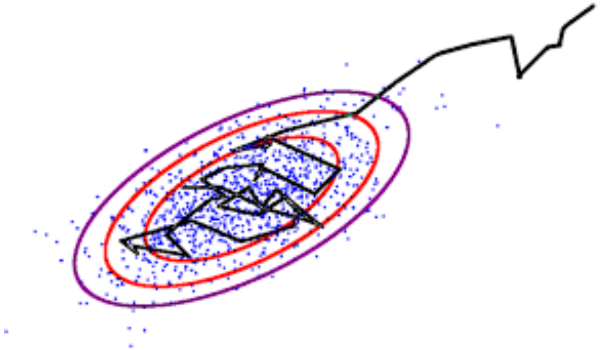
Asymptotically Optimal Exact Minibatch Metropolis-Hastings

Ruqi Zhang A. Feder Cooper Christopher De Sa

Cornell University

Metropolis-Hasting (MH)



Given: dataset $\{x_i\}_{i=1}^N$ and the prior $p(\theta)$

Goal: sample from the posterior

$$\pi(\theta) \propto \exp\left(-\sum_{i=1}^{N} U_i(\theta)\right), \text{where } U_i(\theta) = -\log p(x_i|\theta) - \frac{1}{N}\log p(\theta)$$

Algorithm

- Generate a proposal $\theta' \sim q(\theta'|\theta)$
- Accept it with probability $a(\theta,\theta') = \min\left(1, \exp\left(\sum_{i=1}^{N}(U_i(\theta) - U_i(\theta'))\right) \cdot \frac{q(\theta|\theta')}{q(\theta'|\theta)}\right)$

Challenge: the accept/reject step is costly when dataset is large!

Minibatch to scale MH

Approximate $a(\theta, \theta')$ based on a minibatch of data

Two classes (whether the posterior is preserved):

Inexact methods

- Pros: mild assumptions
- Cons: asymptotic bias

[Korattikara et. al.,2014; Bardenet et.al., 2014; Seita et. al.,2017......]

Exact methods

- Pros: no bias
- Cons: strong assumptions; low efficiency

[Maclaurin et.al., 2015; Cornish et.al., 2019; Zhang et.al. 2019 ......]

Which to use?

Is it important to be exact?

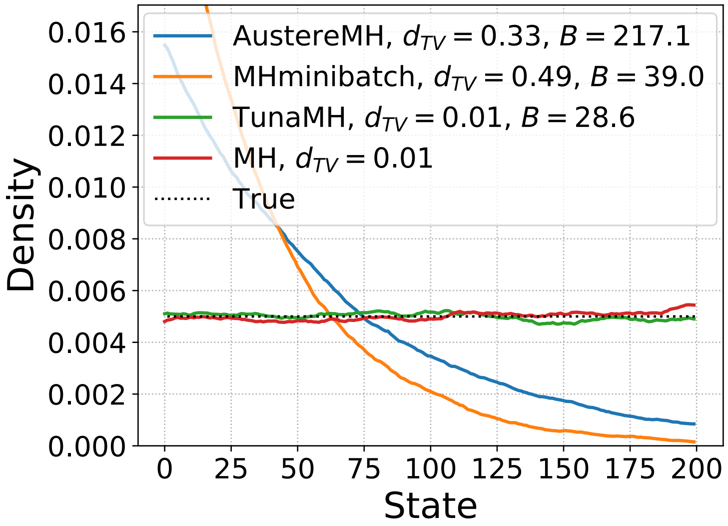
Part I: Inexact methods are unreliable

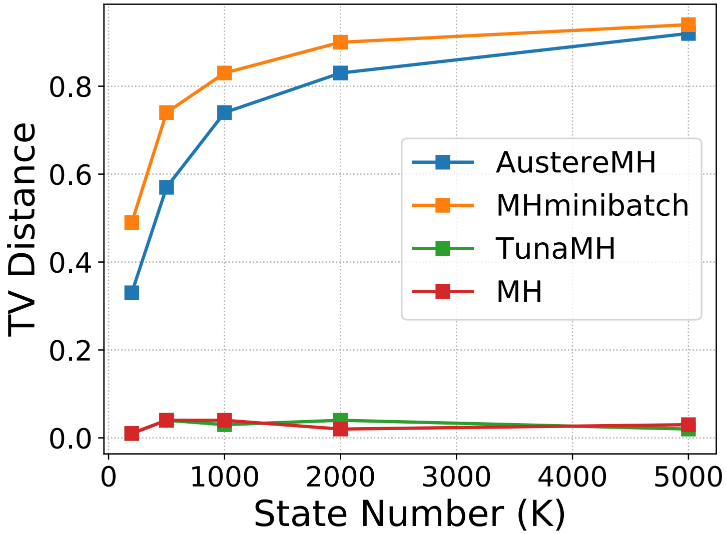
Theorem: we prove that the stationary distribution of any inexact method can be arbitrarily far from the posterior (in terms of TV distance and KL divergence)

Takeaway

- Any inexact minibatch MH can be arbitrarily wrong
- We should use exact methods

Random Walk Example





- The stationary distributions of inexact methods (AustereMH and MHminibatch) diverge significantly from the true distribution
- Divergence can be arbitrarily large

So we should use exact methods for correctness guarantee
But...existing methods are restrictive and inefficient

Part II: Our exact method: TunaMH

- Mild assumptions on the posterior
  (local bounds on the energy: $|U_i(\theta) - U_i(\theta')| \leq c_i M(\theta, \theta')$)

- Convergence rate guarantee: at most a constant factor slower than standard (i.e. full-batch) MH

- A tunable trade-off between scalability (batch size) and efficiency (convergence rate)

Our exact method: TunaMH

Algorithm 2 TunaMH

given: hyperparameter χ → A dial for convergence rate and batch size trade-off

loop

propose $\theta' \sim q(\cdot|\theta)$ and **compute** $M(\theta, \theta')$ → generate a proposal

▷ Form minibatch $\mathcal{I}$

sample $B \sim \text{Poisson}\left(\chi C^2 M^2(\theta, \theta') + C M(\theta, \theta')\right)$ → sample batch size B

initialize minibatch indices $\mathcal{I} \leftarrow \emptyset$ (an initially empty multiset)

for $b \in \{1, \ldots, B\}$ **do**

sample i_b such that $\mathbf{P}(i_b = i) = c_i / C$, for $i = 1 \ldots N$

with probability $\frac{\chi c_{i_b} C M^2(\theta,\theta') + \frac{1}{2}(U_{i_b}(\theta') - U_{i_b}(\theta) + c_{i_b} M(\theta,\theta'))}{\chi c_{i_b} C M^2(\theta,\theta') + c_{i_b} M(\theta,\theta')}$ **add** i_b to $\mathcal{I}$

end for

} select data to the minibatch

▷ Accept/reject step based on minibatch $\mathcal{I}$

compute MH ratio $r \leftarrow \exp\left(2 \sum_{i \in \mathcal{I}} \operatorname{artanh}\left(\frac{U_i(\theta) - U_i(\theta')}{c_i M(\theta,\theta')(1 + 2\chi C M(\theta,\theta'))}\right)\right) \cdot \frac{q(\theta'|\theta)}{q(\theta|\theta')}$

with probability $\min(1, r)$, set $\theta \leftarrow \theta'$

end loop

Compute acceptance rate based on the minibatch

Is it possible to develop a **better** exact method?

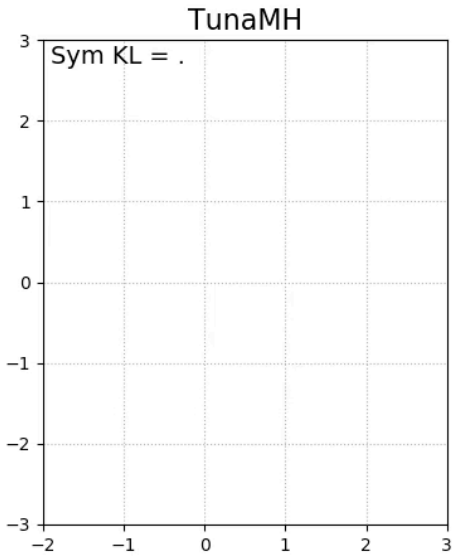
Part III: How efficient can an exact minibatch MH be?

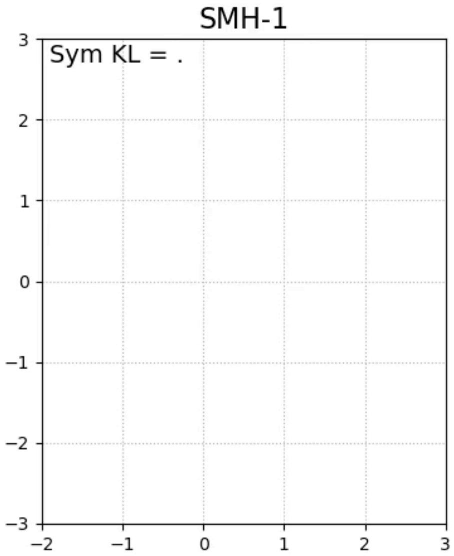
Theorem: given a target convergence rate, we prove a lower bound on the required batch size for any exact minibatch MH

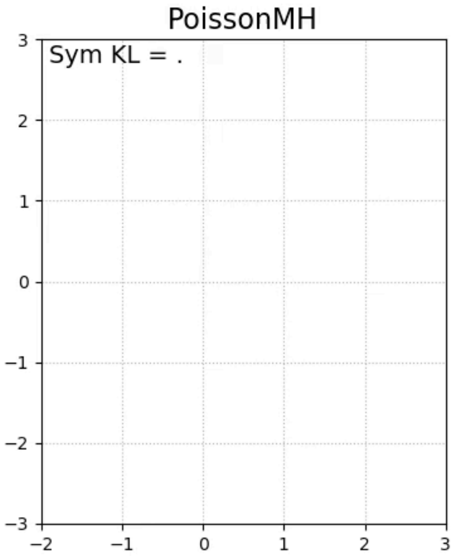
Takeaway

- The first theorem to provide a ceiling for the performance of exact minibatch MH
- TunaMH is asymptotically optimal in the batch size

Gaussian Mixture

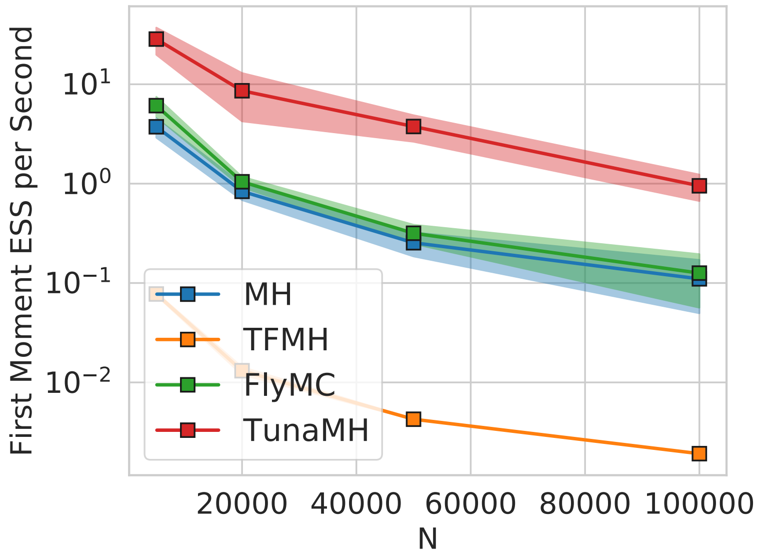


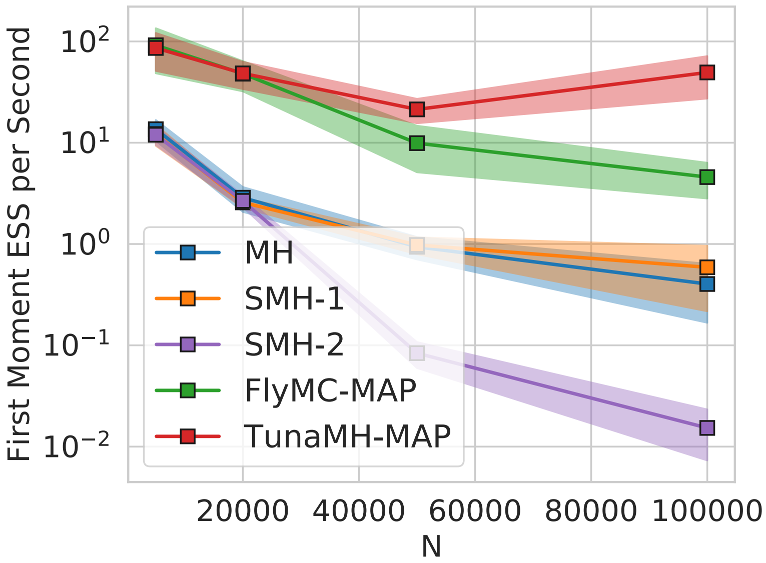




- Compared to SOTA exact methods, TunaMH is the fastest to converge

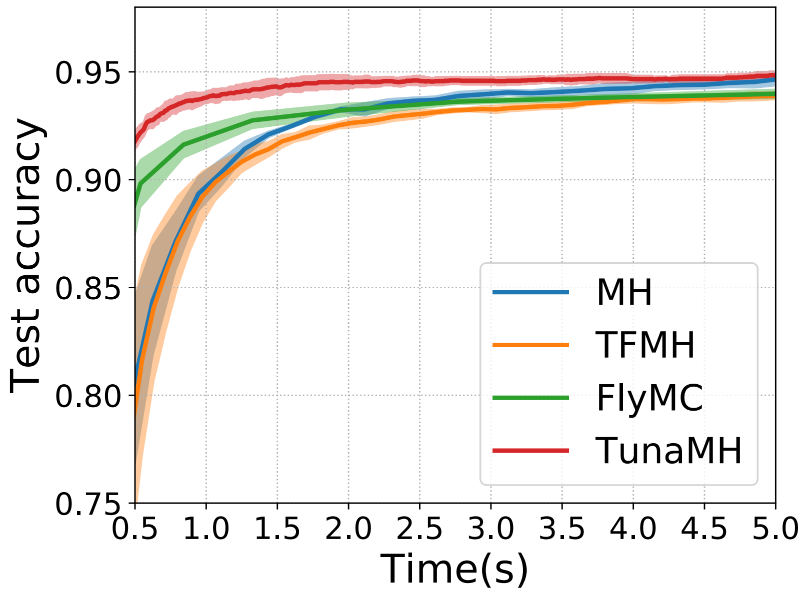
Robust Linear Regression





- TunaMH has the highest ESS per second, even compared to methods using MAP/control variates

Logistic Regression on MNIST



- TunaMH has the highest test accuracy given time

Summary

- We should use exact methods, as any inexact minibatch MH can perform arbitrarily poorly
- TunaMH is an exact minibatch MH, with mild assumptions and convergence rate guarantees
- We provide a lower bound on the batch size of any exact methods and show TunaMH is asymptotically optimal
- Empirical demonstration on common tasks

arXiv.org https://arxiv.org/abs/2006.11677

https://github.com/ruqizhang/tunamh