$, create empirical estimators $\hat{\mu}_t$ and $\hat{\mu}_{t+h}$, the latter by reweighting.
- Estimate a transport map $\hat{T}_t$ by solving the transport problem from $\hat{\mu}_t$ to $\hat{\mu}_{t+h}$ and update $\hat{V}$ accordingly.

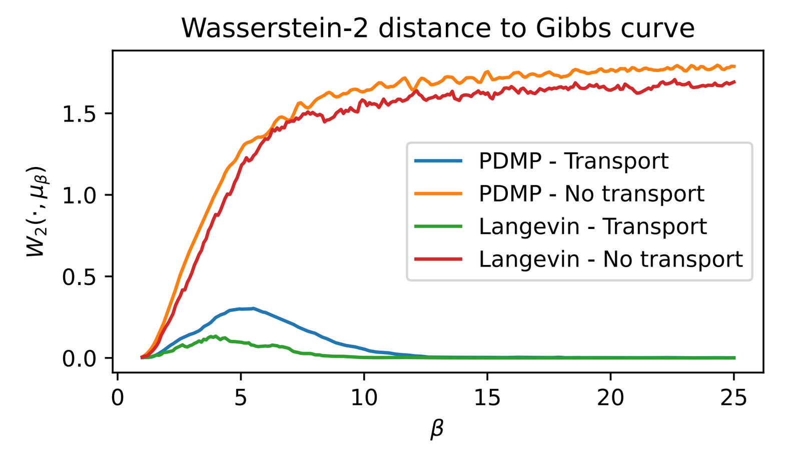
Simulation on a double well potential on $\mathbb{R}$ with linear cooling:



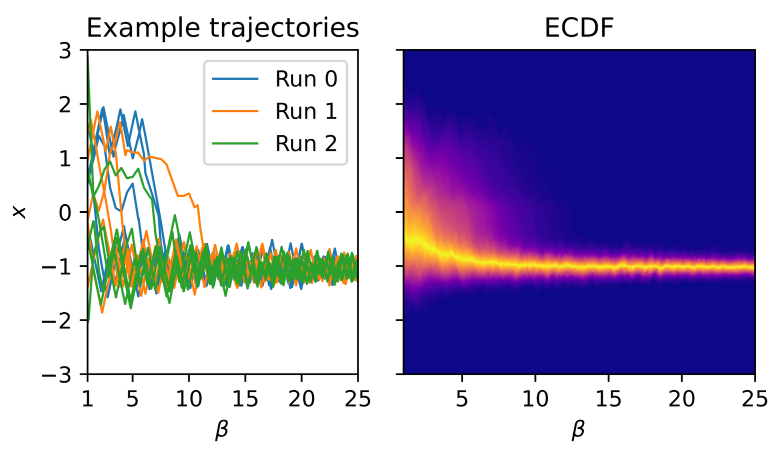
A double well potential



Without v-estimation



W_2-distance to the Gibbs curve



With v-estimation, 5 particles

Contact: molinv@chalmers.se
Made with typst

WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM