

1 We thank all five reviewers for their detailed and incisive feedback. We respond in order below:

2 **R1: Comparison with inexact methods** Aligning with prior exact papers [10, 18], we focus on comparisons with exact
3 methods. We tested AustereMH [16], an inexact method, on robust linear regression in Section 5.1 with $N = 5000$. We
4 computed the MSE between estimated and true parameters. MH, TunaMH and AustereMH obtained MSE 0.149, 0.15
5 and 1.19 respectively, indicating inexact method error can be large on typical problems. We added this to the Appendix.

6 **R1: More difficult problem** Our next step is to apply TunaMH to more difficult problems, such as a threshold testing
7 with a large real-world dataset [Pierson, et al. Fast Threshold Tests for Detecting Discrimination, AISTATS 2018.]

8 **R1: Confusion about $M(\theta, \theta')$** We added additional explanation here.

9 **R2 : What if $\mathbf{E}[B] > N \mathbf{E}[B]$** is typically $\ll N$, and can be decreased using small step sizes. If $\mathbf{E}[B] > N$, we can
10 simply use standard MH in that iteration, similar to [10]. This does not affect the properties of TunaMH. We added this.

11 **R3: Theorem 1 (1)** We believe the example in Theorem 1 is natural. The target (marginal) distribution is uniform,
12 the proposal is a random walk, and the state space is integers on $[0, \dots, K - 1]$ (If $X = 0$, the example can just be
13 a random walk of k without augmenting z). This is one of the simplest types of Markov chains and is encountered
14 in many real applications. (2) Theorem 1 is compatible with the TV bound of MHSUBLHD (Proposition 3.2 in [6]).
15 Proposition 3.2 assumes P_{MH} has a bounded mixing time; it is well-known that this produces a TV bound for any kernel
16 by coupling [17]. Our theorem doesn't have this assumption; it suggests that for MHSUBLHD with given user-specified
17 error, there exists a target distribution and proposal satisfying Theorem 1, on which P_{MH} either does not have bounded
18 mixing time or the mixing time is large enough such that the TV bound is greater than δ . We added this to the Appendix.

19 **R3: PDMPs; Examples of U, C, M** We have (1) weakened the language and added the suggested reference; (2) added
20 an example for logistic regression.

21 **R3: Theorem 2** Given the spectral gap bound, we can immediately show that TunaMH inherits geometric ergodicity
22 from MH and obtain an asymptotic variance bound — a known result [17]. In fact, the spectral gap bound and the
23 asymptotic variance bound are equivalent, so the impact of χ on the variance is similar to that on the spectral gap. We
24 unfortunately do not understand the suggestion for the corollary.

25 **R3: Theorem 3 impact** Similar to Theorem 1, we constructed a random walk example over two states — a simple and
26 natural problem. The impact is 3-fold: it (1) provides an upper bound on performance for algorithms of Algorithm 1's
27 form; (2) shows the upper bound is achievable (e.g. TunaMH); (3) suggests directions for developing new algorithms.
28 To be significantly faster than TunaMH, we either need more assumptions about the problem or new stateful algorithms.

29 **R3: Asymptotics in M and C** In Theorem 3 M and C are the only problem parameters; these two values determine a
30 problem's lowest possible batch size. Thus, asymptotic optimality in M and C indicates asymptotic optimality in any
31 problem parameter (including N). When explicitly assuming the relation of N and M, C (as in prior work [10]), in
32 Corollary 1 we show how the bound depends on N and TunaMH is asymptotically optimal in N .

33 **R3: Claim about control variates** We empirically observe SMH with control variates does not work well in high
34 dimension. We conjecture this is because the batch size bound (Eq.13 in [10]) becomes looser quickly as d increases.

35 **R3&R5: Results in low dimension** We ran experiments (Section 5.1) with $d = 10$ and $N = 10^5$. The ESS/S for SMH,
36 FlyMC, TunaMH are 0.02, 0.75, & 1.7; SMH-1, SMH-2, FlyMC-MAP and TunaMH-MAP are 174.7, 5969.5, 730.8, &
37 730.1 respectively. This suggests TunaMH is significantly better without MAP/control variates. With MAP/control
38 variates, TunaMH is better than SMH-1, similar to FlyMC and worse than SMH-2. We will add this to the Appendix.

39 **R3: \mathcal{A} in Appendix** \mathcal{A} denotes the SubSMH of the minibatch MH method in question. The expectation in L425, L426
40 and L433 are all taken over the randomness in SubSMH. We have clarified this.

41 **R4: Table 2; other distributions for batch size** (1) We fixed Table 2 to make χ values clearer. (2) We were unable to
42 identify a distribution other than Poisson to get nice results. It is possible one does exist, which we leave to future work.

43 **R5: Assumption 1** For consistency we use the same language as in prior work [10] to describe the bounds in Assumption
44 1 and in FlyMC. One can get the required bound in FlyMC by triangle inequality, but the tightness of the bound highly
45 depends on the reference point θ_0 . Thus it is typically harder to get a tight bound on $|U_i(\theta)|$ than $|U_i(\theta) - U_i(\theta')|$. We
46 agreed that Assumption 1 is still a strong requirement. We have clarified these points in the paper.

47 **R5: Reproducibility; Clarity about SMH and FMH** We (1) will release the code and associated documentation upon
48 publication; (2) have cleaned up our SMH/FMH language to be more precise.

49 **R5: "Form minibatch" steps** Thinning is a well-known technique (Lewis and Shedler, 1979) used in many papers [7,
50 8, 24]. We developed the "Form minibatch" steps, including ejection, by replacing the global bounds with the local
51 bounds in Algorithm 4 in the Appendix of [24] (explained in Appendix C). We clarified this and now also cite [10].