Monte Carlo with Approximate Solutions to Dacorogna–Moser Flows

Austin Brown* and Jeffrey S. Rosenthal†

May 13, 2025

Abstract

A fundamental problem in statistics and machine learning is to simulate realizations from a probability measure μ on a Euclidean space with only access to its probability density function. The main technique for approximately solving this problem is Markov chain Monte Carlo, but this has exhibited difficulties in determining the iterations required to mix in many practical problems. In this article, we explore Monte Carlo sampling with approximate solutions to a discretized Dacorogna–Moser ordinary differential equation over the unit time interval. This approach is guaranteed to generate an approximate sample from μ within a fixed number of iterations and readily allows independent simulations in parallel. This method requires estimating one-dimensional integrals at each iteration and simulating a discretized solution to a randomly initialized ordinary differential equation. We develop finite approximation bounds in Wasserstein distances for many discretization schemes with access to an approximation of the normalizing constant. Numerical illustrations are demonstrated on a mixture distribution and Bayesian generalized linear models where the approach shows promise but can suffer some numerical instability.

Keywords: Monte Carlo estimation; Optimal transport; Parallel Monte Carlo

*Department of Statistical Sciences, University of Toronto, email: ad.brown@utoronto.ca
†Department of Statistical Sciences, University of Toronto, email: jeff@math.toronto.edu

1 Introduction

Let μ be a Borel probability measure with support on a compact set K in Euclidean space $\mathbb{R}^d$ of dimension $d \in \mathbb{Z}_+$. We will assume μ has a density function f with respect to Lebesgue measure. An important problem in machine learning, statistics, and stochastic optimization is to generate approximate samples from μ that can be used for Monte Carlo estimation. We will assume we have access to sampling a Borel probability measure ν also supported on K with density g. The goal is to construct a deterministic flow $T_t(\cdot)$ that starts from a random initialization $\xi \sim \nu$ and connects ν to μ so that $T_0(\xi) \sim \nu$ and at time $t = 1$, $T_1(\xi) \sim \mu$. When $T_1(\cdot)$ is invertible and sufficiently smooth, an alternative view of this is solving a deformation of the probability densities so that

$$\int_K \varphi(x) g(T_1^{-1}(x)) \det(\nabla T_1^{-1}(x)) dx = \int_K \varphi(x) f(x) dx$$

holds for all real-valued continuous bounded functions φ. Using a flow $T_t(\xi)$ to find solutions to deformation problems was pioneered by Dacorogna and Moser in the context of partial differential equations [Dacorogna and Moser, 1990]. The main idea we adapt here is to construct a random flow $T_t(\xi)$ so its distribution is the linear interpolation $(1-t)\nu + t\mu$. For a carefully constructed vector-valued function b,

$$T_t(\xi) = \xi + \int_0^t \frac{b(T_t(\xi))}{(1-t)g(T_t(\xi)) + tf(T_t(\xi))} \tag{1}$$

has $(1-t)\nu + t\mu$ as its distribution and therefore $T_1(\xi) \sim \mu$. This idea to use the random flow from Dacorogna–Moser was further investigated and shown to solve Monge's optimal transportation problem between two probability measures [Evans and Gangbo, 1999]. The Dacorogna-Moser flow has also been successfully applied to many applications in traffic optimization by solving minimal flow models [Santambrogio, 2014, 2015].

The focus of this article investigates numerical integration approximations to estimate the continuous Dacorogna–Moser flow (1) with the aim to simulate approximate samples from μ. This has many practical applications in estimating integrals via Monte Carlo in statistics. Our main contribution develops non-asymptotic approximation bounds in Wasserstein distances for many discretization schemes with only access to an approximation of the density f. Optimal transport metrics such as Wasserstein distances are widely

utilized for discrepancies between probability measures as they control the bias among all Lipschitz continuous functions [Villani, 2009]. In particular, we focus on the accuracy of first-order methods or Euler schemes but the techniques developed here can be readily generalized to higher order discretization schemes such as Runge-Kutta. Numerical illustrations are demonstrated on a banana-shaped distribution for Bayesian generalized linear models where the approach shows promise but can also suffer from numerical instability.

Transformation flows between probability measures have received a large amount of attention recently in the generative modeling literature in the context of training a loss function to learn the flow [Liu et al., 2022, Lipman et al., 2023, Chen et al., 2023]. The context of this article is different in that we specifically use flows from Dacorogna–Moser and assume we have access to a probability density f, possibly unnormalized. In comparison, generative modeling approaches use deep neural networks to estimate flows with only access to samples from the target probability measure.

The sampling techniques developed here are only applicable to probability measures on $\mathbb{R}^d$ and requires many regularity assumptions to ensure reliability. At the same time, many benefits such as generating independent samples appear promising in comparison to existing Markov chain Monte Carlo approaches. One apparent issue that can be observed immediately with this method by inspection of (1) is that the initial error propagating through the flow $T_t(\cdot)$ may need to be somewhat irregular if ν and μ are near singular. Intuitively, if the initial sample $\xi \sim \nu$ may need to traverse far regions of the state space through $T_t(\cdot)$ to produce accurate approximate samples from μ. Some other drawbacks to this approach is that we must simulate the path up to time $t = 1$ to generate a single independent sample from the target distribution. In comparison, MCMC uses a sequential simulation run to generate correlated realizations that eventually approximate realizations from the target distribution.

This article is organized as follows. Section 2 develops existence results for the continuous flow defined by (1). Section 3 studies first-order discretization schemes and analyzes non-asymptotic convergence bounds in the Wasserstein distance using coupling techniques adapted from the study of ordinary differential equations. The results are extended in this section into a general framework for higher-order approximation schemes. Section 4.1 studies a numerical example sampling a normal mixture distribution and Section 4.2 for Bayesian generalized linear models with priors supported near the maximum likelihood estimate. Section 5 provides an overview of the results and provides some future research

directions.

2 Monte Carlo with exact Dacorogna–Moser flows

Denote Lebesgue measure on $\mathbb{R}^d$ by λ and let $\|\cdot\|$ denote the Euclidean norm. We use the notation $X \sim \gamma$ to denote a random variable X has distribution γ. Let $\mu = f\lambda$ and $\nu = g\lambda$ be a Borel probability measures supported a compact set $K \subset \mathbb{R}^d$ with continuous density functions $f, g : \mathbb{R}^d \to [0, \infty)$ and $f, g > 0$ on K. Here μ is the desired target measure and ν will serve as an initial measure that we will assume to be able to generate samples from. Let $(w_k)_{k=1}^d$ be a set of weights with $w_k \geq 0$ and $\sum_k w_k = 1$ and define the function $b : \mathbb{R}^d \to \mathbb{R}^d$ through its coordinates $k = 1, \ldots, d$

$$b(x_1, \ldots, x_d)_k = w_k \int_{-\infty}^{x_k} [g(x_1, \ldots, s, \ldots, x_d) - f(x_1, \ldots, s, \ldots, x_d)]\, ds.$$

For example, $w_k = 1/d$ is often a reasonable choice in practice. We have the following general existence result.

Theorem 1. *Assume $f, g : \mathbb{R}^d \to [0, \infty)$ are positive and continuously differentiable on a compact set $K \subset \mathbb{R}^d$. Then for $t \in [0, 1]$, there exists a transformation $T_t : \mathbb{R}^d \to \mathbb{R}^d$ such that $T_0(\xi) \sim \nu$ and $T_1(\xi) \sim \mu$ and satisfying*

$$T_t(\xi) = \xi + \int_0^t \frac{b(T_s(\xi))}{(1-t)g(T_s(\xi)) + tf(T_s(\xi))} ds, \qquad \xi \sim \nu. \tag{2}$$

Proof. The goal is to follow the flow of probability measures $\mu_t = (1-t)\nu + t\mu$ defined for $t \in [0, 1]$ and show $T_t(\xi) \sim \mu_t$. Let $\varphi \in C_c^\infty([0, 1] \times \mathbb{R}^d)$ and then

$$\partial_t \int \varphi d\mu_t = \int \varphi f d\lambda - \int \varphi g d\lambda.$$

Now define the function $v : [0, 1] \times \mathbb{R}^d \to \mathbb{R}^d$ by

$$v(t, x) = v_t(x) = \frac{b(x)}{(1-t)g(x) + tf(x)}.$$

By assumption then

$$\sup_{t\in[0,1],x\in K} \|v_t(x)\| + \sup_{t\in[0,1],x\in K} \|\nabla v_t(x)\| < \infty. \tag{3}$$

Since then

$$\int_0^1 \int_K \|v_t\| \, d\mu_t dt < \infty,$$

we can use integration by parts to get

$$\int \varphi \partial_t d\mu_t(x) + \int \varphi \mathrm{tr}(D_x(v_t(x)d\mu_t(x))) = \int \varphi(f-g)d\lambda + \int \varphi(g-f)d\lambda = 0.$$

Therefore, μ_t solves the continuity equation in the sense of distributions. Then it follows by existence of initial value problems that $T_t : \mathbb{R}^d \to \mathbb{R}^d$ solves the initial value problem almost everywhere ν over the entire unit interval $[0,1]$ defined by

$$\frac{d}{dt}\gamma(t) = v_t(\gamma(t)), \qquad\qquad \gamma(0) = x$$

Now by [Ambrosio et al., 2008, Proposition 8.1.8],

$$\int \varphi d\mu_t = \int \varphi(T_t(x)) d\nu(x).$$

So then $T_t(\xi) \sim \mu_t$ and this implies that $T_1(\xi) \sim \mu$. □

If we can directly simulate $T_1(\xi)$, then it is possible to generate independently $\xi_1, \ldots, \xi_m$ each having distribution ν, and then $T_1(\xi_1), \ldots, T_1(\xi_m)$ are independent samples from μ. We directly have a law of large numbers for all Borel functions $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ with $\int \|\varphi\| \, d\mu < \infty$ so that

$$\lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{m} \varphi(T_1(\xi_k)) = \int_K \varphi d\mu.$$

holds with probability 1.

Remark 2. To simplify the presentation and notation we used a linear interpolation $(1-t)g + tf$, but generalizations to alternative interpolations is possible. In particular, any interpolation if functions $\alpha, \beta : \mathbb{Z}_+ to [0,1]$ such that the pair satisfies $\alpha(t) + \beta(t) = 1$

would suffice to replace $(1-t)$ with $\alpha(t)$ and t with $\beta(t)$. The solution in Theorem 1 would then be

$$T_t(\xi) = \xi + \int_0^t \frac{b_t(T_s(\xi))}{\alpha(t)g(T_s(\xi)) + \beta(t)f(T_s(\xi))} ds, \qquad \xi \sim \nu$$

where for $k = 1, \ldots, d$

$$b_t(x_1, \ldots, x_d)_k = -w_k \int_{-\infty}^{x_k} \left[\frac{d}{dt}\alpha(t)g(x_1, \ldots, s, \ldots, x_d) + \frac{d}{dt}\beta(t)f(x_1, \ldots, s, \ldots, x_d) \right] ds.$$

3 Monte Carlo with first-order discretizations of the inexact Dacorogna–Moser flow

In general, we cannot simulate the continuous time flow $T_1(\xi)$ directly, and will focus on first-order or Euler discretizations in this section. Another potential issue is we may not have access to the normalizing constant of f. To allow for estimation of the normalizing constant, let $\hat{f}$ be an approximation to f and $\hat{b}$ be an approximation to b. Then define the function $\hat{v} : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ by

$$\hat{v}_t(x) = \frac{\hat{b}(x)}{(1-t)g(x) + t\hat{f}(x)}. \tag{4}$$

For example, if the normalizing constant of f is not known, then we may estimate the normalizing constant of $Z = \int f(x)dx$ with Monte Carlo methods such as importance sampling and use this to construct f.

We estimate the flow from ν to μ using a sequence of functions with a first-order discretization starting from $\hat{T}_0(\cdot) = \mathrm{id}$ and defined recursively for $1 \leq k \leq n$ by

$$\hat{T}_k(\cdot) = \hat{T}_{k-1}(\cdot) + \frac{1}{n}\hat{v}_{k/n}(\hat{T}_{k-1}(\cdot)). \tag{5}$$

Then if $\xi \sim \nu$, we can define the marginal distribution for $\hat{T}_k(\xi)$ by $\hat{\mu}_k$. At each iteration, the one-dimensional integrals needed to compute $b(\hat{T}_k(\xi))$ will need to be solved. However, adaptive quadrature can be used to accurately estimate $b(\hat{T}_k(\xi))$ up to arbitrary precision and in computed in parallel to improve computational speed.

Using this approximation, we will measure the accuracy in the Wasserstein distance $\mathcal{W}_p$ of order $p \geq 1$ defined so that

$$\mathcal{W}_p(\hat{\mu}_n, \mu) = \left(\inf_{\Gamma \in \mathcal{C}(\hat{\mu}_n, \mu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|y - x\|^p \right)^{1/p}$$

where $\mathcal{C}(\hat{\mu}_n, \mu)$ is the set of joint Borel measures satisfying $\Gamma(\cdot, \mathbb{R}^d) = \hat{\mu}_n$ and $\Gamma(\mathbb{R}^d, \cdot) = \mu$. The Wasserstein distance is useful as it precisely controls the bias of Lipschitz functions through its rich duality theory. The following result gives a first-order approximation bound in any Wasserstein distance for any $p \geq 1$.

Theorem 3. *Assume $f, g : \mathbb{R}^d \to [0, \infty)$ are positive and continuously differentiable on a compact set $K \subset \mathbb{R}^d$. Assume for some $\delta \in (0, 1)$*

$$\sup_{0 \leq k \leq n} \int_{\mathbb{R}^d} \left\| \hat{v}_{k/n}(\hat{T}_k(x)) - \frac{b(\hat{T}_k(x))}{(1-t)g(\hat{T}_k(x)) + tf(\hat{T}_k(x))} \right\| \leq \delta. \tag{6}$$

Then there is a constant $C > 0$ such that for all $n \in \mathbb{Z}_+$

$$\mathcal{W}_p(\hat{\mu}_n, \mu) \leq C \left(\delta + \frac{1}{n} \right).$$

Satisfying (6) appears reasonable in certain problems and can be accomplished in many ways and one example is through uniform approximations such as

$$\sup_{x \in K, t \in [0,1]} \left\| \hat{v}_t(x) - \frac{b(x)}{(1-t)g(x) + tf(x)} \right\| \leq \delta.$$

The constant appearing in the result of Theorem 3 can be difficult to write in a nice compact form. Moreover, this constant can be large and is exponentially dependent on the regularity of f, g.

Proof of Theorem 3. Given fixed $x \in \mathbb{R}^d$, we have a solution to the ordinary differential equation $\partial_t \gamma_t = v_t(\gamma_t)$ where $\gamma_0 = x$ with

$$T_t(x) = x + \int_0^t \frac{b(T_s(x))}{(1-t)g(T_s(x)) + tf(T_s(x))} ds.$$

The solution exists due to Theorem 1 and moreover, $T_1(\xi) \sim \mu$. We have an upper bound on the optimal coupling with

$$\begin{aligned}\mathcal{W}_p(\hat{\mu}_n, \mu) &\leq \left(\mathbb{E}\left[\left\|\hat{T}_n(\xi) - T_1(\xi)\right\|^p\right]\right)^{1/p} \\ &\leq \left(\int_K \left\|\hat{T}_n(x) - T_1(x)\right\|^p d\nu(x)\right)^{1/p}.\end{aligned}$$

Thus the idea is a Lyapunov technique that comes from studying the stability of discretizing ordinary differential equations. For $x \in \mathbb{R}^d$, define

$$E_k(x) = \left\|T_{k/n}(x) - T_{(k-1)/n}(x) - (1/n)v_{(k-1)/n}(T_{(k-1)/n}(x))\right\|$$

and also define

$$\delta_k(x) = \left\|\frac{\hat{b}(\hat{T}_{k/n}(x))}{(1-t)g(\hat{T}_{k/n}(x)) + t\hat{f}(\hat{T}_{k/n}(x))} - v_{t_k}(\hat{T}_{k/n}(x))\right\|.$$

We then have by the triangle inequality

$$\begin{aligned}&\left\|\hat{T}_{k+1}(x) - \hat{T}_k(x) - T_{(k+1)/n}(x) + T_{k/n}(x)\right\| \\ &\leq \frac{1}{n}\left\|\hat{v}_{k/n}(\hat{T}_k(x)) - v_{k/n}(T_{k/n}(x))\right\| + \left\|T_{(k+1)/n}(x) - T_{k/n}(x) - \frac{1}{n}v_{k/n}(T_{k/n}(x))\right\| \\ &\leq \frac{1}{n}\delta_k(\xi) + \frac{1}{n}\left\|v_{t_k}(\hat{T}_k(x)) - v_{t_k}(T_{t_k}(x))\right\| + E_{k+1}(x).\end{aligned}$$

Since f, g are positive and continuous with compact support, define $\inf f = f_*$ and $\inf g = g_*$ and $\sup f = f^*$ and $\sup g = g^*$. Since v is continuous on its support and that there is an $R > 0$ such that $\sup_{x\in\mathbb{R}^d} \|b(x)\| \leq R$. Since f, g are continuously differentiable, define

$$L = \frac{g^* \vee f^*}{g_* \wedge f_*} + \sup_x \|b(x)\| \sup_x \|\nabla f(x)\| \vee \sup_x \|\nabla g(x)\| < \infty.$$

Taking the derivative, we have that

$$\sup_{x,t}\left\|D_x\left[\frac{b(x)}{(1-t)g(x) + tf(x)}\right]\right\| \leq L$$

and then $b(\cdot)/[(1-t)g(\cdot)+tf(\cdot)]$ is L Lipschitz for every fixed t. It follows then the Lyapunov condition

$$\left\|\hat{T}_{k+1}(x)-T_{(k+1)/n}(x)\right\| \leq \left(1+\frac{L}{n}\right)\left\|\hat{T}_k(x)-T_{k/n}(x))\right\|+\frac{1}{n}\delta_k(\xi)+E_{k+1}(\xi).$$

Applying this recursively

$$\left\|\hat{T}_n(x)-T_1(x)\right\| \leq \sum_{k=1}^{n}\left[E_k(x)+\frac{1}{n}\delta_{k-1}(x)\right](1+L/n)^{n-k}.$$

We have the bound

$$\sum_{k=1}^{n}(1+L/n)^{n-k} \leq \frac{n[(1+L/n)^n-1]}{L} \leq n\frac{[\exp(L)-1]}{L}.$$

Taking the derivative, we have with $c=\sup_x \|b(x)\|^2 \frac{f^*\vee g^*}{(f_*\wedge g_*)^3}$

$$\partial_z \|v_z(y_z(y))\| \leq \|\partial_z v_z(y_z(y))\| \leq c.$$

It follows that

$$\sup_x E_k(x) \leq \frac{c}{2n^2}.$$

Taking the expectation and by assumption on the error (6), we have

$$\begin{aligned} W_p(\hat{\mu}_n,\mu) &\leq \sup_{0\leq k\leq n}\left[\int E_k(x)d\nu(x)+\frac{1}{n}\int \delta_k(x)d\nu(x)\right] n\frac{[\exp(L)-1]}{L} \\ &\leq \frac{\exp(L)-1}{L}\left\{n\sup_{0\leq k\leq n}\int E_k(x)d\nu(x)+\delta\right\} \\ &\leq \frac{\exp(L)-1}{L}\left\{\frac{c}{2n}+\delta\right\}. \end{aligned}$$

□

3.1 Higher order approximations for Monte Carlo

In this section, we put the previous results of Section 3 into a general framework that allows for further approximation accuracy. We first construct an abstract generalization

of the discretized flow (5). Let $\hat{u}_t : \mathbb{R}^d \to \mathbb{R}^d$ be a Borel function defined for all $t \in [0,1]$ and define for $k \in \mathbb{Z}_+$,

$$\hat{W}_k(\xi) = \hat{W}_{k-1}(\xi) + \frac{1}{n}\hat{u}_{k/n}(\hat{W}_{k-1}(\xi)). \tag{7}$$

In the special case when $\hat{u}_{k/n} = \hat{v}_{k/n}$, we return to the definition (5). Let $\tilde{\mu}_k$ denote the distribution of $\hat{W}_k(\xi)$. Let $u_t : \mathbb{R}^d \to \mathbb{R}^d$ be a Borel function defined for all $t \in [0,1]$ and define

$$E_k(x) = \left\| T_{k/n}(x) - T_{(k-1)/n}(x) - (1/n)u_{(k-1)/n}(T_{(k-1)/n}(x)) \right\|$$

where $T_t(x) = x + \int_0^t v_s(T_s(x))ds$ is defined previously by (2).

We are specifically interested in an improved Euler discretization to achieve a higher rate of convergence than provided in Theorem 3. More specifically, consider the special case when

$$u_{k/n}(x) = \frac{1}{2}\left[v_{k/n}(x) + v_{(k+1)/n}\left(x + \frac{1}{n} v_{k/n}(x) \right) \right] \tag{8}$$

and $\hat{u}_{k/n}$ is defined by replacing $v_{k/n}$ by an approximation $\hat{v}_{k/n}$. However, the definitions of u_t and $\hat{u}_t$ are more general and we have the following abstract result.

Theorem 4. *Assume $f, g : \mathbb{R}^d \to [0,\infty)$ are positive and continuously differentiable on a compact set $K \subset \mathbb{R}^d$ and additionally assume for some $\delta \in (0,1)$,*

$$\sup_{0 \le k \le n} \int_{\mathbb{R}^d} \left\| \hat{u}_{k/n}(\hat{W}_{k/n}(x)) - u_{k/n}(\hat{W}_{k/n}(x))) \right\| d\nu(x) \le \delta. \tag{9}$$

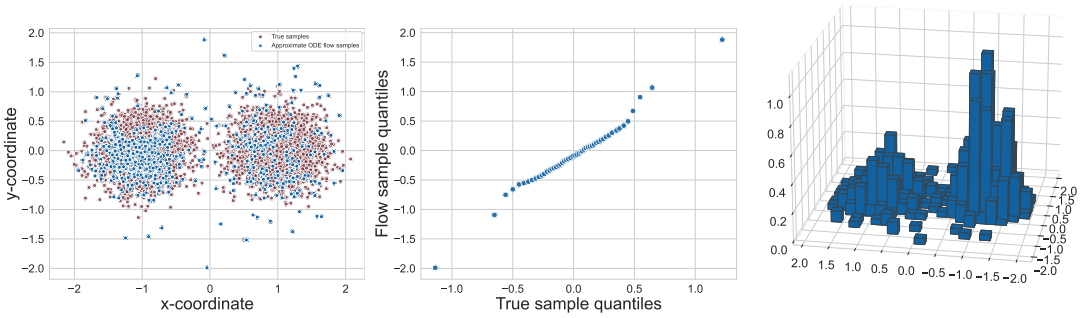
Then there is a constant $C_1 > 0$ such that for all $n \in \mathbb{Z}_+$

$$W_p(\tilde{\mu}_n, \mu) \le C_1 \left\{ n \sup_{0 \le k \le n} \int_K E_k(x) d\nu(x) + \delta \right\}.$$

Proof. The proof is a direct generalization of the technique in Theorem 3. □

Now we apply Theorem 4 in the case when u_t is defined as in (8) and $\hat{u}_t$ satisfies the approximation (9).

Corollary 5. *Assume the conditions of Theorem 4 hold with u_t defined by (8). Then*



(a) (b) (c)

Figure 1: Plot (a) is the approximate samples from the discretized flow versus true samples for the normal mixture distribution. Plot (b) is a quantile versus quantile plot of the approximate flow samples to the true samples. Plot (c) is a histogram of the approximate flow samples.

there is a constant $C_2 > 0$ *such that for all* $n \in \mathbb{Z}_+$

$$W_p(\tilde{\mu}_n, \mu) \leq C_2 \left\{ \frac{1}{n^2} + \delta \right\}.$$

Proof. The proof is a direct application of Taylor expansion on u_t and applying the conclusion of Theorem 4. □

4 Empirical performance examples

4.1 Mixture distributions

We now shift to look at empirical Monte Carlo simulations using the discretized transformation flow defined in (5). Here the goal is to simulate independent realizations by interpolating starting from a normal distribution and ending at a normal mixture distribution with separated modes. This problem is generally difficult for many Markov chain Monte Carlo algorithms even in low dimensions and we focus on dimension 2. Let $\mathrm{N}(\mu, \Sigma)$ denote a normal distribution on Euclidean space with mean vector μ and covariance matrix Σ. In this case, the initial distribution is $\mathrm{N}(0, \sigma_0^2 I)$ with scaling $\sigma_0^2 > 0$ and

the target distribution is a normal mixture

$$\pi \equiv \frac{1}{2}\mathrm{N}((-1,0)^T, \sigma^2 I) + \frac{1}{2}\mathrm{N}((1,0)^T, \sigma^2 I). \tag{10}$$

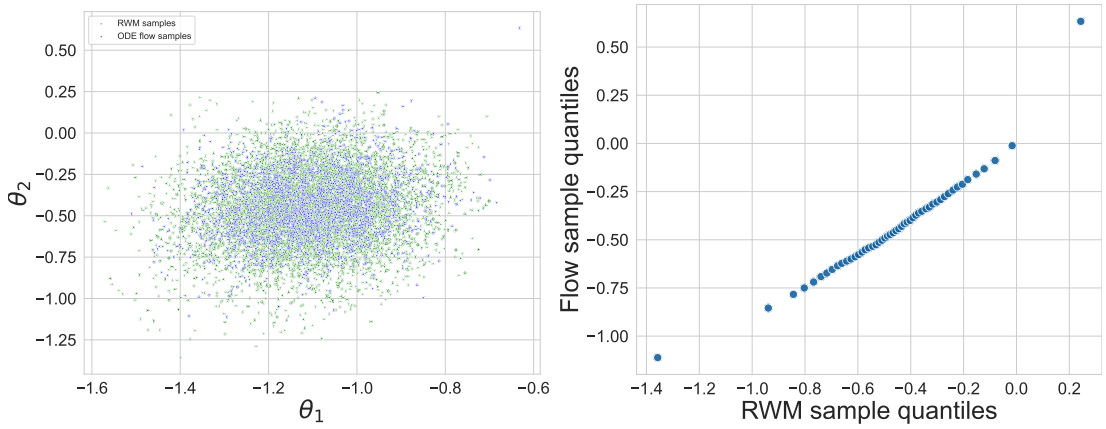
with scaling $\sigma^2 = 1/10$.

We first simulate 5000 samples independently from this mixture distribution (10) and compare these to samples generated from the discretized transformative flow approximations developed in Section 3. For the flow, we generate 5000 independent samples where we start with initial distribution $\mathrm{N}(0, 4I)$ and uniform weights, and employ an adaptive numerical ODE solver [Petzold, 1983] using max step size $h = 10^{-4}$ to generate each sample. We also experienced similar simulation results with alternative scalings of the initial distribution. The implementation of the method suffers some numerical instability and we filter samples that experienced any instability and were realized outside of the ball of radius 2, which seems to appear due to division by small numbers. Figure 1 (a) shows the results of the numerical simulation with realizations sampled from the approximate flow compared to true Monte Carlo realizations. We can see the transformative flow can approximately sample the mixture distribution starting from a standard normal, but can produce more outlier samples and slight skewness to the mode at coordinate $(1,0)^T$. Figure 1 (b) compares the empirical quantiles of the second coordinate of both methods against each other, and Figure 1 (c) plots the histogram of the approximate samples generated by the flow. In these plots, we also see slight skewness to one of the modes.

4.2 Bayesian generalized linear models

Generalized linear models such as logistic regression are widely-utilized applied models across statistics. For logistic regression, the data consists of binary responses $y_i \in \{0, 1\}$ and vector-valued predictors $x_i \in \mathbb{R}^d$ for $i = 1, \ldots, N$ so that the probability of a positive response given certain predictors follows

$$\mathbb{P}(y_i = 1 | x_i, \theta) = \frac{\exp(\theta^T x_i)}{1 + \exp(\theta^T x_i)}.$$



(a) (b)

Figure 2: Plot (a) is the approximate samples from the ODE flow versus random-walk Metropolis MCMC samples and (b) plots the quantiles of the approximate flow versus the quantiles of the random-walk MCMC samples.

The likelihood in this model is $L_n(\theta) = \exp(-\ell_n(\theta))$ defined by the loss function

$$\ell_n(\theta) = \sum_{i=1}^{n} \left[\log(1 + \exp(\theta^T x_i) - y_i \theta^T x_i\right]$$

and we assume there exists a point θ_n that minimizes the loss. We consider a truncated normal prior π_0 from a normal distribution with mean θ_n and covariance $\sigma_0^2 (X^T X)^{-1}$ with hyper-parameter $\sigma_0^2 > 0$ and support $B_r(\theta_n) = \{\theta : \|\theta - \theta_n\| \leq r\}$ for some hyper-parameter $r > 0$. The posterior for this generalized linear model has density is

$$\pi_n(\theta) = \exp\left(-\ell_n(\theta)\right) \pi_0(\theta)/Z$$

supported on $B_r(\theta_n)$ where $Z = \int \exp\left(-\ell_n(\theta)\right) \pi_0(\theta) d\theta$.

We artificially generate a data set for this posterior with $N = 50$ and parameter dimension $d = 2$ being the intercept and slope parameter. The most common method for simulation from this posterior is Markov chain Monte Carlo with Metropolis-Hastings and we simulate $n = 10^4$ realizations from a Random-walk Metropolis Markov chain Monte Carlo algorithm tuned with normal proposal and identity covariance scaled by $1/(nd)$, and run for sufficiently long to satisfy burn-in. We will compare these Monte Carlo samples to using the approximate transformative flow (5) developed here with

uniform weights. In this model, we can estimate the normalizing constant with $\hat{Z}$ using Monte Carlo by sampling from π_0. We simulate $n = 10^4$ realizations of the flow starting with initial measure $\mu = \pi_0$ being the prior and step size $h = 10^{-5}$. We filter any realizations that experience numerical instability. Figure 2 (a) shows the independent realizations from the approximate flow against the correlated realizations of the RWM MCMC samples and Figure 2 (b) shows the quantiles of the first coordinates plotted against each other. We see similar performance from both algorithms where one benefit is that the flow produces independent samples and may be readily run in parallel. On the other hand, some drawbacks are that some simulations experienced numerical instability and these points need to be removed.

5 Conclusion

Discretizing solutions to initial value problems appears like a promising research direction for some new alternative approaches to some classical sampling methods such as Markov chain Monte Carlo. In this article, we studied discretizing the solution to the initial value problem initally used by Dacorogna-Moser for solving partial differential equations [Dacorogna and Moser, 1990]. One benefit is that this method readily allows parallel computing and produces independent samples compared to Markov chain Monte Carlo which generates correlated samples with a sequential algorithm that is impossible to fully parallelize. Another large advantage of this approach is the discretized flow is run for a fixed number of iterations known beforehand and therefore, no convergence diagnostics are needed. In comparison to Markov chain Monte Carlo, the convergence is often unable to be verified as the required simulation length is generally unknown.

While there are some interesting advantages, there are also many limitations to this new Monte Carlo method. The first is that we required access to a "good" estimate of the normalizing constant of the target measure which can potentially be unreasonable for many practical problems. This appears to be the largest drawback to the approach, but can hopefully be remedied in future research. We also found numerical instability issues likely due to the fraction in the flow defined in (1) that can arise from the division of small numbers on a computer. It is not understood if this numerical instability is due to the current implementation and may be fixed in the future or is inherent to the method. Despite these drawbacks, Monte Carlo through approximating transformative flows (2)

is an interesting future research direction.

References

Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

Sitan Chen, Sinho Chewi, Holden Lee, Yuanzhi Li, Jianfeng Lu, and Adil Salim. The probability flow ODE is provably fast. *arXiv preprint arXiv:2305.11798*, 2023.

Bernard Dacorogna and Jürgen Moser. On a partial differential equation involving the Jacobian determinant. *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, 7(1):1–26, February 1990. ISSN 0294-1449, 1873-1430. doi: 10.1016/s0294-1449(16)30307-9.

L. C. Evans and W. Gangbo. Differential equations methods for the Monge-Kantorovich mass transfer problem. *Memoirs of the American Mathematical Society*, 137(653):0–0, 1999. ISSN 0065-9266, 1947-6221. doi: 10.1090/memo/0653.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow Matching for Generative Modeling, February 2023. arXiv:2210.02747 [cs].

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

Linda Petzold. Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *SIAM Journal on Scientific and Statistical Computing*, 4(1):136–148, 1983. doi: 10.1137/0904010.

Filippo Santambrogio. A dacorogna-moser approach to flow decomposition and minimal flow problems. *ESAIM: Proceedings and Surveys*, 45:265–274, September 2014. ISSN 2267-3059. doi: 10.1051/proc/201445027.

Filippo Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, volume 87 of *Progress in Nonlinear Differential Equations and Their Applications*. Springer International Publishing, Cham, 2015. ISBN 978-3-319-20827-5 978-3-319-20828-2. doi: 10.1007/978-3-319-20828-2.

Cédric Villani, editor. *Optimal Transport: Old and New*. Number 338 in Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. ISBN 978-3-540-71049-3 978-3-540-71050-9. doi: 10.1007/978-3-540-71050-9.