Concave-Convex PDMP-based Sampling

Matthew Sutton a and Paul Fearnhead b

aCentre for Data Science, Queensland University of Technology, Brisbane, Australia; bDepartment of Mathematics and Statistics, Lancaster University, Lancaster, UK

ABSTRACT
Recently nonreversible samplers based on simulating piecewise deterministic Markov processes (PDMPs) have shown potential for efficient sampling in Bayesian inference problems. However, there remains a lack of guidance on how to best implement these algorithms. If implemented poorly, the computational costs of simulating event times can outweigh the statistical efficiency of the nonreversible dynamics. Drawing on the adaptive rejection literature, we propose the concave-convex thinning approach for simulating a piecewise deterministic Markov process, which we call CC-PDMP. This approach provides a general guide for constructing bounds that may be used to facilitate PDMP-based sampling. A key advantage of this method is its additive structure—adding concave-convex decompositions yields a concave-convex decomposition. This makes the construction of bounds modular, as given a concave-convex decomposition for a class of likelihoods and a family of priors, they can be combined to construct bounds for the posterior. We show that constructing our bounds is simple and leads to computationally efficient thinning. Our approach is well suited to local PDMP simulation where conditional independence of the target can be exploited for potentially huge computational gains. We provide an R package and compare with existing approaches to simulating events in the PDMP literature. Supplementary materials for this article are available online.

1. Introduction
Monte Carlo methods based on continuous-time Markov processes have shown potential for efficient sampling in Bayesian inference challenges (Fearnhead et al. 2018; Goldman and Singh 2021). These new sampling algorithms are based on simulating piecewise-deterministic Markov processes (PDMPs). Such processes evolve deterministically between random event times with possibly random dynamics at event times (see Davis 1993, for an introduction to PDMPs). Monte Carlo methods based on simulating a PDMP with a desired invariant distribution were first proposed in the computational physics literature where they were motivated as examples of nonreversible Markov processes (Peters and de With 2012; Michel, Kapfer, and Krauth 2014). These ideas transitioned to the statistics community as an alternative to traditional Markov chain Monte Carlo (MCMC) samplers. Popular algorithms in this development include the bouncy particle sampler (Bouchard-Côté, Vollmer, and Doucet 2018) and the Zig-Zag sampler (Bierkens, Fearnhead and Roberts 2019) amongst others (Vanetti et al. 2017; Wu and Robert 2020; Michel, Durmus, and Sénécal 2020; Bierkens et al. 2020).

There is substantial theoretical and empirical evidence to support the claim that nonreversible MCMC samplers offer more efficient sampling (Diaconis, Holmes, and Neal 2000; Bierkens and Roberts 2017; Bouchard-Côté, Vollmer, and Doucet 2018) as nonreversibility reduces the chance of moving back to areas that have recently been visited. Samplers based on PDMPs simulate from a density \( \pi(\theta) \) by introducing a velocity component \( v \) which is used to evolve the position of the state, \( \theta \), with deterministic dynamics. At stochastic event times the velocity component is updated using a Markov kernel after which the process continues with the new velocity. The main barrier to implementing a PDMP-based sampler is the simulation of the random event times. This involves simulating the next event time in a time-inhomogeneous Poisson process with rate function that depends on the current position of the sampler.

Simulation of event times in a Poisson process is often facilitated through a method known as thinning (Lewis and Shedler 1979). The aim of this article is to provide a general framework to aid practitioners in implementing PDMP-based samplers. Specifically we introduce the concave-convex thinning approach to facilitate simulating a PDMP-based sampler, which we call CC-PDMP. This method can be applied whenever the rate function of the PDMP can be decomposed into the sum of concave and convex functions, and can be used to facilitate thinning from polynomial rate functions, allowing for broad applicability of the method. We discuss the efficiency of the method compared to alternative approaches in the literature for exact sampling of event times.

Related to our work is the concave-convex adaptive rejection sampler (CC-ARS) of Görür and Teh (2011) which uses the same upper-bounding approach to construct bounds on
the log density within a rejection sampler that draws independent samples from univariate density functions. In contrast, by applying this technique within PDMP sampling, the bounds are constructed on the gradient of the log density (i.e., the rate) and used within the nonreversible Markov process framework to generate samples from multivariate densities. Moreover, the CC-PDMP framework allows for both subsampling and local updating schemes to facilitate these methods for high dimensional sampling. A short comparison between concave-convex adaptive rejection sampling and CC-PDMP for sampling from a univariate density is given in Appendix C, supplementary materials.

Some authors have proposed using automatic but approximate methods for simulating a PDMP-based sampler. Among these approaches, Cotter, House, and Pagani (2020) propose a novel numerical integration and root-finding algorithms to facilitate these methods for high dimensional sampling. Others have considered using approximate local updating schemes to facilitate these methods for high dimensional sampling. These authors have used within the nonreversible Markov process framework to generate samples exactly (Pakman et al. 2017; Goldman and Singh 2021; Goan et al. 2023). Both of these approaches sacrifice exact sampling of the posterior and involve a tradeoff between the computational cost and the approximation of the sampling distribution.

The article is organized as follows. In Section 2, we introduce the technical details of simulating from a PDMP-based sampler and the details of some popular samplers. Section 3 reviews the literature on thinning for PDMP-based samplers. We introduce the concave-convex PDMP approach for adaptive thinning in Sections 4 and 5. Empirical evaluation of CC-PDMP, and comparison to existing methods, is given in Section 6. We conclude with discussion of limitations and extensions that are possible for the method. Code for implementing our method and replicating our results is available at https://github.com/matt-sutton/ccpdmp.

2. Sampling using a PDMP

2.1. Piecewise Deterministic Markov Processes

Before we explore PDMP-based samplers, we first review PDMPs. A PDMP is a stochastic process, and we will denote its state at time $t$ by $z_t$. There are three key components that define the dynamics of a PDMP:

1. A set of deterministic dynamics defined as $z_{t+s} = \psi(z_t, s)$.
2. Event times that are driven by an inhomogeneous Poisson process with rate $\lambda(z_t)$.
3. A Markov transition kernel $q(z|z_{t-})$ where $z_{t-} := \lim_{t\downarrow t} z_t$.

A PDMP with these specifications will evolve according to its deterministic dynamics between event times and at event times will update according to the Markov transition kernel $q(z|z_{t-})$. The event times must be simulated from a Poisson process with rate $\lambda(z_t)$, which depends on the state of the process at time $t$. The result of simulating this process is a PDMP skeleton $\{(t_k, z_k)\}_{k=1}^n$ where for $k = 1, \ldots, n$, $t_k$ denotes the $k$th event time and we have stored the state of the process immediately before this event time, $z_{t_k}$. Given this skeleton it is possible to fill in the values of the state between the events using the deterministic dynamics.

2.2. PDMP-based Samplers

Assume we wish to construct a PDMP process to sample from a target distribution of interest, $\pi(\theta)$. Current PDMP-based samplers are defined on an augmented space $z \in \mathcal{E}$, where $\mathcal{E} = \mathcal{X} \times \mathcal{V}$, that can be viewed as having a position, $\theta$, and a velocity, $v$. As described below, the dynamics of the PDMP can be chosen so that the PDMP's invariant distribution has the form $\pi(z) = \pi(\theta) p(v|\theta)$, for some conditional distribution of the velocities, $p(v|\theta)$. As a result, the $\theta$-marginal of the invariant distribution is the target distribution we wish to sample from. We will consider target distributions of the form

$$\pi(\theta) \propto \exp(-U(\theta))$$

where $\theta \in \mathbb{R}^p$ and $U(\theta)$ is the associated potential.

There are two ways to evaluate Monte Carlo estimates of $\int g(z) \pi(z)dz$ using the output of a PDMP. The first estimator averages over the continuous time path of the PDMP,

$$\frac{1}{t_n - t_1} \sum_{k=1}^{n-1} \int_{t_k}^{t_{k+1}} g\left(\psi(z_{t_k}, s)\right)ds.$$

This requires that the integral of $g(\psi(z_{t_k}, s))$ with respect to $s$ may be computed. The second, simpler, approach is to “discretise” the PDMP path. This involves taking an integer $M > 0$, defining $s = (t_n - t_1)/M$, calculating the value of the state at discrete-time points, $z_{1}, \ldots, z_{n}$, and then using the familiar Monte Carlo estimator

$$\frac{1}{M} \sum_{j=1}^{M} g(z_{t_j}).$$

The points $z_{t_j}$ may be found by evolving the process along its deterministic dynamics for the appropriate event time interval.

2.3. Bouncy Particle Sampler

The bouncy particle sampler (BPS) takes the velocity space $\mathcal{V}$ to be either $\mathbb{R}^p$ or the unit sphere $S^{p-1}$ and targets an invariant distribution $\nu(z) = \pi(\theta) p(v)$ where $p(v)$ is either the standard $p$-dimensional Normal distribution, or a uniform distribution on $S^{p-1}$. The sampler moves with linear deterministic dynamics $\psi((\theta_t, v_t), s) = (\theta_t + sv_t, v_t)$ until a random time $t$ occurs according to a Poisson process with canonical event rate given by

$$\lambda^c(z_t) = \max \{0, \langle v_t, \nabla U(\theta_t) \rangle \}.$$

At an event time $t$ the velocity is updated as $v_t = R_{\theta_t-}(v_{t-})$ where

$$R_{\theta}(v) = v - 2 \frac{\langle v, \nabla U(\theta) \rangle}{\|\nabla U(\theta)\|^2} \nabla U(\theta).$$

This update can be interpreted as reflecting the velocity off the hyperplane tangential to the gradient of $U(\theta)$. Additionally, at random times $t$ an additional event occurs according to a homogeneous Poisson process with rate $\lambda^M > 0$ in which the velocity is refreshed, drawing $v_t \sim p(v)$. Refreshment ensures that the Bouncy Particle Sampler samples the invariant distribution and
does not get “stuck” on contours of the potential (Bouchard-
Côté, Vollmer, and Doucet 2018). Formally the Bouncy Particle
Sampler has event rate \( \lambda(z_t) = \lambda^c(z_t) + \lambda^{ref} \) with Markov transition kernel
\[
q(z|z_{t-}) = \frac{\lambda^c(z_{t-})}{\lambda(z_{t-})} \delta_{z_{t-}}(\theta) \delta_{\theta_{t-}}(v_{t-}) + \frac{\lambda^{ref}}{\lambda(z_{t-})} \delta_{\theta_{t-}}(\theta) p(v),
\]
where \( \delta_{a}(A) \) denotes the Dirac delta function with point mass at \( a \). The BPS is an example of a \textit{global PDMP} as all components of the velocity \( v \) are changed according to the Markov transition kernel.

### 2.4. Zig-Zag Sampler

The Zig-Zag sampler takes the space of velocities, \( V \), to be \( \{-1,1\}^p \) with \( p(v) \) uniform on this space. Similarly to BPS, the Zig-Zag sampler uses linear dynamics \( \psi((\theta_t, v_t), s) = (\theta_t + sv_t, v_t) \) between events, but unlike the BPS only the single velocity vector is updated at event times. The overall Zig-Zag process can be viewed as a composition of \( p \) \textit{local event rates} where each event rate has a corresponding Markov transition kernel which operates on a single component of the velocity. Specifically, the \( i \)th component of the velocity is updated with local event rate
\[
\lambda_i(z_t) = \max \left\{ 0, \frac{v_i \partial \theta_i}{\partial \theta_i} U(\theta_i) \right\}
\]
for \( i = 1, \ldots, p \), where \( \partial \theta_i \) denotes the partial derivative of \( U(\theta_i) \) with respect to \( \theta_i \). Simulating an event time in this local framework consists of simulating an event time for each local event rate and then taking the minimum time as the event time \( t \) for the overall process. At an event time \( t \) triggered by event rate \( \lambda_i \), the velocity is updated as \( v_t = F_i(v_{t-}) \)
\[
F_i(v) = (v_1, \ldots, v_{i-1}, -v_i, v_{i+1}, \ldots, v_p)^T.
\]

Once the velocity is updated for component \( i \) local event times must be re-simulated only for rates \( \lambda_i(z_t) \) which are functions of \( \theta_i \). This fact can induce massive computational savings when there is conditional independence between the components of \( \theta \). To see this, consider the extreme case where the target is fully independent across dimensions. In this case each event rate, \( \lambda_i(z_t) \), will only depend on the \( i \)th components of \( \theta_i \) and \( v_i \). Thus, at each event, it is only the time of the next event for the component whose velocity changes that needs to be recalculated. The computational cost of simulation in this case will scale as \( O(1) \) per event as opposed to \( O(d) \) for the global BPS. This computational advantage was explored in the context of simulating from diffusion bridges in Bierkens et al. (2021).

### 2.5. Global and Local PDMP Methods

As described above, we can divide PDMP samplers into two main classes: \textit{global methods} and \textit{local methods}. These PDMPs differ in how they operate on the velocity vector when an event is triggered. Global methods are PDMP-based methods where the transition kernel acts on the entire velocity vector. By contrast, local methods are defined so that the transition kernel only affects a subset of the velocity. In this section we will give a general algorithm for constructing a local PDMP-based sampler (though see also Bouchard-Côté, Vollmer, and Doucet 2018).

Let \( v^S \) denote the sub-vector of \( v \) with elements indexed by \( S \subseteq \{1, \ldots, p\} \) and \( v^{-S} \) denote the sub-vector of \( v \) without the elements indexed in \( S \). For some set \( S \) define the local kernel as one that satisfies
\[
q^S_\theta(v|v_{t-}) = \delta_{v^{-S}}(v^{-S}) q^S_\theta(v|v_{t-}),
\]
that is, \( q^S_\theta(v|v_{t-}) \) only updates the sub-vector \( v^S \). Suppose we have a partition of the components of \( \theta_1, \ldots, \theta_p \) given by \( S = \{S_1, S_2, \ldots, S_F\} \). Subset \( S_f \) of \( v \) is updated with local kernel \( q^S_\theta(v_{t-}) \) and associated event rate
\[
\lambda_f(z_t) = \max \left\{ 0, (v^S_f, \nabla_{\theta^S_f} U(\theta)) \right\},
\]
for \( f = 1, \ldots, F \). Simulating a local PDMP with this structure involves simulating an event time for each of the \( F \) rates and applying the local transition kernel corresponding to the smallest simulated event time. For the partition \( S \) define \( \mathcal{N} = \{N_1, \ldots, N_F\} \)
\[
N_f = \{ m \in \{1, \ldots, F\} \mid \theta(S_m) \neq \theta(S_f) \}
\]
for \( f = 1, \ldots, F \) where \( \theta(S_m) \neq \theta(S_f) \) denotes that the rate of events associated with \( S_m \) depends on some element of \( \theta(S_f) \). When an event is triggered for a local rate \( f \) the velocity \( v^S_f \) is updated and new event times are simulated for rates which depend on the value of one or more \( \theta_i \) for \( i \in S_f \). This simulation process is summarized in \textit{alg1}.

In this framework the Zig-Zag sampler is a local PDMP with partition \( S = \{\{1\}, \{2\}, \ldots, \{p\}\} \) and local kernels \( q^S_\theta(v|v_{t-}) = \delta_{F_i(v_{t-})}(v) \). The local BPS uses kernel \( q^S_{\theta}(v|v_{t-}) = \delta_{\theta_{t-}}(v) \) where \( v^S_{-} = v^{-S}_{t-} \) and \( v^S_{t} = R^S_{\theta}(v^S_{t-}) \)
\[
R^S_{\theta}(v) = v^S - \frac{\nabla_{\theta^S} U(\theta)}{\|\nabla_{\theta^S} U(\theta)\|^2} \nabla_{\theta^S} U(\theta).
\]
If there is a single factor \( S = \{S\} \) containing all indices \( S = \{1, \ldots, p\} \) we recover the global BPS.

#### Algorithm 1: Simulating a PDMP with local structure

**Inputs:** \( t_1 = 0, \theta_1, v_1 \), factorization \((S, \mathcal{N})\) with \( S = \{S_1, S_2, \ldots, S_F\} \) and \( \mathcal{N} = \{N_1, N_2, \ldots, N_F\} \).

**Sample** \( \tau_f \) for \( f = 1, \ldots, F \) where
\[
\mathbb{P}(\tau_f \geq t) = \exp \left( -\int_0^t \lambda_f(\theta_0 + v_0 s, v_0) ds \right)
\]

For \( k = 1, \ldots, n \)
(a) Let \( f^* = \arg\min_{f} (\tau_f) \) and update \( t_{k+1} = t_k + \tau_f^* \)
(b) Update states: \( \theta_{t_{k+1}} = \theta_{t_k} + \tau_f^* v_{t_k} \)
(c) Update velocity: \( v_{t_{k+1}} = q_{\theta_{t_{k+1}}}^{S_f}(v_{t_k}) \)
(d) For \( f \in N_f^*, \) update times, that is, resample \( \tau_f \) where
\[
\mathbb{P}(\tau_f \geq t) = \exp \left( -\int_0^t \lambda_f(\theta_{t_{k+1}} + v_{t_{k+1}} s, v_{t_{k+1}}) ds \right)
\]

**Outputs:** PDMP skeleton \( \{(t_k, \theta_{t_k}, v_{t_k})\}_{k=1}^n \)
3. Algorithms for Event-Time Simulation

Simulating the first event time, \( \tau \), from a Poisson process with event rate \( \lambda(t) \) is equivalent to simulating \( u \sim \text{Uniform}[0, 1] \) and solving

\[
\tau = \text{argmin}_t \left\{ -\log(u) = \int_0^t \lambda(z_s) \, ds \right\}.
\]

If the event rate is sufficiently simple this can be done exactly. For example if the rate is constant or linear, then there are analytical solutions for \( \tau \) in terms of \( u \). If the rate function is convex a solution may be found using numerical methods (Bouchard-Côté, Vollmer, and Doucet 2018). For more complex functions, the rate can be simulated by a process known as thinning. This process makes use of Theorem 1.

**Theorem 1 (Lewis and Shedler (1979)).** Let \( \lambda(t) \) and \( \tilde{\lambda}(t) \) be continuous functions where \( \lambda(t) \leq \tilde{\lambda}(t) \) for all \( 0 \leq t \leq \tau_{\text{max}} \). Let \( \tau_1, \tau_2, \ldots, \tau_n \) be event times of the Poisson process with rate \( \lambda(t) \) over the interval \([0, \tau_{\text{max}}]\). For \( i = 1, \ldots, n \) retain the event time \( \tau_i \) with probability \( \lambda(\tau_i)/\tilde{\lambda}(\tau_i) \) to obtain a set of event times from a nonhomogeneous Poisson process with rate \( \lambda \) over the interval \([0, \tau_{\text{max}}]\).

The efficiency of simulating via thinning depends on how tightly the rate \( \lambda \) upper-bounds \( \tilde{\lambda} \) and how costly it is to simulate from \( \tilde{\lambda} \). Another key tool for enabling the simulation of event times is known as super-positioning and the general process is outlined in Theorem 2.

**Theorem 2 (Kingman (1992)).** Suppose that \( \Lambda_1, \ldots, \Lambda_n \) are a set of independent Poisson processes with rates \( \lambda_1(t), \ldots, \lambda_n(t) \) with first arrival times \( \tau_1^{[1]}, \ldots, \tau_n^{[n]} \). The Poisson process with rate \( \lambda(t) = \sum_{i=1}^n \lambda_i(t) \) may be simulated by returning the first arrival time as \( \tau = \text{min}_{i=1,\ldots,n} \tau_i^{[i]} \).

Both super-positioning and thinning provide useful tools in the simulation of event times from a nonhomogeneous Poisson process. However, constructing the upper-bounds required for thinning is largely an ad hoc and time consuming process for the practitioner. A common desire for a practitioner is to reuse simulation knowledge. For example if an event rate can be written as the sum of several sub-event rates, that can be exactly simulated, ideally this knowledge should be used to simulate from the sum. We refer to this class of event rates as additive rates. Specifically, we will refer to a Poisson process with a rate

\[
\lambda(z_t) = \text{max}(0, f_1(t) + f_2(t))
\]
as process with an additive rate. The super-positioning method from Theorem 2 facilitates additive event-time simulation by thinning using an upper-bounding event rate

\[
\tilde{\lambda}(t) = \text{max}(0, f_1(t) + \text{max}(0, f_2(t)));
\]

which satisfies \( \lambda(z_t) \leq \tilde{\lambda}(t) \). We can simulate the first event from a process with this rate as \( \tau = \text{min}(\tau_1, \tau_2) \), where \( \tau_1 \) is simulated with rate \( \tilde{\lambda}_1(t) = \text{max}(0, f_1(t)) \) and \( \tau_2 \) with rate \( \tilde{\lambda}_2(t) = \text{max}(0, f_2(t)) \). This simulated time will be accepted with probability

\[
\frac{\text{max}(0, f_1(\tau) + f_2(\tau))}{\text{max}(0, f_1(\tau)) + \text{max}(0, f_2(\tau))},
\]

which will be one when both \( f_1(\tau) > 0 \) and \( f_2(\tau) > 0 \). This procedure is useful since it allows the reuse of thinning procedures for \( f_1 \) and \( f_2 \). Thinning via super-positioning is often recommended in the literature; notable examples include Bouchard-Côté, Vollmer, and Doucet (2018) who illustrate this approach when simulating from an exponential family and Wu and Robert (2020) who discuss the approach generally. The approach was also recommended broadly by Sen et al. (2020) where it was suggested for practical implementation of the Zig-Zag where \( f_1 \) and \( f_2 \) correspond to the terms from the likelihood and prior, respectively.

4. Concave-Convex Adaptive Thinning

Our general proposal for simulating events in a PDMP is based on concave-convex adaptive thinning. As the dynamics between events are deterministic, conditional on the current state of the PDMP, we can rewrite the rate of the next event as a function of time. With slight abuse of notation we will denote this rate as \( \lambda(t) \) with the argument of the function making it clear whether we are viewing it as a function of the state, or as here, as a function of time. Suppose \( \lambda(t) = \max(0, f(t)) \) where \( f(t) \) can be decomposed as

\[
f(t) = f_u(t) + f_r(t)
\]
on a finite interval \( t \in [0, \tau_{\text{max}}] \) where \( f_u(t) \) is a convex function and \( f_r(t) \) is a concave function in time. We will later discuss general conditions where such a decomposition is possible. The problem of upper-bounding \( f(t) \) is recast to finding upper-bounding piecewise linear functions \( \ell_u(t) \) and \( \ell_r(t) \) such that \( f_u(t) \leq \ell_u(t) \) and \( f_r(t) \leq \ell_r(t) \). We may then apply the bound \( f(t) \leq \ell(t) \) where \( \ell(t) = \ell_u(t) + f_r(t) \). Since \( \ell(t) \) is the sum of piecewise linear functions it will also be a piecewise linear function and direct simulation from a Poisson process with rate \( \ell(t) \) is possible (see Appendix A, supplementary materials). Concave-convex adaptive thinning proceeds by constructing a piecewise linear function \( \ell(t) \) over a set of \( m \) abscissae \( t_0 < t_2 < \cdots < t_m = \tau_{\text{max}} \) at which the evaluation of \( f_u(t_1), f_r(t_1) \) and derivative \( f_r(t_1) \) are known. An event time \( \tau \) is simulated from the Poisson process with rate \( \tilde{\lambda}(\tau) = \max(0, \ell(t)) \) and is accepted with probability

\[
\frac{\max(0, f_u(\tau) + f_r(\tau))}{\tilde{\lambda}(\tau)}.
\]

If the event is rejected, we can reuse the information from the evaluation of \( f_u(t) \) and \( f_r(t) \) with an additional evaluation of \( f_r(t) \) to refine the simulation on the abscissae \( t_1 = \tau < t_2 < \cdots < t_m = \tau_{\text{max}} \) may be reused. If an event does not occur on the range \( t \in [0, \tau_{\text{max}}] \) the PDMP process is evolved by \( \tau_{\text{max}} \) and the thinning process is repeated. We give the general construction of the piecewise linear function \( \ell(t) \) for \( t \in [t_1, t_2] \) and note this construction may be applied iteratively over the abscissae. This construction is also depicted visually in Figure 1. A convex function \( f_u(t) \), is by definition a function that can be upper-bounded by the line segment connecting any two function evaluations. So, for any \( t \in [t_1, t_2] \) we have the upper-bound \( f_u(t) \leq \ell_u(t) \) where

\[
\ell_u(t) = f_u(t_1) + \frac{f_u(t_2) - f_u(t_1)}{t_2 - t_1}(t - t_1).
\]
For a concave function $f(t)$ with derivative $f'_c(t)$ the line segment corresponding to the tangent of $f_c(t)$ will upper-bond the function. Thus, for $t \in [t_1, t_2]$, $f_c(t)$ will be upper-bounded by

$$
\ell_u(t) = \min \left\{ f_c(t_1) + f'_c(t_1)(t - t_1), f_c(t_2) + f'_c(t_2)(t - t_1) \right\}.
$$

This minimum will switch at the point of intersection between the lines $f_c(t_1) + f'_c(t_1)(t - t_1)$ and $f_c(t_2) + f'_c(t_2)(t - t_1)$. It is simple to find this intersection point \cite{GörürTeh2011},

$$
t^* = \frac{f_c(t_2) - f'_c(t_2)t_2 - f_c(t_1) + f'_c(t_1)t_1}{f'_c(t_1) - f'_c(t_2)},
$$

which will be a point on the interval $[t_1, t_2]$ provided the derivatives $f'_c(t_1)$ and $f'_c(t_2)$ are not equal. If $f'_c(t_1) = f'_c(t_2)$, the linear segment will not change over the interval and we take $t^* = t_2$.

So we take

$$
\ell_u(t) = \begin{cases} 
  f_c(t_1) + f'_c(t_1)t & t \in [t_1, t^*) \\
  f_c(t_2) + f'_c(t_2)(t_2 - t) & t \in [t^*, t_2]
\end{cases},
$$

and combining these bounds we have $\ell_u(t) = \tilde{\ell}_u(t) + \ell_n(t)$ which is a piecewise linear function upper-bounding $f(t) \leq \ell(t)$ for $t \in [t_1, t_2]$.

### 4.1. Concave-Convex Decompositions

The proposition below gives simple conditions for the class of nonhomogeneous Poisson processes that admit thinning using the concave-convex approach.

**Proposition 1.** If a Poisson process with rate function $\lambda(t)$ can be written as $\lambda(t) = \max\{0, f(t)\}$ where $f(t)$ is continuous then the process admits thinning using concave-convex adaptive thinning.

**Proof.** Consider $f(t)$ on the interval $[0, \tau_{\text{max}}]$ where $\tau_{\text{max}}$ is some arbitrary value $\tau_{\text{max}} > 0$. On this closed compact interval $f(t)$ is continuous and by the Stone-Weierstrass theorem \cite{Stone1948}, for any $\epsilon > 0$ there exists a polynomial $g(t)$ on the interval with $\|f - g\| < \epsilon$. The function $g$ is a polynomial so it admits the concave-convex decomposition

$$
g(t) = \sum_{\{i \in \mathbb{N}, a_i > 0\}} a_i t^i + \sum_{\{i \in \mathbb{N}, a_i < 0\}} a_i t^i
$$

with convex function $g_0(t) = \sum_{\{i \in \mathbb{N}, a_i < 0\}} a_i t^i$ and concave function $g_0(t) = \sum_{\{i \in \mathbb{N}, a_i > 0\}} a_i t^i$. The Process with rate $\lambda(t)$ admits thinning on the interval $[0, \tau_{\text{max}}]$ using the Poisson process with rate $\tilde{\lambda}(t) = \max\{0, g(t) + \epsilon\}$ which has a concave-convex decomposition. If an event does not occur on the interval $[0, \tau_{\text{max}}]$ then the process is evolved to $\tau_{\text{max}}$ and the thinning process repeated.

**Remark 1.** While continuity of $f(t)$ is required for the Stone-Weierstrass theorem, the concave-convex process may also be applied more generally. If $f(t)$ is continuous for all $0 < t < t^*$ and $\lim_{t \to (t^*)^-} f(t) = \infty$ then concave-convex thinning can be applied. Consider $f(t)$ using the abscissae $t_0 = 0$, $t_1 = \frac{t^*}{2}$, $\tau_{\text{max}} = t^*$. **Proposition 1** ensures a concave-convex decomposition on the interval $[0, t^*/2]$. On $[t^*/2, t^*]$ we upper bound the rate by infinity. So if a time is not simulated on $[0, t^*/2)$, we will propose an event at $t_1$, and this event will be rejected. The concave-convex bound will then be updated based on $f(t^*/2)$, and used to simulate an event on $[t^*/2, t^*)$. The abscissae on this increment will be $t^*/2$, $3t^*/4$, and $t^*$, and the process will be repeated. Importantly, as $\lim_{t \to (t^*)^-} f(t) = \infty$ there will be an actual event before $t^*$ with probability one, thus, repeating this process will yield an (accepted) event time.

The proof of **Proposition 1** relies on finding a polynomial that can upper-bound the process over a finite interval. If a bound can be given for higher order derivatives of $f(t)$ then there are two main approaches for finding an upper-bounding polynomial. Using Taylor’s expansion of $f(t)$ about zero, we get

$$
f(t) = f(0) + f'(0)t + \frac{f''(0)}{2!} t^2 + \cdots + \frac{f^{(k)}(0)}{k!} t^k + \int_0^t f^{(k+1)}(s) \frac{t^k}{k!} ds.
$$

If there is a known constant $M$ such that $|f^{(k+1)}(t)| \leq M$ we can employ thinning based on the polynomial bound $g(t) = f(0) + f'(0)t + \frac{f''(0)}{2!} t^2 + \cdots + \frac{f^{(k)}(0)}{k!} t^k + \frac{M}{k+1} t^{k+1}$. Alternatively polynomial upper-bounds can be constructed based on interpolation where the error is controlled. For example using Lagrange polynomial interpolation on $[0, \tau_{\text{max}}]$ with $k + 1$ evaluations of the function has error bounded by $\tau_{\text{max}}^{k+1} M$. Adding
4.2. Concave-Convex Adaptive Thinning and PDMP-based Samplers

A key advantage to CC-PDMP is that it can easily and efficiently simulate from rates with an additive structure. That is, if \( f_1(t) \) and \( f_2(t) \) both have concave-convex decompositions then \( f(t) = f_1(t) + f_2(t) \) will also have a concave-convex decomposition. The rates associated to a PDMP-based sampler with \( \pi(\theta) \propto \exp(-U(\theta)) \) as its target will have an additive structure. In particular, a local PDMP sampler as described in Section 2.4 will have rates

\[
\lambda_f(z_t) = \max \left\{ 0, \sum_{j \in S_f} f_j(t) \right\}
\]

where \( f_j(t) = v_j \partial \theta_j U(\theta + tv) \). This means that the concave-convex decompositions for the partial derivatives \( f_j(t) \) can be trivially combined to simulate from new local or global implementations of the PDMP. There is a subtle tradeoff between computational and statistical efficiency offered by global and local PDMP-based samplers which we explore in Section 6.

Additive rates are also present when using a PDMP to simulate from rates with an additive structure. That is, if \( f(t) = f_1(t) + f_2(t) \) both have concave-convex decompositions then \( \frac{1}{U(t)} \) will also have a concave-convex decomposition. For the rate function to be split into more manageable pieces.

5. Tuning Parameters

5.1. Choice of Abscissae

A key user choice in CC-PDMP sampling is that of the position and number of abscissae on the interval \([0, \tau_{\max}]\). Suppose that CC-PDMP simulation is implemented with abscissae \( \ell_0 = 0 < \ell_1 < \cdots < \ell_n = \tau_{\max} \) if an event does not occur on the interval \([0, \tau_{\max}]\) the PDMP will be evolved by \( \tau_{\max} \) and the thinning process will repeat where the function evaluations at \( \tau_{\max} \) can be reused at \( \ell_0 = 0 \). Thus, one simulation with \( n \) abscissae would be computationally equivalent to \( n \) evolutions of the CC-PDMP approach using abscissae with only two points \( t_0 = 0 \) and \( t_1 = \tau_{\max} \). This encourages choosing a minimal number of abscissae and more carefully choosing the length of the interval. There are two main issues that can occur when tuning the parameter \( \tau_{\max} \). If \( \tau_{\max} \) is too small, then the proposal frequently lie outside the interval \([0, \tau_{\max}]\) and simulating an event will require many iterations of bounding the event rate. While if \( \tau_{\max} \) is too large, then the bound on the rate may be poor, leading to many events that are rejected. In this article, we refer to an iteration of the CC-PDMP that does not generate an event time (i.e., a rejected proposal event or not generating on the interval) as a shadow event. The total number of iterations will be the sum of the number of events and shadow events. Efficiency is measured as the proportion of iterations that are events. Adapting the value of \( \tau_{\max} \) will not change the sampling dynamics. Consequently, unlike in adaptive MCMC, this parameter may be adapted throughout the course of simulating the PDMP. Ideally this parameter should be a little larger than the average event time so that events are proposed on the interval. A simple automatic approach for choosing this parameter is to set \( \tau_{\max} \) equal to the \( q \) th percentile of previous simulated event times, and update \( \tau_{\max} \) every 100 iterations of the sampler. We found that setting \( q = 80 \)th percentile worked well for automatically selecting \( \tau_{\max} \). We investigate the dependency on this tuning parameter by implementing the Zig-Zag sampler on the two-dimensional Banana distribution, with

\[
U(\theta) = (\theta_1 - 1)^2 + \kappa (\theta_2 - \theta_1^2)^2,
\]

where \( \kappa \) controls how much mass concentrates around the region \( \theta_2 \approx \theta_1^2 \). Since the potential is polynomial in both parameters it is trivial to see that the Zig-Zag rate functions will also be polynomial. Further implementation details may be found in the supplementary material and associated GitHub.

Figure 2 shows the efficiency of the Zig-Zag sampler on the Banana distribution with \( \kappa = 1 \) when changing the interval length \( \tau_{\max} \) from 0.1 to 4 with two abscissae. The Zig-Zag algorithm was simulated for 10,000 events at each value of \( \tau_{\max} \). The effect of selecting \( \tau_{\max} \) too small or too large is clearly seen and the red line shows the performance using the adaptive approach. The main computational cost is due to an artifact of the otherwise fast C++ code having to call an R function each time the concave-convex decomposition is evaluated. For simple event rates such as polynomials this computation can be implemented in C++ as well. Appendix D, supplementary materials illustrates the effect of \( \tau_{\max} \) on computation time when the convex-concave decomposition is evaluated in C++.

5.2. Choice of Decomposition

The choice of concave-convex decomposition is not unique (Göür and Teh 2011). Arbitrary convex functions can be added to \( f_1(t) \) and subtracted from \( f_2(t) \) to give a new valid decomposition. A concave-convex decomposition is minimal if there is no non-affine convex function that can be added to \( f_1(t) \) and subtracted from \( f_2(t) \) while preserving the convexity of the decomposition (Hartman 1959). For example, consider the function

\[
f(t) = -t^3 + 3t^2 - 3t + 3,
\]

where \( f''(t) = -6t + 6 \). It is simple to see that the function is convex for \( t < 1 \) and concave for \( t > 1 \). A minimal decomposition is given by the piecewise functions
We now present empirical evaluation of CC-PDMP and comparison with other approaches to simulate PDMPs. Our experiments were implemented using the R package ccpdmp available at https://github.com/matt-sutton/ccpdmp.

The package enables you to simulate a PDMP provided one specifies the concave and convex decomposition of the rate function. If the rate function is polynomial (or bounded by a polynomial) the practitioner may parse this instead and the concave-convex decomposition will be handled internally. The package contains some basic example use cases and code to reproduce the experiments.

6. Experiments

We now present empirical evaluation of CC-PDMP and comparison with other approaches to simulate PDMPs. Our experiments were implemented using the R package ccpdmp available at https://github.com/matt-sutton/ccpdmp.

The package enables you to simulate a PDMP provided one specifies the concave and convex decomposition of the rate function. If the rate function is polynomial (or bounded by a polynomial) the practitioner may parse this instead and the concave-convex decomposition will be handled internally. The package contains some basic example use cases and code to reproduce the experiments.

6.1. Application in Generalized Linear Models

Generalized Linear Models (GLMs) provide a rich class of models that are frequently used in Bayesian inference. Let \( \{ (y_i, x_i) \}_{i=1}^{n} \) be the observed data, where \( y_i \) is an observed response and \( x_i \in \mathbb{R}^p \) is a vector of associated covariates for \( i = 1, \ldots, n \). The expected value of \( y_i \) is modeled by \( g^{-1}(x_i^T \theta) \) where \( g^{-1} : \mathbb{R} \rightarrow \mathbb{R} \) is the inverse link function. The potential of the likelihood has the form

\[
U(t) = \sum_{i=1}^{n} \phi(x_i^T \theta, y_i)
\]

where \( \phi : \mathbb{R}^2 \rightarrow \mathbb{R} \) is the GLM mapping function \( \phi(a, y) = -\log p(y|g^{-1}(a)) \) returning the negative log-likelihood for observation \( y \) given \( a \). The partial derivative with respect to \( \theta_k \) has the form

\[
\partial_{\theta_k} U(t) = \sum_{i=1}^{n} \phi'(x_i^T \theta, y_i)x_{ik}
\]

where \( \phi'(a, y) = \partial_{a} \phi(a, y) \) and higher order derivatives are defined similarly. Over the time interval \( t \in [0, \tau] \) let \( f_k(t) = \nu_k \partial_{\theta_k} U(t + tv) \) which is

\[
f_k(t) = \nu_k \sum_{i=1}^{n} \phi'(a_i(t), y_i)x_{ik}
\]

where \( a_i(t) = x_i^T (\theta + tv) \). We can use this to define local rates as \( \lambda_k(t) = \max(0, f_k(t)) \) for \( k = 1, \ldots, p \). For GLMs, repeated application of the chain rule yields:

\[
f^{(j)}_k(t) = \nu_k \sum_{i=1}^{n} \phi^{(j+1)}(a_i(t), y_i) \left( \frac{\partial a_i}{\partial t} \right)^j x_{ik}
\]

where \( \frac{\partial a_i}{\partial t} = x_i^T v \) and \( f^{(j)}_k(t) \) denotes the \( j \)th derivative of \( f_k(t) \). When there are bounds on \( \phi(a) \) we can use the upper-bound Taylor polynomial. In Appendix B, supplementary materials we provide bounds for several modeling choices. Here we consider logistic regression, which has the mapping function \( \phi(a, y) = \log(1 + \exp(a)) - ya \) where \( y \in \{0, 1\} \). We look at the efficiency of CC-PMGP thinning for a five dimensional logistic regression problem with \( n = 200 \) observations. The covariates were generated from a multivariate normal, \( x_i \sim \mathcal{N}(0, V^{-1}) \) for \( i = 1, \ldots, 200 \), with mean zero, precision matrix

\[
V = \begin{bmatrix}
1 & \rho & 0 & 0 & 0 \\
\rho & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]
Figure 3. The upper-bounds resulting from two different concave-convex decompositions of the function \( f(t) = -t^3 + 3t^2 - 3t + 3 \), based on abscissae at 0 and 1. The top row corresponds to the decomposition from Proposition 1 and the bottom row corresponds to an optimal decomposition which recognizes \( f(t) \) as a convex function on \([0, 1]\). The columns show the functions \( f_u, f_n \) and \( f = f_u + f_n \) as well as their piece-wise linear upper-bounds.

Table 1. Efficiency of thinning in Zig-Zag sampling of logistic regression for increasing order of Taylor series polynomial thinning bound.

| Polynomial Order | Correlation 0.00 | Correlation 0.25 | Correlation 0.50 | Correlation 0.65 | Correlation 0.75 | Correlation 0.85 | Correlation 0.95 |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| 1                | 0.64(4.13)       | 0.58(4.41)       | 0.49(4.87)       | 0.41(5.57)       | 0.34(6.47)       | 0.26(8.06)       | 0.13(14.80)      |
| 2                | 0.77(4.13)       | 0.76(4.12)       | 0.75(4.16)       | 0.72(4.26)       | 0.68(4.42)       | 0.62(4.75)       | 0.39(6.56)       |
| 3                | 0.80(5.56)       | 0.80(5.56)       | 0.79(5.58)       | 0.78(5.67)       | 0.75(5.76)       | 0.70(6.03)       | 0.47(7.91)       |

NOTE: Efficiency is measured as the average proportion of iterations that are events over 20 repetitions of the sampler. Average runtime, in seconds, of each sampler is shown in parentheses.

and data generated using \( \theta = (-1.25, 0.5, -0.4, -0.4, -0.4)^T \).

We take a Gaussian prior \( \theta_j \sim N(0, 1) \) for \( j = 1, \ldots, 5 \). As the correlation is increased the thinning becomes more challenging. We investigate the performance of CC-PDMP for increasing polynomial bounds with increasing correlation \( \rho \).

To date only linear bounds (polynomial order 1) for logistic regression have been used for thinning (Bierkens et al. 2018; Bouchard-Côté, Vollmer, and Doucet 2018). These are based on the bound \( |\phi''(a, y)| \leq 1/4 \). Table 1 shows the efficiency for thinning using polynomials of order 1–3 (using bounds \( |\phi''(a, y)| \leq 1/4, |\phi'''(a, y)| \leq 1/(6\sqrt{3}) \) and \( |\phi^{(4)}(a, y)| \leq 1/8 \). The linear 1st order bound matches the bound used in Bierkens et al. (2018) for logistic regression using the Zig-Zag sampler. Table 1 shows that higher order polynomial thinning facilitated through the CC-PDMP allows more efficient thinning. However, these bounds can incur higher computational cost as they require evaluation of the higher order derivatives. For this example an order 2 polynomial appears to have the tradeoff in computation and thinning efficiency giving the best performance. For the linear bounds we used a large fixed \( \tau_{\text{max}} \) to ensure the linear rate proposal is simulated on the interval \([0, \tau_{\text{max}}]\). For other polynomial orders we used the adaptive procedure described in the tuning section to select \( \tau_{\text{max}} \).

### 6.2. Comparison to Thinning via Super-Positioning

In this example we demonstrate the advantages of CC-PDMP when the event rate is additive. In particular, we compare thinning using the CC-PDMP approach with thinning using super-positioning as outlined in Section 3. Let \( \{y_i\}_{i=1}^n \) be the observed data with the following model

\[
y_i | \theta_i \sim \text{Poisson}(\exp(\theta_i)), \quad \theta_i \sim N(0, 1)
\]

independently for \( i = 1, \ldots, n \). The derivative of the potential is

\[
\partial_{\theta_j} U(\theta) = \theta_k - y_k + \exp(\theta_k) \quad \text{and} \quad f_k(t) = v_k \partial_{\theta_k} U(\theta + t\nu) = v_k (\theta_k + v_k t) - y_k v_k + v_k \exp(\theta_k + v_k t).
\]

This has the concave-convex decomposition \( f_k(t) = f_k(t) \) when \( v_k > 0 \) and \( f_k(t) = v_k (\theta_k + v_k t) - y_k v_k f_k(t) = v_k \exp(\theta_k + v_k t) \).
and total computation time (right) is averaged over 20 repeated runs. The BPS samplers were simulated for 1000 event times on each run.

Efficiency of thinning for the global BPS using thinning via super-positioning and CC-PDMP thinning with increasing dimension. Efficiency of the thinning (left) and total computation time (right) is averaged over 20 repeated runs. The BPS samplers were simulated for 1000 event times on each run.

when \( v_k < 0 \). The global BPS rate can be defined as

\[
\lambda(t) = \max \left\{ 0, \sum_{k=1}^{n} v_k \partial_{\theta_k} U(\theta + t\nu) \right\} = \max \left\{ 0, \sum_{k=1}^{n} f_k(t) \right\},
\]

which has concave-convex decomposition defined by the decompositions of the individual \( f_k \) functions. In contrast, thinning via super-positioning involves simulating an event time for the linear and exponential terms of the rate individually. The proposed event time is the minimum of all simulated times. This approach to simulating event times for exponential-like terms, leading to more efficient thinning proposals. This is seen in Figure 4 where the thinning efficiency remains roughly constant with increasing dimension.

6.3. Local Methods

Local PDMP methods take advantage of the conditional independence between parameters. Consider the following extension to the previous Poisson example

\[
y_i \mid \theta_i \sim \text{Poisson}(\exp(\theta_i)) \quad \text{independently for } i = 1, \ldots, n
\]

\[
\theta_i \mid \theta_{i-1} \sim \mathcal{N}(0, 1/(1 - \rho^2))
\]

where we fix \( \rho = 0.5 \). The prior on \( \theta \) corresponds to an AR(1) process. The partial derivative of the potential is

\[
\partial_{\theta_k} U(\theta) = \begin{cases} 
(1 + \rho^2)\theta_k - \rho \theta_{k+1} - y_k + \exp(\theta_k) & k = 1 \\
(1 + \rho^2)\theta_k - \rho \theta_{k-1} - \rho \theta_k - y_k + \exp(\theta_k) & 1 < k < n \\
\theta_k - \rho \theta_{k-1} - y_k + \exp(\theta_k) & k = n
\end{cases}
\]

which has the same linear and exponential form as the rate in Section 6.2 so we can use an analogous concave-convex decomposition for this rate. In this section we consider local PDMP implementations. The CC-PDMP implementation facilitates simple construction of thinning bounds for local PDMP factorizations. For the partition \( \mathcal{S} = \{S_1, \ldots, S_F\} \) the local rate for factor \( f \) is

\[
\lambda_f(z_f) = \max \left\{ 0, \sum_{k \in S_f} v_k \partial_{\theta_k} U(\theta_f) \right\} = \max \left\{ 0, \sum_{k \in S_f} f_k(t) \right\},
\]

for \( f = 1, \ldots, F \). In this section we consider a number of local decompositions of the form \( \mathcal{S}^{(i)} = \{S_1, \ldots, S_F\} \) where \( S_1 = \{1, 2, \ldots, j\}, S_2 = \{j + 1, j + 2, \ldots, 2j\}, \ldots, S_F = \{n - j, \ldots, n - 1, n\} \). We assume \( n \) is divisible by \( F \) so each \( S_f \) contains \( j \) elements. For these decompositions the conditional independence between parameters gives the following neighbors \( \mathcal{N} = \{N_1, \ldots, N_F\} \) where

\[
N_f = \begin{cases} 
\{f, f + 1\} & f = 1 \\
\{f - 1, f, f + 1\} & 1 < f < F \\
\{f - 1, f\} & f = F
\end{cases}
\]
This local decomposition offers computational advantages since updating a single rate requires recalculation of at most three rates regardless of the dimension of the problem. However, this computational efficiency comes at a loss of statistical efficiency. In particular, a global PDMP will be able to move further in stochastic time than local methods before requiring a new event simulation as the global rate function will lower bound the local rate. The overall efficiency of the PDMP method should therefore be considered in terms of both its computational and statistical efficiency.

In Figure 5 we measure overall efficiency as the effective sample size (ESS) per second, the computational efficiency as the number of events or MCMC samples per second and the statistical efficiency as the ESS per MCMC sample or PDMP event. This allows a clear description of the scaling performance of local methods in comparison with alternative state-of-the-art MCMC methods. Linear trends on the log-log scale are stated in the figure legend and used to give approximate asymptotic scaling rates with dimension. We compare local PDMP methods with well-tuned implementations of Metropolis adjusted Langevin algorithm (MALA) and Hamiltonian Monte Carlo (HMC). MALA is tuned to obtain acceptance rate approximately equal to 0.5 and HMC tuned to scale with acceptance probability approximately equal to 0.6. These are close to the optimal values from the MCMC optimal scaling literature (Roberts and Rosenthal 1998; Beskos et al. 2013). For MALA this involves scaling the variance in the proposal and for HMC this involves adjusting the number of leapfrog steps per iteration. These implementations are in line with theoretical scaling results when the parameters of the distribution are independent.

The computational efficiency of the local PDMP methods remains (approximately) constant with increasing dimension, while for MALA and BPS the computational efficiency scales approximately as $O(d^{-1})$. The computational efficiency for HMC scales at a rate worse than $O(d^{-1})$ due to the increase in the number of leapfrog iterations. The statistical efficiency for the local methods improves with larger local factors and is highest for the Global BPS. The statistical efficiency for MALA drops at a rate roughly equal to $O(d^{-1/3})$. The Zig-Zag and local BPS methods attain the best overall efficiency scaling rates around $O(d)$ or better. However, it is clear that there is a tradeoff between the statistical efficiency of larger local factors and the increased computational cost. This decomposition can be thought of as an additional choice for PDMP implementation that can easily be tuned using the CC-PDMP approach for thinning. Despite a poorer scaling with dimension, HMC remains competitive with local methods to a very high dimension.

7. Conclusion

PDMP-based samplers have shown advantages over traditional MCMC samplers, but their use has been limited by the perceived difficulty in simulating the event times. We have introduced CC-PDMP as a general approach that can simulate the PDMP provided we can specify a concave-convex decomposition of the event rate. This method has broad applicability, enables simple implementation of local PDMP methods, and empirically outperforms alternative simulation approaches. Additional generalizations of the CC-PDMP approach are also possible by making use of other ideas for adaptive rejection sampling. For example, to bound the concave term, $f_\cap(t)$, of the rate function we require that the derivative $f_\cap'(t)$ is known on the interval $t \in [0, t_{\text{max}}]$. However, we could instead use a looser bound that does not require this derivative information based on the adaptive rejection bounds proposed in Gilks (1992). If the rate function is particularly computationally intensive, a lower bound based on the abscissae where $f_\cap$ and $f_\cup$ are evaluated may be used to perform early rejection in the thinning. Additional generalizations may also be found in the difference of convex functions programming literature (Le Thi and Pham Dinh 2018).

A general approach to facilitate thinning using a concave-convex decomposition was described in Section 4. This approach advocates finding an upper-bounding polynomial approximation for the rate function. This upper-bounding polynomial may be thinned using the concave-convex adaptive
thinning as described in Proposition 1. It is natural to consider an approximate version of our algorithm where this polynomial is estimated via interpolation. Without controlling for the potential error in the polynomial the overall algorithm would be biased. Recently Corbella, Spencer, and Roberts (2022) have proposed an automatic Zig-Zag implementation which uses optimization methods to obtain a valid linear thinning bound. Future research may consider how to automate the concave-convex thinning procedure for higher order polynomial thinning based on this new work. This may enable more efficient automatic PDMP-based samplers. Finally while our CC-PDMP approach has been illustrated only on PDMPs with linear dynamics the approach could be used more generally on PDMPs with nonlinear dynamics, such as the Boomerang sampler of Bierkens et al. (2020). The only requirement is that the rate function can be bounded by a function with a concave-convex decomposition.

**Supplementary Materials**

Appendix: Appendix containing additional coding and mathematical details for the work (appendix.pdf)

R-package: R-package “ccpdmp” containing code to perform methods described in the article. The package also contains all data generating code to generate the examples in the article. (ccpdmp.zip, zipped file)

Readme: A markdown file giving instructions for installing the package and replicating the results (README.md)

**Funding**

This research was supported by EPSRC grant EP/R018561 and EP/R034710/1

**ORCID**

Matthew Sutton https://orcid.org/0000-0003-1900-2957

**References**

Beskos, A., Pillai, N., Roberts, G., maria Sanz-serna, J., and Stuart, A. (2013), “Optimal Tuning of the Hybrid Monte Carlo Algorithm,” Bernoulli, 19, 1501–1534. [1434]

Bierkens, J., and Roberts, G. (2017), “A Piecewise Deterministic Scaling Limit of Lifted Metropolis–Hastings in the Curie–Weiss Model,” The Annals of Applied Probability, 27, 846–882. [1425]

Bierkens, J., Bouchard-Côté, A., Doucet, A., Duncan, A. B., Fearnhead, P., Lienart, T., Roberts, G., and Vollmer, S. J. (2018), “Piecewise Deterministic Markov Processes for Scalable Monte Carlo on Restricted Domains,” Statistics & Probability Letters, 136, 148–154. [1432]

Bierkens, J., Fearnhead, P., and Roberts, G. (2019), “The zig-zag Process and Super-Efficient Sampling for Bayesian Analysis of Big Data,” The Annals of Statistics, 47, 1288–1320. [1425]

Bierkens, J., Grazzi, S., Kamatani, K., and Roberts, G. (2020), “The Boomerang Sampler,” in: International Conference on Machine Learning, pp. 908–918, PMLR. [1425,1435]

Bierkens, J., Grazzi, S., van der Meulen, F., and Schauer, M. (2021), “A Piecewise Deterministic Monte Carlo Method for Diffusion Bridges,” Statistics and Computing, 31, 37. [1427]

Bouchard-Côté, A., Vollmer, S. J., and Doucet, A. (2018), “The Bouncy Particle Sampler: A Nonreversible Rejection-Free Markov Chain Monte Carlo Method,” Journal of the American Statistical Association, 113, 855–867. [1425,1427,1428,1432,1433]

Corbella, A., Spencer, S. E. F., and Roberts, G. O. (2022), “Automatic zig-zag Sampling in Practice,” Statistics and Computing, 32, 107. [1435]

Cotter, S., House, T., and Panagia, F. (2020), “The NuZZ: Numerical zigzag Sampling for General Models,” arXiv:2003.03636. [1426]

Davis, M. (1993), Markov Models and Optimization, New York: Chapman Hall. [1425]

Diaconis, P., Holmes, S., and Neal, R. M. (2000), “Analysis of a Nonreversible Markov Chain Sampler,” Annals of Applied Probability, 10, 726–752. [1425]

Fearnhead, P., Bierkens, J., Pollock, M., and Roberts, G. O. (2018), “Piecewise Deterministic Markov Processes for Continuous-Time Monte Carlo,” Statistical Science, 33, 386–412. [1425]

Gilks, W. R. (1992), “Derivative-Free Adaptive Rejection Sampling for Gibbs Sampling,” Bayesian Statistics, 4, 641–649. [1434]

Goan E., Perrin D., Mengersen K., and Fookes C. (2023), “Piecewise Deterministic Markov Processes for Bayesian Neural Networks,” arXiv:2302.08724. [1426]

Goldman, J. V., and Singh, S. S. (2021), “Spatiotemporal Blocking of the Bouncy Particle Sampler for Efficient Inference in State-Space Models,” Statistics and Computing, 31, 68. [1425,1426]

Görür, D., and Teh, Y. W. (2011), “Convex-Converse Adaptive Rejection Sampling,” Journal of Computational and Graphical Statistics, 20, 670–691. [1425,1429,1430,1431]

Hartman, P. (1959), “On Functions Representable as a Difference of Convex Functions,” Pacific Journal of Mathematics, 9, 707–713. [1430]

Kingman, J. F. C. (1992), Poisson Processes, Oxford: Clarendon Press. [1428]

Le Thi, H. A., and Pham Dinh, T. (2012), “DC Programming and DCA: Thirty Years of Developments,” Mathematical Programming. A Publication of the Mathematical Programming Society, 169, 5–68. [1434]

Lewis, P. A. W., and Shedler, G. S. (1979), “Simulation of Nonhomogeneous Poisson Processes by Thinning,” Naval Research Logistics Quarterly, 26, 403–413. [1425,1427]

Michel, M., Kapfer, S. C., and Krauth, W. (2014), “Generalized Event-Chain Monte Carlo: Constructing Rejection-Free Global-Balance Algorithms from Infinitesimal Steps,” The Journal of Chemical Physics, 140, 054116. [1425]

Michel, M., Durmus, A., and Sénac, S. (2020), “Forward Event-Chain Monte Carlo: Fast Sampling by Randomness Control in Irreversible Markov Chains,” Journal of Computational and Graphical Statistics, 29, 689–702. [1425]

Pakman, A., Gilboa, D., Carlson, D., and Paninski, L. (2017), “Stochastic Bouncy Particle Sampler,” in Proceedings of the 34th International Conference on Machine Learning (Vol. 70), pp. 2741–2750, PMLR. [1426]

Peters, E. A., and de With, G. (2012), “Rejection-Free Monte Carlo Sampling for General Potentials,” Physical Review E, 85, 026703. [1425]

Roberts, G. O., and Rosenthal, J. S. (1998), “Optimal Scaling of Discrete Markov Chain SAMplers,” The Annals of Applied Probability, 8, 1509–1532. [1426]

Sen, D., Sachs, M., Lu, J., and Dunson, D. (2020), “Efficient Posterior Sampling for High-Dimensional Imbalanced Logistic Regression,” Biometrika, 107, 1005–1012. [1428]

Stone, M. H. (1948), “The Generalized Weierstrass Approximation Theorem,” Mathematics Magazine, 21, 237–254. [1429]

Vanetti, P., Bouchard-Côté, A., Deligiannidis, G., and Doucet, A. (2017), “Piecewise-Deterministic Markov Chain Monte Carlo,” arXiv:1705.05296. [1425,1433]

Wu, C., and Robert, C. P. (2020), “Coordinate Sampler: A Non-Reversible Gibbs-like MCMC Sampler,” Statistics and Computing, 30, 721–730. [1425,1428]