# Gibbs flow transport for Bayesian inference

Jeremy Heng,
Department of Statistics, Harvard University

Joint work with Arnaud Doucet (Oxford) & Yvo Pokern (UCL)

Computational Statistics and Molecular Simulation:
A Practical Cross-Fertilization
Casa Matemática Oaxaca (CMO)
13 November 2018

# Problem specification

- **Target** distribution on $\mathbb{R}^d$

$$\pi(dx) = \frac{\gamma(x)\,dx}{Z}$$

where $\gamma : \mathbb{R}^d \to \mathbb{R}_+$ can be evaluated pointwise and

$$Z = \int_{\mathbb{R}^d} \gamma(x)\,dx$$

is unknown

# Problem specification

- **Target** distribution on $\mathbb{R}^d$

$$\pi(dx) = \frac{\gamma(x)\,dx}{Z}$$

where $\gamma : \mathbb{R}^d \to \mathbb{R}_+$ can be evaluated pointwise and

$$Z = \int_{\mathbb{R}^d} \gamma(x)\,dx$$

is unknown

- **Problem 1**: Obtain consistent estimator of $\pi(\varphi) := \int_{\mathbb{R}^d} \varphi(x)\,\pi(dx)$

# Problem specification

- **Target** distribution on $\mathbb{R}^d$

$$\pi(dx) = \frac{\gamma(x)\,dx}{Z}$$

where $\gamma : \mathbb{R}^d \to \mathbb{R}_+$ can be evaluated pointwise and

$$Z = \int_{\mathbb{R}^d} \gamma(x)\,dx$$

is unknown

- **Problem 1**: Obtain consistent estimator of $\pi(\varphi) := \int_{\mathbb{R}^d} \varphi(x)\,\pi(dx)$

- **Problem 2**: Obtain unbiased and consistent estimator of $Z$

# Motivation: Bayesian computation

- **Prior** distribution $\pi_0$ on unknown parameters of a model

# Motivation: Bayesian computation

- **Prior** distribution $\pi_0$ on unknown parameters of a model
- **Likelihood** function $L : \mathbb{R}^d \to \mathbb{R}_+$ of data $y$

# Motivation: Bayesian computation

- **Prior** distribution $\pi_0$ on unknown parameters of a model
- **Likelihood** function $L : \mathbb{R}^d \to \mathbb{R}_+$ of data $y$
- Bayes update gives **posterior** distribution on $\mathbb{R}^d$

$$\pi(dx) = \frac{\pi_0(x)L(x)\,dx}{Z},$$

where $Z = \int_{\mathbb{R}^d} \pi_0(x)L(x)\,dx$ is the **marginal likelihood** of $y$

# Motivation: Bayesian computation

- **Prior** distribution $\pi_0$ on unknown parameters of a model
- **Likelihood** function $L : \mathbb{R}^d \to \mathbb{R}_+$ of data $y$
- Bayes update gives **posterior** distribution on $\mathbb{R}^d$

$$\pi(dx) = \frac{\pi_0(x)L(x)\,dx}{Z},$$

where $Z = \int_{\mathbb{R}^d} \pi_0(x)L(x)\,dx$ is the **marginal likelihood** of $y$

- $\pi(\varphi)$ and $Z$ are typically intractable

# Monte Carlo methods

- Typically sampling from $\pi$ is intractable, so we rely on **Markov chain Monte Carlo** (MCMC) methods

# Monte Carlo methods

- Typically sampling from $\pi$ is intractable, so we rely on **Markov chain Monte Carlo** (MCMC) methods

- MCMC constructs a $\pi$-**invariant Markov kernel** $K : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \to [0, 1]$

# Monte Carlo methods

- Typically sampling from $\pi$ is intractable, so we rely on **Markov chain Monte Carlo** (MCMC) methods

- MCMC constructs a $\pi$-**invariant Markov kernel** $K : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \to [0,1]$

- Sample $X_0 \sim \pi_0$ and iterate $X_n \sim K(X_{n-1}, \cdot)$ until convergence

# Monte Carlo methods

- Typically sampling from $\pi$ is intractable, so we rely on **Markov chain Monte Carlo** (MCMC) methods
- MCMC constructs a $\pi$-**invariant Markov kernel** $K : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \to [0, 1]$
- Sample $X_0 \sim \pi_0$ and iterate $X_n \sim K(X_{n-1}, \cdot)$ until convergence
- MCMC can fail in practice, for e.g. when $\pi$ is highly multi-modal

# Annealed importance sampling

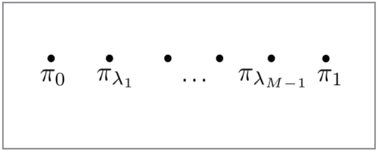
- If $\pi_0$ and $\pi$ are distant,

# Annealed importance sampling

- If $\pi_0$ and $\pi$ are distant, define **bridges**

$$\pi_{\lambda_m}(dx) = \frac{\pi_0(x)L(x)^{\lambda_m}\,dx}{Z(\lambda_m)},$$

with $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_M = 1$ so that $\pi_1 = \pi$

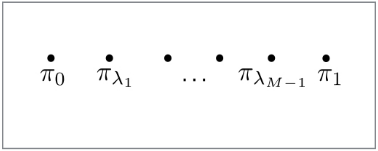


# Annealed importance sampling

- If $\pi_0$ and $\pi$ are distant, define **bridges**

$$\pi_{\lambda_m}(dx) = \frac{\pi_0(x) L(x)^{\lambda_m}\, dx}{Z(\lambda_m)},$$

with $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_M = 1$ so that $\pi_1 = \pi$



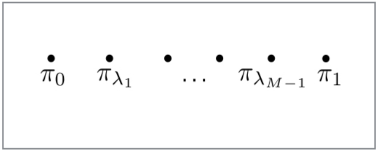

  - Initialize $X_0 \sim \pi_0$ and move $X_m \sim K_m(X_{m-1}, \cdot)$ for $m = 1, \ldots, M$, where $K_m$ is $\pi_{\lambda_m}$-**invariant**

# Annealed importance sampling

- If $\pi_0$ and $\pi$ are distant, define **bridges**

$$\pi_{\lambda_m}(dx) = \frac{\pi_0(x)L(x)^{\lambda_m}\,dx}{Z(\lambda_m)},$$

with $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_M = 1$ so that $\pi_1 = \pi$




  - Initialize $X_0 \sim \pi_0$ and move $X_m \sim K_m(X_{m-1}, \cdot)$ for $m = 1, \ldots, M$, where $K_m$ is $\pi_{\lambda_m}$-**invariant**

  - **Annealed importance sampling** constructs $w : (\mathbb{R}^d)^{M+1} \to \mathbb{R}_+$ so that

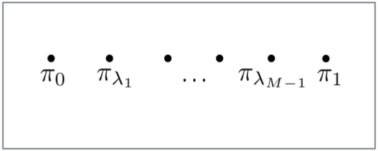
$$\pi(\varphi) = \frac{\mathbb{E}\left[\varphi(X_M)w(X_{0:M})\right]}{\mathbb{E}\left[w(X_{0:M})\right]}, \quad Z = \mathbb{E}\left[w(X_{0:M})\right]$$

# Annealed importance sampling

- If $\pi_0$ and $\pi$ are distant, define **bridges**

$$\pi_{\lambda_m}(dx) = \frac{\pi_0(x)L(x)^{\lambda_m}\,dx}{Z(\lambda_m)},$$

with $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_M = 1$ so that $\pi_1 = \pi$




  - Initialize $X_0 \sim \pi_0$ and move $X_m \sim K_m(X_{m-1}, \cdot)$ for $m = 1, \ldots, M$, where $K_m$ is $\pi_{\lambda_m}$-**invariant**

  - **Annealed importance sampling** constructs $w : (\mathbb{R}^d)^{M+1} \to \mathbb{R}_+$ so that

$$\pi(\varphi) = \frac{\mathbb{E}\left[\varphi(X_M)w(X_{0:M})\right]}{\mathbb{E}\left[w(X_{0:M})\right]}, \quad Z = \mathbb{E}\left[w(X_{0:M})\right]$$

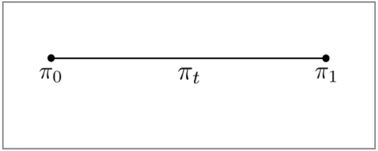
  - AIS (Neal, 2001) and SMC samplers (Del Moral et al., 2006) are considered state-of-the-art in statistics and machine learning

# Jarzynski nonequilibrium equality

- Consider $M \to \infty$,

# Jarzynski nonequilibrium equality

- Consider $M \to \infty$, i.e. define the curve of distribution $\{\pi_t\}_{t\in[0,1]}$

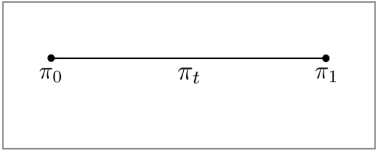


# Jarzynski nonequilibrium equality

- Consider $M \to \infty$, i.e. define the curve of distribution $\{\pi_t\}_{t \in [0,1]}$

$$\pi_t(dx) = \frac{\pi_0(x) L(x)^{\lambda(t)}\, dx}{Z(t)},$$

where $\lambda : [0,1] \to [0,1]$ is a strictly increasing $C^1$ function

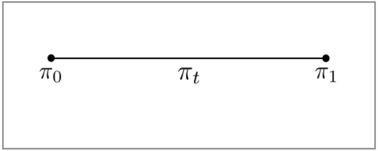


# Jarzynski nonequilibrium equality

- Consider $M \to \infty$, i.e. define the curve of distribution $\{\pi_t\}_{t\in[0,1]}$

$$\pi_t(dx) = \frac{\pi_0(x)L(x)^{\lambda(t)}\,dx}{Z(t)},$$

where $\lambda : [0,1] \to [0,1]$ is a strictly increasing $C^1$ function




  - Initialize $X_0 \sim \pi_0$ and run **time-inhomogenous Langevin dynamics**

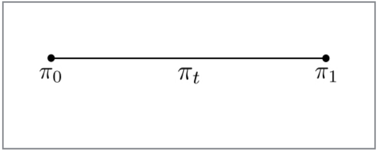
$$dX_t = \frac{1}{2}\nabla \log \pi_t(X_t)dt + dW_t, \quad t \in [0,1]$$

# Jarzynski nonequilibrium equality

- Consider $M \to \infty$, i.e. define the curve of distribution $\{\pi_t\}_{t\in[0,1]}$

$$\pi_t(dx) = \frac{\pi_0(x)L(x)^{\lambda(t)}\,dx}{Z(t)},$$

where $\lambda : [0,1] \to [0,1]$ is a strictly increasing $C^1$ function




- Initialize $X_0 \sim \pi_0$ and run **time-inhomogenous Langevin dynamics**

$$dX_t = \frac{1}{2}\nabla \log \pi_t(X_t)dt + dW_t, \quad t \in [0,1]$$

- **Jarzynski equality** (Jarzynski, 1997; Crooks, 1998) constructs $w : C([0,1],\mathbb{R}^d) \to \mathbb{R}_+$ so that

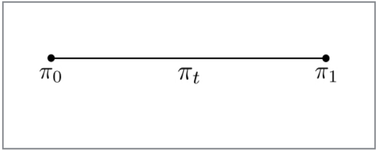
$$\pi(\varphi) = \frac{\mathbb{E}\left[\varphi(X_1)w(X_{[0,1]})\right]}{\mathbb{E}\left[w(X_{[0,1]})\right]}, \quad Z = \mathbb{E}\left[w(X_{[0,1]})\right]$$

# Jarzynski nonequilibrium equality

- Consider $M \to \infty$, i.e. define the curve of distribution $\{\pi_t\}_{t\in[0,1]}$

$$\pi_t(dx) = \frac{\pi_0(x)L(x)^{\lambda(t)}\, dx}{Z(t)},$$

where $\lambda : [0,1] \to [0,1]$ is a strictly increasing $C^1$ function




- Initialize $X_0 \sim \pi_0$ and run **time-inhomogenous Langevin dynamics**

$$dX_t = \frac{1}{2}\nabla \log \pi_t(X_t)dt + dW_t, \quad t \in [0,1]$$

- **Jarzynski equality** (Jarzynski, 1997; Crooks, 1998) constructs $w : C([0,1],\mathbb{R}^d) \to \mathbb{R}_+$ so that

$$\pi(\varphi) = \frac{\mathbb{E}\left[\varphi(X_1)w(X_{[0,1]})\right]}{\mathbb{E}\left[w(X_{[0,1]})\right]}, \quad Z = \mathbb{E}\left[w(X_{[0,1]})\right]$$

- To what extent is this state-of-the-art in molecular dynamics?

# Optimal dynamics

- **Dynamical lag** $\|\mathrm{Law}(X_t) - \pi_t\|$ impacts variance of estimators

# Optimal dynamics

- **Dynamical lag** $\|\mathrm{Law}(X_t) - \pi_t\|$ impacts variance of estimators
- Vaikuntanathan & Jarzynski (2011) considered **adding drift** $f : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ to reduce lag

$$dX_t = f(t, X_t)dt + \frac{1}{2}\nabla \log \pi_t(X_t)dt + dW_t, \quad t \in [0,1], X_0 \sim \pi_0$$

# Optimal dynamics

- **Dynamical lag** $\|\mathrm{Law}(X_t) - \pi_t\|$ impacts variance of estimators
- Vaikuntanathan & Jarzynski (2011) considered **adding drift** $f : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ to reduce lag

$$dX_t = f(t, X_t)dt + \frac{1}{2}\nabla \log \pi_t(X_t)dt + dW_t, \quad t \in [0,1], X_0 \sim \pi_0$$

- An optimal choice of $f$ results in **zero lag**, i.e. $X_t \sim \pi_t$ for $t \in [0,1]$, and **zero variance** estimator of $Z$

# Optimal dynamics

- **Dynamical lag** $\|\mathrm{Law}(X_t) - \pi_t\|$ impacts variance of estimators
- Vaikuntanathan & Jarzynski (2011) considered **adding drift** $f : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ to reduce lag

$$dX_t = f(t, X_t)dt + \frac{1}{2}\nabla \log \pi_t(X_t)dt + dW_t, \quad t \in [0,1], X_0 \sim \pi_0$$

- An optimal choice of $f$ results in **zero lag**, i.e. $X_t \sim \pi_t$ for $t \in [0,1]$, and **zero variance** estimator of $Z$
- Any optimal choice $f$ satisfies **Liouville PDE**

$$-\nabla \cdot (\pi_t(x) f(t,x)) = \partial_t \pi_t(x)$$

# Optimal dynamics

- **Dynamical lag** $\|\mathrm{Law}(X_t) - \pi_t\|$ impacts variance of estimators
- Vaikuntanathan & Jarzynski (2011) considered **adding drift** $f : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ to reduce lag

$$dX_t = f(t, X_t)dt + \frac{1}{2}\nabla \log \pi_t(X_t)dt + dW_t, \quad t \in [0,1], X_0 \sim \pi_0$$

- An optimal choice of $f$ results in **zero lag**, i.e. $X_t \sim \pi_t$ for $t \in [0,1]$, and **zero variance** estimator of $Z$
- Any optimal choice $f$ satisfies **Liouville PDE**

$$-\nabla \cdot (\pi_t(x)f(t,x)) = \partial_t \pi_t(x)$$

- Zero lag also achieved by running **deterministic dynamics**

$$dX_t = f(t, X_t)dt, \quad X_0 \sim \pi_0$$

# Time evolution of distributions

- Time evolution of $\pi_t$ is given by

$$\partial_t \pi_t(x) = \lambda'(t) \left(\log L(x) - I_t\right) \pi_t(x),$$

where

$$I_t = \frac{1}{\lambda'(t)} \frac{d}{dt} \log Z(t) \stackrel{!}{=} \mathbb{E}_{\pi_t}[\log L(X_t)] < \infty$$

# Time evolution of distributions

- Time evolution of $\pi_t$ is given by

$$\partial_t \pi_t(x) = \lambda'(t)\left(\log L(x) - I_t\right)\pi_t(x),$$

where

$$I_t = \frac{1}{\lambda'(t)}\frac{d}{dt}\log Z(t) \overset{!}{=} \mathbb{E}_{\pi_t}[\log L(X_t)] < \infty$$

- Integrating recovers **path sampling** (Gelman and Meng, 1998) or **thermodynamic integration** (Kirkwood, 1935) identity

$$\log\left(\frac{Z(1)}{Z(0)}\right) = \int_0^1 \lambda'(t) I_t \, dt.$$

# Liouville PDE

- Dynamics governed by ODE

$$dX_t = f(t, X_t)dt, \quad X_0 \sim \pi_0$$

# Liouville PDE

- Dynamics governed by ODE

$$dX_t = f(t, X_t)dt, \quad X_0 \sim \pi_0$$

- For sufficiently regular $f$, ODE admits a unique solution

$$t \mapsto X_t, \quad t \in [0, 1]$$

# Liouville PDE

- Dynamics governed by ODE

$$dX_t = f(t, X_t)dt, \quad X_0 \sim \pi_0$$

- For sufficiently regular $f$, ODE admits a unique solution

$$t \mapsto X_t, \quad t \in [0, 1]$$

# Liouville PDE

- Dynamics governed by ODE

$$dX_t = f(t, X_t)dt, \quad X_0 \sim \pi_0$$

- For sufficiently regular $f$, ODE admits a unique solution

$$t \mapsto X_t, \quad t \in [0, 1]$$

inducing a curve of distributions

$$\{\tilde{\pi}_t := \mathrm{Law}(X_t),\ t \in [0, 1]\}$$

# Liouville PDE

- Dynamics governed by ODE

$$dX_t = f(t, X_t)dt, \quad X_0 \sim \pi_0$$

- For sufficiently regular $f$, ODE admits a unique solution

$$t \mapsto X_t, \quad t \in [0, 1]$$

inducing a curve of distributions

$$\{\tilde{\pi}_t := \mathrm{Law}(X_t),\ t \in [0, 1]\}$$

satisfying Liouville PDE

$$-\nabla \cdot (\tilde{\pi}_t f) = \partial_t \tilde{\pi}_t, \quad \tilde{\pi}_0 = \pi_0$$

# Defining the flow transport problem

- Set $\tilde{\pi}_t = \pi_t$, for $t \in [0, 1]$ and solve Liouville equation

$$-\nabla \cdot (\pi_t f) = \partial_t \pi_t, \quad (L)$$

for a drift $f$... but not all solutions will work!

# Defining the flow transport problem

- Set $\tilde{\pi}_t = \pi_t$, for $t \in [0, 1]$ and solve Liouville equation

$$-\nabla \cdot (\pi_t f) = \partial_t \pi_t, \quad (L)$$

for a drift $f$... but not all solutions will work!

- Validity relies on following result:

# Defining the flow transport problem

- Set $\tilde{\pi}_t = \pi_t$, for $t \in [0,1]$ and solve Liouville equation

$$-\nabla \cdot (\pi_t f) = \partial_t \pi_t, \quad (L)$$

for a drift $f$... but not all solutions will work!

- Validity relies on following result:

**Theorem. Ambrosio et al. (2005)**

Under the following assumptions:

Eulerian Liouville PDE $\Longleftrightarrow$ Lagrangian ODE

# Defining the flow transport problem

- Set $\tilde{\pi}_t = \pi_t$, for $t \in [0,1]$ and solve Liouville equation

$$-\nabla \cdot (\pi_t f) = \partial_t \pi_t, \quad (L)$$

for a drift $f$... but not all solutions will work!

- Validity relies on following result:

**Theorem. Ambrosio et al. (2005)**

Under the following assumptions:

A1 $f$ is locally Lipschitz;

Eulerian Liouville PDE $\Longleftrightarrow$ Lagrangian ODE

# Defining the flow transport problem

- Set $\tilde{\pi}_t = \pi_t$, for $t \in [0, 1]$ and solve Liouville equation

$$-\nabla \cdot (\pi_t f) = \partial_t \pi_t, \quad (L)$$

for a drift $f$... but not all solutions will work!

- Validity relies on following result:

**Theorem. Ambrosio et al. (2005)**

Under the following assumptions:

A1 $f$ is locally Lipschitz;

A2 $\int_0^1 \int_{\mathbb{R}^d} |f(t,x)| \pi_t(x)\, dx\, dt < \infty$;

Eulerian Liouville PDE $\Longleftrightarrow$ Lagrangian ODE

# Defining the flow transport problem

- Set $\tilde{\pi}_t = \pi_t$, for $t \in [0,1]$ and solve Liouville equation

$$-\nabla \cdot (\pi_t f) = \partial_t \pi_t, \quad (L)$$

for a drift $f$... but not all solutions will work!

- Validity relies on following result:

**Theorem. Ambrosio et al. (2005)**

Under the following assumptions:

A1 $f$ is locally Lipschitz;

A2 $\int_0^1 \int_{\mathbb{R}^d} |f(t,x)| \pi_t(x) \, dx \, dt < \infty$;

Eulerian Liouville PDE $\Longleftrightarrow$ Lagrangian ODE

- Define **flow transport problem** as solving Liouville $(L)$ for $f$ that satisfies [A1] & [A2]

# Ill-posedness and regularization

- **Under-determined**: consider $\pi_t = \mathcal{N}((0,0), I_2)$ for $t \in [0,1]$,

$$f(x_1, x_2) = (0,0) \text{ and } f(x_1, x_2) = (-x_2, x_1)$$

are both solutions

# Ill-posedness and regularization

- **Under-determined**: consider $\pi_t = \mathcal{N}((0,0), I_2)$ for $t \in [0,1]$,

$$f(x_1, x_2) = (0,0) \text{ and } f(x_1, x_2) = (-x_2, x_1)$$

  are both solutions

- Regularization: seek **minimal kinetic energy** solution

$$\operatorname{argmin}_f \left\{ \int_0^1 \int_{\mathbb{R}^d} |f(t,x)|^2 \pi_t(x)\, dx\, dt : f \text{ solves Liouville} \right\}$$

$$\overset{EL}{\Longrightarrow} f^* = \nabla \varphi \text{ where } -\nabla \cdot (\pi_t \nabla \varphi) = \partial_t \pi_t$$

# Ill-posedness and regularization

- **Under-determined**: consider $\pi_t = \mathcal{N}((0,0), I_2)$ for $t \in [0,1]$,

$$f(x_1, x_2) = (0,0) \text{ and } f(x_1, x_2) = (-x_2, x_1)$$

  are both solutions
- Regularization: seek **minimal kinetic energy** solution

$$\text{argmin}_f \left\{ \int_0^1 \int_{\mathbb{R}^d} |f(t,x)|^2 \pi_t(x)\, dx\, dt : f \text{ solves Liouville} \right\}$$

$$\overset{EL}{\Longrightarrow} f^* = \nabla\varphi \text{ where } -\nabla \cdot (\pi_t \nabla\varphi) = \partial_t \pi_t$$

- Analytical solution available when distributions are (mixtures of) Gaussians (Reich, 2012)

# Flow transport problem on $\mathbb{R}$

- Minimal kinetic energy solution

$$f(t,x) = \frac{-\int_{-\infty}^{x} \partial_t \pi_t(u)\, du}{\pi_t(x)}$$

# Flow transport problem on $\mathbb{R}$

- Minimal kinetic energy solution

$$f(t,x) = \frac{-\int_{-\infty}^{x} \partial_t \pi_t(u)\, du}{\pi_t(x)}$$

- Checking **Liouville**

$$-\nabla \cdot (\pi_t f) = \partial_x \int_{-\infty}^{x} \partial_t \pi_t(u)\, du = \partial_t \pi_t(x)$$

# Flow transport problem on $\mathbb{R}$

- Minimal kinetic energy solution

$$f(t,x) = \frac{-\int_{-\infty}^{x} \partial_t \pi_t(u)\, du}{\pi_t(x)}$$

- Checking **Liouville**

$$-\nabla \cdot (\pi_t f) = \partial_x \int_{-\infty}^{x} \partial_t \pi_t(u)\, du = \partial_t \pi_t(x)$$

A1 For $f$ to be **locally Lipschitz**, assume

$$\pi_0, L \in C^1(\mathbb{R}, \mathbb{R}_+) \Longrightarrow f \in C^1([0,1] \times \mathbb{R}, \mathbb{R})$$

# Flow transport problem on $\mathbb{R}$

- Minimal kinetic energy solution

$$f(t,x) = \frac{-\int_{-\infty}^{x} \partial_t \pi_t(u)\,du}{\pi_t(x)}$$

- Checking **Liouville**

$$-\nabla \cdot (\pi_t f) = \partial_x \int_{-\infty}^{x} \partial_t \pi_t(u)\,du = \partial_t \pi_t(x)$$

A1 For $f$ to be **locally Lipschitz**, assume

$$\pi_0, L \in C^1(\mathbb{R}, \mathbb{R}_+) \Longrightarrow f \in C^1([0,1] \times \mathbb{R}, \mathbb{R})$$

A2 For **integrability** of $\int_0^1 \int_{\mathbb{R}^d} |f(t,x)| \pi_t(x)\,dx\,dt < \infty$, necessarily

$$|\pi_t f|(t,x) = \left| \int_{-\infty}^{x} \partial_t \pi_t(u)\,du \right| \to 0 \text{ as } |x| \to \infty$$

since $\int_{-\infty}^{\infty} \partial_t \pi_t(u)\,du = 0$

# Flow transport problem on $\mathbb{R}$

- Minimal kinetic energy solution

$$f(t,x) = \frac{-\int_{-\infty}^{x} \partial_t \pi_t(u)\, du}{\pi_t(x)}$$

- Checking **Liouville**

$$-\nabla \cdot (\pi_t f) = \partial_x \int_{-\infty}^{x} \partial_t \pi_t(u)\, du = \partial_t \pi_t(x)$$

A1 For $f$ to be **locally Lipschitz**, assume

$$\pi_0, L \in C^1(\mathbb{R}, \mathbb{R}_+) \Longrightarrow f \in C^1([0,1] \times \mathbb{R}, \mathbb{R})$$

A2 For **integrability** of $\int_0^1 \int_{\mathbb{R}^d} |f(t,x)| \pi_t(x)\, dx\, dt < \infty$, necessarily

$$|\pi_t f|(t,x) = \left| \int_{-\infty}^{x} \partial_t \pi_t(u)\, du \right| \to 0 \text{ as } |x| \to \infty$$

since $\int_{-\infty}^{\infty} \partial_t \pi_t(u)\, du = 0$

- **Optimality**: $f = \nabla \varphi$ holds trivially

# Flow transport problem on $\mathbb{R}$

- Re-write solution as

$$f(t,x) = \frac{\lambda'(t) I_t \left\{F_t(x) - I_t^x / I_t\right\}}{\pi_t(x)}$$

where $I_t^x = \mathbb{E}_{\pi_t}[\mathbb{1}_{(-\infty,x]} \log L(X_t)]$ and $F_t$ is CDF of $\pi_t$

# Flow transport problem on $\mathbb{R}$

- Re-write solution as

$$f(t,x) = \frac{\lambda'(t) I_t \left\{ F_t(x) - I_t^x / I_t \right\}}{\pi_t(x)}$$

where $I_t^x = \mathbb{E}_{\pi_t}[\mathbb{1}_{(-\infty,x]} \log L(X_t)]$ and $F_t$ is CDF of $\pi_t$

- **Speed** is controlled by $\lambda'(t)$ and $\pi_t(x)$

# Flow transport problem on $\mathbb{R}$

- Re-write solution as

$$f(t,x) = \frac{\lambda'(t) I_t \left\{ F_t(x) - I_t^x / I_t \right\}}{\pi_t(x)}$$

where $I_t^x = \mathbb{E}_{\pi_t}[\mathbb{1}_{(-\infty,x]} \log L(X_t)]$ and $F_t$ is CDF of $\pi_t$

- **Speed** is controlled by $\lambda'(t)$ and $\pi_t(x)$

- **Sign** is given by difference between $F_t(x)$ and $I_t^x / I_t \in [0,1]$

# Flow transport problem on $\mathbb{R}^d, d \geq 1$

- Multivariate solution for $d = 3$

$$(\pi_t f_1)(t, x_{1:3}) = -\int_{-\infty}^{x_1} \partial_t \pi_t(u_1, x_2, x_3)\, du_1 + g_1(t, x_1) \int_{-\infty}^{\infty} \partial_t \pi_t(u_1, x_2, x_3)\, du_1$$

$$(\pi_t f_2)(t, x_{1:3}) = -g_1'(t, x_1) \int_{-\infty}^{\infty} \int_{-\infty}^{x_2} \partial_t \pi_t(u_1, u_2, x_3)\, du_{1:2} + g_1'(t, x_1) g_2(t, x_2) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \partial_t \pi_t(u_1, u_2, x_3)\, du_{1:2}$$

$$(\pi_t f_3)(t, x_{1:3}) = -g_1'(t, x_1) g_2'(t, x_2) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{x_3} \partial_t \pi_t(u_1, u_2, u_3)\, du_{1:3}$$

where $g_1, g_2 \in C^2([0,1] \times \mathbb{R}, [0,1])$

# Flow transport problem on $\mathbb{R}^d, d \geq 1$

- Checking **Liouville**

$$
\begin{aligned}
\partial_{x_1}(\pi_t f_1)(t, x_{1:3}) = &- \partial_t \pi_t(x_1, x_2, x_3) \\
&+ g_1'(t, x_1) \int_{-\infty}^{\infty} \partial_t \pi_t(u_1, x_2, x_3)\, du_1 \\
\partial_{x_2}(\pi_t f_2)(t, x_{1:3}) = &- g_1'(t, x_1) \int_{-\infty}^{\infty} \partial_t \pi_t(u_1, x_2, x_3)\, du_{1:2} \\
&+ g_1'(t, x_1) g_2'(t, x_2) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \partial_t \pi_t(u_1, u_2, x_3)\, du_{1:2} \\
\partial_{x_3}(\pi_t f_3)(t, x_{1:3}) = &- g_1'(t, x_1) g_2'(t, x_2) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \partial_t \pi_t(u_1, u_2, x_3)\, du_{1:2}
\end{aligned}
$$

# Flow transport problem on $\mathbb{R}^d, d \geq 1$

- Checking **Liouville**

$$\partial_{x_1}(\pi_t f_1)(t, x_{1:3}) = -\partial_t \pi_t(x_1, x_2, x_3) + g_1'(t, x_1)\int_{-\infty}^{\infty} \partial_t \pi_t(u_1, x_2, x_3)\, du_1$$

$$\partial_{x_2}(\pi_t f_2)(t, x_{1:3}) = -g_1'(t, x_1)\int_{-\infty}^{\infty} \partial_t \pi_t(u_1, x_2, x_3)\, du_{1:2} + g_1'(t, x_1) g_2'(t, x_2)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \partial_t \pi_t(u_1, u_2, x_3)\, du_{1:2}$$

$$\partial_{x_3}(\pi_t f_3)(t, x_{1:3}) = -g_1'(t, x_1) g_2'(t, x_2)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \partial_t \pi_t(u_1, u_2, x_3)\, du_{1:2}$$

- Taking divergence gives telescopic sum

$$-\nabla \cdot (\pi_t f)(t, x_{1:3}) = -\sum_{i=1}^{3} \partial_{x_i}(\pi_t f_i)(t, x_{1:3}) = \partial_t \pi_t(x_{1:3})$$

# Flow transport problem on $\mathbb{R}^d, d \geq 1$

A1 For $f$ to be **locally Lipschitz**, assume

$$\pi_0, L \in C^1(\mathbb{R}^d, \mathbb{R}_+) \Longrightarrow f \in C^1([0,1] \times \mathbb{R}^d, \mathbb{R}^d)$$

# Flow transport problem on $\mathbb{R}^d, d \geq 1$

A1 For $f$ to be **locally Lipschitz**, assume

$$\pi_0, L \in C^1(\mathbb{R}^d, \mathbb{R}_+) \Longrightarrow f \in C^1([0,1] \times \mathbb{R}^d, \mathbb{R}^d)$$

A2 For **integrability** of $\int_0^1 \int_{\mathbb{R}^d} |f(t,x)| \pi_t(x)\, dx\, dt < \infty$, necessarily

$$|\pi_t f|(t,x)| \to 0 \text{ as } |x| \to \infty$$

if $\{g_i\}$ are **non-decreasing** functions with **tail behaviour**

$$g_i(t, x_i) \to 0 \quad \text{as } x_i \to -\infty,$$
$$g_i(t, x_i) \to 1 \quad \text{as } x_i \to \infty$$

# Flow transport problem on $\mathbb{R}^d, d \geq 1$

A1 For $f$ to be **locally Lipschitz**, assume

$$\pi_0, L \in C^1(\mathbb{R}^d, \mathbb{R}_+) \Longrightarrow f \in C^1([0,1] \times \mathbb{R}^d, \mathbb{R}^d)$$

A2 For **integrability** of $\int_0^1 \int_{\mathbb{R}^d} |f(t,x)|\pi_t(x)\, dx\, dt < \infty$, necessarily

$$|\pi_t f|(t,x)| \to 0 \text{ as } |x| \to \infty$$

if $\{g_i\}$ are **non-decreasing** functions with **tail behaviour**

$$\begin{aligned} g_i(t,x_i) &\to 0 \quad \text{as } x_i \to -\infty, \\ g_i(t,x_i) &\to 1 \quad \text{as } x_i \to \infty \end{aligned}$$

- Choosing $g_i(t,x_i) = F_t(x_i)$ as **marginal CDF** of $\pi_t$ allows $f$ to decouple if distributions are independent

# Approximate Gibbs flow transport

- Solution involved integrals of **increasing dimension** as it tracks **increasing conditional** distributions

$$\pi_t(x_1|x_{2:d}), \pi_t(x_2|_{3:d}), \ldots, \pi_t(x_d), \quad x_i \in \mathbb{R}$$

# Approximate Gibbs flow transport

- Solution involved integrals of **increasing dimension** as it tracks **increasing conditional** distributions

$$\pi_t(x_1|x_{2:d}), \pi_t(x_2|_{3:d}), \ldots, \pi_t(x_d), \quad x_i \in \mathbb{R}$$

- Trade-off accuracy for computational tractability: track **full conditional** distributions

$$\pi_t(x_i|x_{-i}), \quad x_i \in \mathbb{R}^{d_i}$$

# Approximate Gibbs flow transport

- Solution involved integrals of **increasing dimension** as it tracks **increasing conditional** distributions

$$\pi_t(x_1|x_{2:d}), \pi_t(x_2|_{3:d}), \dots, \pi_t(x_d), \quad x_i \in \mathbb{R}$$

- Trade-off accuracy for computational tractability: track **full conditional** distributions

$$\pi_t(x_i|x_{-i}), \quad x_i \in \mathbb{R}^{d_i}$$

- **System of Liouville equations**

$$-\nabla_{x_i} \cdot \left\{ \pi_t(x_i|x_{-i}) \tilde{f}_i(t,x) \right\} = \partial_t \pi_t(x_i|x_{-i}),$$

each defined on $(0,1) \times \mathbb{R}^{d_i}$

# Approximate Gibbs flow transport

- For $d_i = 1$, solution is

$$\tilde{f}_i(t, x) = \frac{-\int_{-\infty}^{x_i} \partial_t \pi_t(u_i | x_{-i})\, du_i}{\pi_t(x_i | x_{-i})}$$

# Approximate Gibbs flow transport

- For $d_i = 1$, solution is

$$\tilde{f}_i(t,x) = \frac{-\int_{-\infty}^{x_i} \partial_t \pi_t(u_i|x_{-i})\, du_i}{\pi_t(x_i|x_{-i})}$$

- If $\pi_0, L \in C^1(\mathbb{R}^d, \mathbb{R}_+)$ and $\lim_{|x|\to\infty} L(x) = 0$, the ODE

$$dX_t = \tilde{f}(t, X_t)dt, \quad X_0 \sim \pi_0$$

admits a **unique solution** on $[0,1]$, referred to as **Gibbs flow**

# Approximate Gibbs flow transport

- For $d_i = 1$, solution is

$$\tilde{f}_i(t,x) = \frac{-\int_{-\infty}^{x_i} \partial_t \pi_t(u_i|x_{-i})\, du_i}{\pi_t(x_i|x_{-i})}$$

- If $\pi_0, L \in C^1(\mathbb{R}^d, \mathbb{R}_+)$ and $\lim_{|x|\to\infty} L(x) = 0$, the ODE

$$dX_t = \tilde{f}(t, X_t)dt, \quad X_0 \sim \pi_0$$

  admits a **unique solution** on $[0,1]$, referred to as **Gibbs flow**

- For $d_i > 1$, can often exploit **analytical tractability** of $\pi_t(x_i|x_{-i})$ to solve for $\tilde{f}_i(t,x)$; or apply **multivariate extension**

# Approximate Gibbs flow transport

- For $d_i = 1$, solution is

$$\tilde{f}_i(t,x) = \frac{-\int_{-\infty}^{x_i} \partial_t \pi_t(u_i|x_{-i})\, du_i}{\pi_t(x_i|x_{-i})}$$

- If $\pi_0, L \in C^1(\mathbb{R}^d, \mathbb{R}_+)$ and $\lim_{|x|\to\infty} L(x) = 0$, the ODE

$$dX_t = \tilde{f}(t, X_t)dt, \quad X_0 \sim \pi_0$$

  admits a **unique solution** on $[0,1]$, referred to as **Gibbs flow**

- For $d_i > 1$, can often exploit **analytical tractability** of $\pi_t(x_i|x_{-i})$ to solve for $\tilde{f}_i(t,x)$; or apply **multivariate extension**

- Otherwise, analogous to Metropolis-within-Gibbs, **split into one dimensional components** and apply above

# Error control

- Define **local error**

$$\begin{aligned}\varepsilon_t(x) &= \left|\partial_t \pi_t(x) + \nabla \cdot (\pi_t(x)\tilde{f}(t,x))\right| \\ &= \left|\partial_t \pi_t(x) - \sum_{i=1}^{p} \partial_t \pi_t(x_i|x_{-i})\pi_t(x_{-i})\right|\end{aligned}$$

# Error control

- Define **local error**

$$\begin{aligned}\varepsilon_t(x) &= \left|\partial_t \pi_t(x) + \nabla \cdot (\pi_t(x)\tilde{f}(t,x))\right| \\ &= \left|\partial_t \pi_t(x) - \sum_{i=1}^{p} \partial_t \pi_t(x_i|x_{-i})\pi_t(x_{-i})\right|\end{aligned}$$

- Gibbs flow exploits **local independence**:

$$\pi_t(x) = \prod_{i=1}^{p} \pi_t(x_i) \Longrightarrow \varepsilon_t(x) = 0$$

# Error control

- Define **local error**

$$\begin{aligned}\varepsilon_t(x) &= \left|\partial_t \pi_t(x) + \nabla \cdot (\pi_t(x)\tilde{f}(t,x))\right| \\ &= \left|\partial_t \pi_t(x) - \sum_{i=1}^{p} \partial_t \pi_t(x_i|x_{-i})\pi_t(x_{-i})\right|\end{aligned}$$

- Gibbs flow exploits **local independence**:

$$\pi_t(x) = \prod_{i=1}^{p} \pi_t(x_i) \Longrightarrow \varepsilon_t(x) = 0$$

- If Gibbs flow induces $\{\tilde{\pi}_t\}_{t\in[0,1]}$ with $\tilde{\pi}_0 = \pi_0$

$$\|\tilde{\pi}_t - \pi_t\|_{L^2}^2 \leq t\int_0^t \|\varepsilon_u\|_{L^2}^2\, du \cdot \exp\left(1 + \int_0^t \|\nabla \cdot \tilde{f}\,(u,\cdot)\,\|_\infty\, du\right)$$

# Numerical integration of Gibbs flow

- Previously, we considered the **forward Euler scheme**

$$Y_m = Y_{m-1} + \Delta t\, \tilde{f}(t_{m-1}, Y_{m-1}) = \Phi_m(Y_{m-1})$$

# Numerical integration of Gibbs flow

- Previously, we considered the **forward Euler scheme**

$$Y_m = Y_{m-1} + \Delta t \, \tilde{f}(t_{m-1}, Y_{m-1}) = \Phi_m(Y_{m-1})$$

- To get $\mathrm{Law}(Y_m)$, we need **Jacobian determinant** of $\Phi_m$ which typically costs $O(d^3)$ for $d_i = 1$

# Numerical integration of Gibbs flow

- Previously, we considered the **forward Euler scheme**

$$Y_m = Y_{m-1} + \Delta t\, \tilde{f}(t_{m-1}, Y_{m-1}) = \Phi_m(Y_{m-1})$$

- To get $\mathrm{Law}(Y_m)$, we need **Jacobian determinant** of $\Phi_m$ which typically costs $O(d^3)$ for $d_i = 1$

- In contrast, this scheme mimicking a **systematic Gibbs scan**

$$Y_m[i] = Y_{m-1}[i] + \Delta t\, \tilde{f}(t_{m-1}, Y_m[1:i-1], Y_{m-1}[i:p])$$
$$Y_m = \Phi_{m,d} \circ \cdots \circ \Phi_{m,1}(Y_{m-1})$$

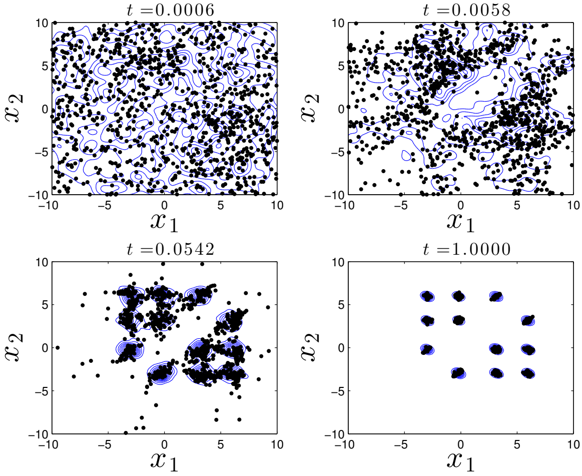
is also **order one**, and costs $O(d)$

# Mixture modelling example

- Lack of identifiability induces $\pi$ on $\mathbb{R}^4$ with $4! = 24$ **well-separated** and **identical modes**

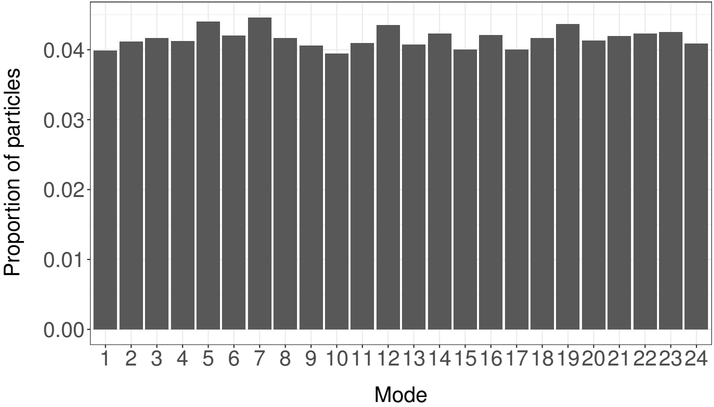
# Mixture modelling example

- Lack of identifiability induces $\pi$ on $\mathbb{R}^4$ with $4! = 24$ **well-separated** and **identical modes**
- Gibbs flow approximation



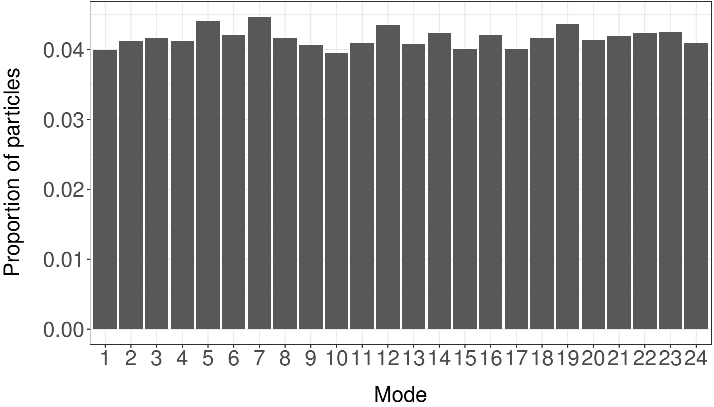
# Mixture modelling example

- Proportion of particles in each of the 24 modes



# Mixture modelling example

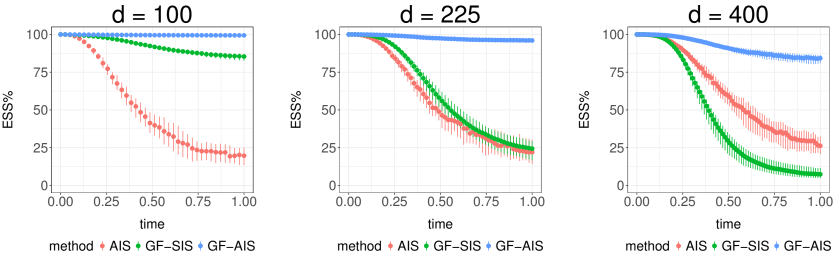
- Proportion of particles in each of the 24 modes




- Pearson's Chi-squared test for uniformity gives p-value of 0.85

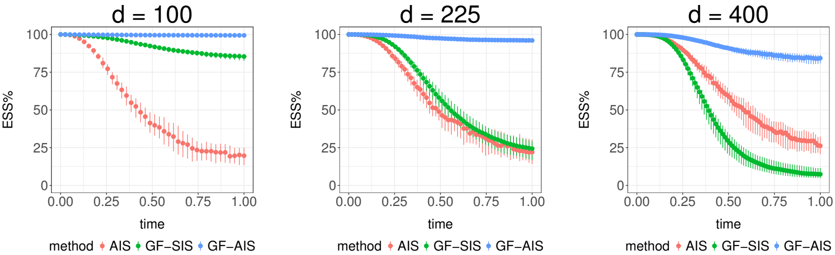
# Cox point process model

- Effective sample size % in dimension $d$



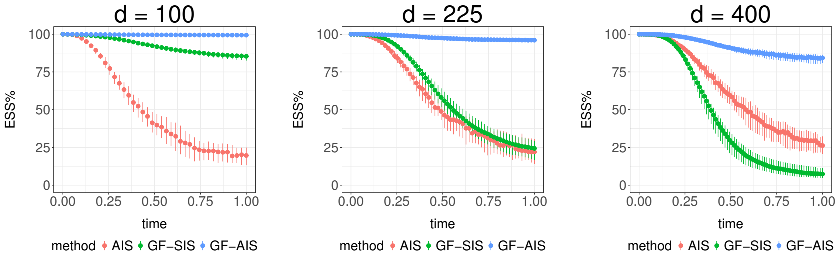
# Cox point process model

- Effective sample size % in dimension $d$
- AIS: AIS with HMC moves



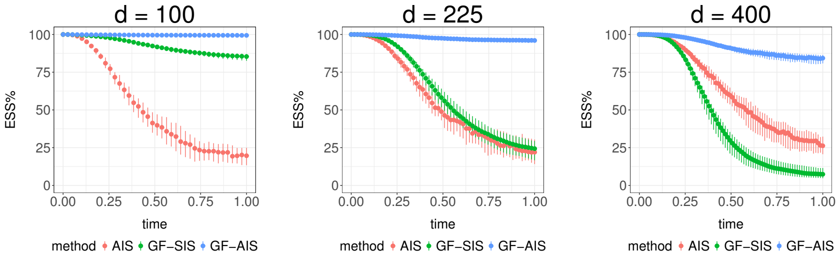
# Cox point process model

- Effective sample size % in dimension $d$
- AIS: AIS with HMC moves
- GF-SIS: Gibbs flow



# Cox point process model

- Effective sample size % in dimension $d$
- AIS: AIS with HMC moves
- GF-SIS: Gibbs flow
- GF-AIS: Gibbs flow with HMC moves



# End


- Heng, J., Doucet, A., & Pokern, Y. (2015). Gibbs Flow for Approximate Transport with Applications to Bayesian Computation. arXiv preprint arXiv:1509.08787.


- Updated article and `R` package coming soon!