Can Stochastic Gradient Langevin Dynamics Provide Differential Privacy for Deep Learning?

Omitted for submission
Omitted for submission

***Abstract*—Bayesian learning via Stochastic Gradient Langevin Dynamics (SGLD) has been suggested for differentially private learning. While previous research provides differential privacy bounds for SGLD at the initial steps of the algorithm or when close to convergence, the question of what differential privacy guarantees can be made in between remains unanswered. This interim region is of great importance, especially for Bayesian neural networks, as it is hard to guarantee convergence to the posterior. This paper shows that using SGLD might result in unbounded privacy loss for this interim region, even when sampling from the posterior is as differentially private as desired.**

***Index Terms*—Differential Privacy, Stochastic Gradient Langevin Dynamics, Bayesian Inference, Deep Learning**

I. Introduction

Machine learning models, specifically deep neural networks, achieve state-of-the-art results in various fields such as computer vision, natural language processing, and signal processing (e.g., [1]–[3]). Training these models requires data, which in some domains, e.g., healthcare and finance, can include sensitive information that should not be made public. Unfortunately, information from the training data can, in some cases, be extracted from the trained model [4], [5]. One common approach to handle this issue is Differential Privacy (DP). DP framework ensures that the distribution of the training output would remain approximately the same when we switch one of the training examples, thus ensuring we cannot extract information specific to a unique individual.

As privacy is usually obtained by adding random noise, it is natural to investigate whether Bayesian inference, which uses a distribution over models, can yield private predictions. Previous works have shown that sampling from the posterior is differentially private under certain mild conditions [6]–[8]. The main disadvantage of this method is that sampling from the posterior can be challenging. The posterior generally does not have a closed-form solution, so iterative methods such as Markov Chain Monte Carlo (MCMC), whose sample distribution converges to the posterior, are commonly used. While theoretical bounds on the convergence of MCMC methods for non-convex problems exist [9], they usually require an infeasible number of steps to guarantee convergence in practice.

Stochastic Gradient Langevin Dynamics (SGLD) [10] is a popular MCMC algorithm, as it avoids the accept-reject step. There are good reasons to believe that this specific sampling algorithm can provide private predictions. First, SGLD returns an approximate sample from the posterior, which can be private. Second, the SGLD process of stochastic gradient descent with Gaussian noise mirrors the common Gaussian mechanism in DP.

Previous work [6] gives two separate privacy analyses related to SGLD: The first is based on the Gaussian mechanism and the Advanced Composition theorem [11]. Therefore, it only applies to a limited number of steps and is not connected to Bayesian sampling.

The second is for approximate sampling from the Bayesian posterior, which is only relevant when SGLD nearly converges. Neither of these results is suitable for deep learning and many other problems: one would limit the model's accuracy, and the other is unattainable in a reasonable time. Consequently, the privacy properties of SGLD in the interim region (between these two private sections) remain unknown even though they are of great interest.

Our Contributions:

- We provide a rigorous analysis of a counter-example based on a Bayesian linear regression problem, showing that approximate sampling using SGLD might result in unbounded loss of privacy in the interim region, even if sampling from the posterior is as private as desired.
- We further empirically show that SGLD can result in nonprivate models.

These results imply that special care should be given when using SGLD for private predictions, especially for problems for which it is infeasible to guarantee convergence.

II. Related Work

Several previous works investigate the connection between Bayesian inference and differential privacy [6]–[8], [12]–[15]. None of these papers guarantees SGLD differential privacy in the interim region. However, The closest work to ours is [6], which specifically investigates stochastic MCMC algorithms such as SGLD. As mentioned, its analysis only covers the initial phase and when approximate convergence is achieved.

A concurrent work [16] studies the DP guarantees of SGLD under assumptions of convex, Lipschitz, and smooth loss function on a convex set with a bounded diameter. Under these assumptions, [16] shows a bound over the privacy, which plateaus after a certain number of iterations.

When training machine learning models in a differentially private way via Stochastic Gradient Descent, a common practice is to apply the Gaussian Mechanism by clipping the gradients of the loss with respect to the weights and adding a matching noise (see [17], for example). SGLD learning step

resembles the resulting learning step but does not include gradients clipping. Reference [15] suggests incorporating gradient clipping in the SGLD step. However, clipping the gradients changes the algorithm properties, and it is not obvious if it converges to the posterior. As such, we do not consider it SGLD. Reference [6] circumvents this issue by assuming the log-likelihood of the model is Lipschitz continuous.

Another related work on the privacy of SGLD is [18], although they investigate a weaker type of privacy called membership privacy.

As many of the Bayesian methods' privacy bounds require sampling from the posterior, if SGLD is to be used, it requires non-asymptotic convergence bounds. Reference [19] provides non-asymptotic bounds on the approximation error for a smooth and log-concave target distribution by Langevin Monte Carlo. Reference [20] studies the non-asymptotic bounds on the error of approximating a target density p^* where $\log p^*$ is smooth and strongly convex.

For the non-convex setting, [21] shows non-asymptotic bounds on the 2-Wasserstein distance between SGLD and the invariant distribution solving Itô stochastic differential equation. However, the 2-Wasserstein metric is ill-suited for differential privacy - it is easy to create two distributions with 2-Wasserstein distance as small as desired but with disjoint support.

Total Variation (for details about Total Variation, see [22]) is a more suitable distance for working with differential privacy. Reference [9] examines a target distribution p^*, which is strongly log-concave outside of a region of radius R, and where $-\ln p^*$ is L-Lipschitz. They provided a bound on the number of steps needed for the Total Variation distance between the distribution at the final step and p^* to be smaller than ϵ. This bound is proportional to $O(e^{32LR^2}\frac{d}{\epsilon^2})$, where d is the model dimension. This result suggests that it is impractical to run SGLD until convergence is guaranteed in the non-convex setting.

A conclusion from this work is that basing the differential privacy of SGLD on the proximity to the posterior is impractical for non-convex settings.

III. Background

A. Differential Privacy

Differential Privacy [11], [23]–[25] is a definition and a framework that enables performing data analysis on a dataset while reducing one's risk posed by disclosing its personal data to the dataset. In a nutshell, an algorithm is differentially private if it does not change its output distribution by much due to a single record change in its dataset. Approximate Differential Privacy, Definition III.1, is an extension of pure Differential Privacy, where pure differential privacy is Approximate Differential Privacy with $\delta = 0$.

Definition III.1. *Approximate Differential Privacy: A randomized algorithm $f : \mathcal{D} \to Range(f)$ is (ϵ, δ)-differentially private if $\forall S \subseteq Range(f)$ and $\forall D, \hat{D} \in \mathcal{D} : d(D, \hat{D}) \leq 1$ eq. 1 holds, where d is the distance between D and $\hat{D}$. $D, \hat{D}$ are called neighboring datasets, and while the metric can change per application, Hamming distance is typically used.*

$$p(f(D) \in S) \leq \exp(\epsilon) p(f(\hat{D}) \in S) + \delta \tag{1}$$

Rényi Divergence [26], which generalizes the Kullback-Leibler divergence, is defined as follows:

Definition III.2. *Rényi Divergence: For two probability distributions Z and Q, the Réyni divergence of order $\nu > 1$ is*

$$\mathrm{D}_\nu(Z||Q) \triangleq \frac{1}{\nu - 1} \log \mathbb{E}_{x \sim Q} \left[\left(\frac{Z(x)}{Q(x)} \right)^\nu \right].$$

Reference [27] suggested a relaxation of differential privacy based on the Rényi divergence, termed Rényi Differential Privacy:

Definition III.3. *(ν, ϵ)-RDP: A randomized algorithm $f : \mathcal{D} \to Range(f)$ is said to have ϵ-Rényi differential privacy of order ν, or (ν, ϵ)-RDP in short, if for any neighbouring datasets $D, \hat{D} \in \mathcal{D}$ eq. 2 holds, where D_ν is Rényi divergence of order ν.*

$$\mathrm{D}_\nu \left(f(D) || f(\hat{D}) \right) \leq \epsilon \tag{2}$$

In this paper, we utilize the fact that RDP has a closed-form solution when both $f(D)$ and $f(\hat{D})$ are Normal distributions (see [28] and the proof of Lemma A.1 in the appendix for details).

By Proposition III.4, RDP guarantees can be translated into approximate differential privacy guarantees.

Proposition III.4. *From RDP to (ϵ, δ)-DP [27]: If f is (ν, ϵ)-RDP, it also satisfies $(\epsilon + \frac{\log \frac{1}{\delta}}{\nu - 1}, \delta)$-differential privacy for any $0 < \delta < 1$.*

B. Stochastic Gradient Langevin Dynamics

Stochastic Gradient Langevin Dynamics (SGLD) is an MCMC method commonly used for Bayesian Inference [10]. Given a Bayesian model parameterized by θ, a dataset $D = \{x_i, y_i\}_{i=1}^n$, a prior distribution $p(\theta)$, the likelihood function $p(y_i|\theta, x_i)$, and a batch size b, SGLD can be used for approximate sampling from the posterior $p(\theta|D)$. The update step of SGLD is shown in eq. 3, where θ_j is the parameter vector at step j, and η_j is the step size at step j. SGLD can be seen as a Stochastic Gradient Descent with Gaussian noise, where the variance of the noise is calibrated to the step size.

$$\begin{aligned} &\theta_{j+1} = \\ &\theta_j + \frac{\eta_j}{2} \left(\nabla_{\theta_j} \ln p(\theta_j) + \frac{n}{b} \sum_{i=1}^{b} \nabla_{\theta_j} \ln p(y_{i_j}|\theta_j, x_{i_j}) \right) + \sqrt{\eta_j} \xi_j \\ &i_j \sim uniform\{1, ..., n\} \\ &\xi_j \sim \mathcal{N}(0, 1) \end{aligned} \tag{3}$$

A common practice in deep learning is to use *cyclic* Stochastic Gradient Descent. This modification to SGD first randomly shuffles the dataset samples and then cyclically uses the samples in this order. For optimization, there is empirical evidence

that it works as well or better than SGD with reshuffling, and it was conjectured that it converges at a faster rate [29]. Cyclic-SGLD[1] is the analog of cyclic-SGD for SGLD, where the difference is the use of the SGLD step instead of the SGD step. For simplicity, we will consider cyclic-SGLD in this work. While this assumption simplifies the proof, we expect the general behavior to be equivalent.

IV. Theoretical Results

Our goal is to prove that even when sampling from the posterior is as private as desired, approximate sampling using SGLD can be as nonprivate as desired in the interim region. This requires analysing the distribution of SGLD in the interim region, which is hard in the general case. To circumvent this difficulty, we investigate the Bayesian linear regression problem, where the distributions are a mixture of Gaussians and thus have closed-form expressions. Our result is summarized in Theorem IV.1.

Theorem IV.1. $\forall\ 0 < \delta < 0.5$ *and* $\forall \epsilon, \epsilon' > 0$*, there exists a number* T*, a domain, and a Bayesian inference problem for which a single sample from the posterior distribution is* (ϵ, δ) *differentially private. However, performing approximate sampling by running SGLD for* T *steps is not* (ϵ', δ) *differentially private.*

The Bayesian inference problem, mentioned in Theorem IV.1, refers to sampling from the posterior for a dataset and a model defined by likelihood and prior distributions. The specific model and dataset we will analyse in our work are defined in eq. 4 and eq. 5, respectively.

An example of the behavior described by the theorem is depicted in Fig. 1. In this case, given the model defined in eq. 4 and the domain defined in eq. 5, a single sample from the posterior is $(\epsilon = 0.5, \delta = 0.001)$ DP; however, approximate sampling from the posterior by running SGLD for 48 epochs ($T = 48\cdot$ dataset size) is not $(\epsilon' = 38, \delta = 0.001)$ DP (further details for the figure are provided below).

As Theorem IV.1 allows ϵ' to be as big as desired and ϵ to be as small as desired, a corollary of Theorem IV.1 is that we could always find a problem for which the posterior is (ϵ, δ) differentially private, but there will be a step in which SGLD will result in an unbounded loss of privacy. Therefore, SGLD alone can not provide any privacy guarantees in the interim region, even if the posterior is private.

Theorem IV.1 is presented and proved for a fixed and equal δ for both the posterior and the SGLD privacy analysis. This is done for simplicity; however, the proof could be augmented to prove a lower bound on SGLD privacy for all $\epsilon' > 0$ and $0 < \delta' < 0.25$ (i.e., approximately sampling via SGLD is not (ϵ', δ')-DP for all $\epsilon' > 0$ and $0 < \delta' < 0.25$).

[1]Cyclic SGLD, which cycles through examples, should be distinguished from cSGLD [30], which uses a cyclic step size schedule.

To prove our theorem, we consider a Bayesian regression problem for a 1D linear model with Gaussian noise, as defined in eq. 4.

$$\begin{aligned} &y = \theta x + \xi \\ &\xi \sim \mathcal{N}(0, \beta^{-1}) \\ &\theta \sim \mathcal{N}(0, \alpha^{-1}) \\ &p(y|x, \theta) \sim \mathcal{N}(\theta x, \beta^{-1}) \end{aligned} \tag{4}$$

We assume our input domain is

$$\begin{aligned} &\mathcal{D}(n, \gamma_1, x_h, x_l, c) = \\ &\{(x_i, y_i) || \frac{y_i}{x_i} - c| \leq n^{\gamma_1};\ x_i, y_i, c, \gamma_1 \in \mathbb{R}_{>0}; \\ &n \in \mathbb{Z}_{>0}; x_l \leq x_i \leq x_h\}_{i=1}^{n} \end{aligned} \tag{5}$$

, where $x_h^2\beta > 3$ and $\gamma_1 < \frac{1}{2}$. The constants n, c, x_l, x_h, and γ_1 are parameters of the problem (c, x_l, x_h, and γ_1 are used, together with the dataset size - n, to bound the dataset samples to a chosen region). For every ϵ, ϵ', and δ, we will show the existence of parameters n, c, x_l, x_h, γ_1 values that have the privacy properties required to prove Theorem IV.1. The restrictions on the dataset simplify the proof but are a bit unnatural as it assumes we approximately know c, the parameter we are trying to estimate. Later we show in subsection IV-C that they can be replaced with a Propose-Test-Release phase.

For simplicity, we will address the problem of sampling (or approximately sampling via SGLD) from the posterior for the model described in eq. 4 and a dataset from domain $\mathcal{D}(n, \gamma_1, x_h, x_l, c)$ as a *Bayesian linear regression problem* on domain $\mathcal{D}(n, \gamma_1, x_h, x_l, c)$. This problem has a closed-form solution for both the posterior distribution and the distribution at each SGLD step, thus enabling us to get tight bounds on the differential privacy in each case.

In essence, our proof shows that for a big enough n, sampling from the posterior is (ϵ, δ) differentially private, with $\epsilon \sim \mathcal{O}(\frac{c^2}{n^3})$. However, for the same problem instance, there exists an SGLD step in which releasing a sample will not be (ϵ', δ) differentially private for $\epsilon' = \Omega(\frac{c^2}{n^2})$. Therefore, for problem instances where $c \sim \mathcal{O}(n^{\frac{3}{2}}\sqrt{\epsilon})$ and n is big enough, sampling from the posterior will be (ϵ, δ) differentially private, while there will be an SGLD step in which releasing a sample will not be (ϵ', δ) differentially private for $\epsilon' = \Omega(n\epsilon)$. We note that the bounds dependend on δ, but since we are using a fixed and equal δ for both the posterior and SGLD privacy analysis, we omit it from the bounds for simplicity.

Fig. 1 depicts a lower bound over the DP of SGLD for the *Bayesian Linear Regression Problem* on domain $\mathcal{D}(n, \gamma_1, x_h, x_l, c)$. The values of n, γ_1, x_h, x_l, c ensure $(0.5, 0.001)$-DP when sampling from the posterior. However, we can see that sampling via SGLD in the interim region causes a significant privacy breach (sampling via SGLD at epoch 48 is not $(38, 0.001)$-DP). For the derivation of the lower bound in Fig. 1, see subsection A-E in the appendix.

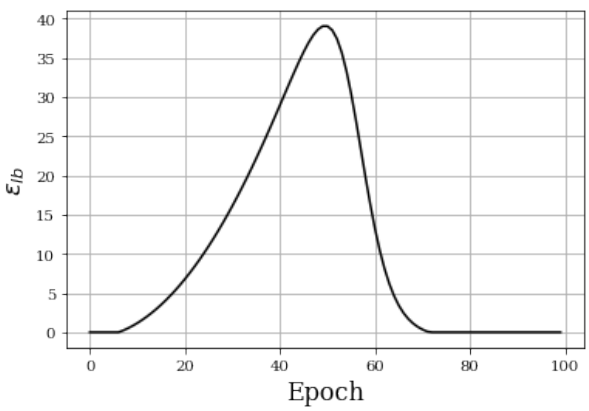

Fig. 1. Lower bound over the DP of SGLD for the model defined in eq. 4 and a dataset from domain $\mathcal{D}(n, \gamma_1, x_h, x_l, c)$ (defined in eq. 5), given that $\delta = 0.001$. The domain parameters values are $n = 17864389, \alpha = 2, \beta = 1, \gamma_1 = 0.1, x_h = 1.8, x_l = 0.9$, which ensure $(0.5, 0.001)$-DP when sampling from the posterior.

A. Posterior Sampling Privacy

To prove Theorem IV.1, we need to show that sampling from the posterior is private, while there is an SGLD sample that is not private at some intermediate step. In this section, we prove the first part - that a single sample from the posterior for the *Bayesian linear regression problem* on domain $\mathcal{D}(n, \gamma_1, x_h, x_l, c)$ is differentially private.

We begin by using a well-known result for the closed-form solution of the posterior distribution for a Bayesian linear regression problem (see [31] for further details). By incorporating the parameters of our problem in this result, we get Lemma IV.2.

Lemma IV.2. *The posterior distribution for the model defined in eq. 4 on dataset $D = \{(y_i, x_i)\}_{i=1}^n$ is*

$$p(\theta|D) = \mathcal{N}(\theta; \mu, \sigma^2);$$
$$\mu = \frac{\sum_{i=1}^n x_i y_i \beta}{\alpha + \sum_{i=1}^n x_i^2 \beta}; \sigma^2 = \frac{1}{\alpha + \sum_{i=1}^n x_i^2 \beta}, \tag{6}$$

As the posterior distribution is a Normal distribution, the Rényi divergence between every two posterior distributions has a closed-form solution. For two neighbouring datasets, $D, \hat{D} \in \mathcal{D}(n, \gamma_1, x_h, x_l, c)$, and matching posterior distributions $p(\theta|D) = \mathcal{N}(\theta; \mu, \sigma^2), p(\theta|\hat{D}) = \mathcal{N}(\theta; \hat{\mu}, \hat{\sigma}^2)$, the Rényi divergence of order ν is

$$\mathrm{D}_\nu \left(p\left(\theta|D\right) || p(\theta|\hat{D}) \right) =$$
$$\ln \frac{\sigma}{\hat{\sigma}} + \frac{1}{2}(\nu - 1) \ln \frac{\hat{\sigma}^2}{(\sigma^2)_\nu^*} + \frac{1}{2} \frac{\nu(\mu - \hat{\mu})^2}{(\sigma^2)_\nu^*},$$
$$(\sigma^2)_\nu^* = \nu \hat{\sigma}^2 + (1 - \nu)\sigma^2.$$

By bounding $\mathrm{D}_\nu(p(\theta|D)||p(\theta|\hat{D}))$ for every two neighbouring datasets, one can prove RDP. The first and second terms of $\mathrm{D}_\nu(p(\theta|D)||p(\theta|\hat{D}))$ can be bounded by $\mathcal{O}(\frac{1}{n})$ using Taylor Theorem and the fact that the natural logarithm is monotonically increasing. By using direct computation, the third term can be bounded by $\mathcal{O}(\frac{1}{n^{1-2\gamma_1}}) + \mathcal{O}(\frac{c+n^{\gamma_1}}{n^{2-\gamma_1}}) + \mathcal{O}(\frac{(c+n^{\gamma_1})^2}{n^3})$. This gives way to Lemma IV.3. For the full proof, see subsection A-B in the appendix.

Lemma IV.3. *For the Bayesian linear regression problem on domain $\mathcal{D}(n, \gamma_1, x_h, x_l, c)$, such that $n > \max\{1 + 10\frac{x_h^2}{x_l^2}\frac{\nu}{\beta}, 1 + \nu\frac{x_h^2}{x_l^2}\}$, one sample from the posterior is (ν, ϵ_1)-Rényi differentially private, and ϵ_1 is*

$$\epsilon_1 = \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}\left(\frac{1}{n^{1-2\gamma_1}}\right) + \mathcal{O}\left(\frac{c + n^{\gamma_1}}{n^{2-\gamma_1}}\right) + \mathcal{O}\left(\frac{(c + n^{\gamma_1})^2}{n^3}\right). \tag{7}$$

We can show that for $c >> n^{1+\gamma_1}$, each of the terms in the right hand side of eq. 7 is bounded by $\mathcal{O}(\frac{c^2}{n^3})$. The first term is trivially bounded by $\mathcal{O}(\frac{1}{n})$. For the second term, noticing that $n^{2\gamma_1 - 1} = \frac{n^{2(1+\gamma_1)}}{n^3} < \frac{c^2}{n^3}$, we get that it is bounded by $\mathcal{O}(\frac{c^2}{n^3})$. As $c >> n^{\gamma_1}$, the third term is bounded by $\mathcal{O}(\frac{cn^{\gamma_1}}{n^2})$, and since $\frac{cn^{\gamma_1}}{n^2} = \frac{cn^{1+\gamma_1}}{n^3} < \frac{c^2}{n^3}$, the term is bounded by $\mathcal{O}(\frac{c^2}{n^3})$. Lastly, since $c >> n^{\gamma_1}$ the last term is bounded by $\mathcal{O}(\frac{c^2}{n^3})$.

Translating the Rényi differential privacy guarantees of Lemma IV.3 into approximate differential privacy terms can be done according to Lemma III.4, which gives Lemma IV.4.

Lemma IV.4. *With the conditions of Lemma IV.3, one sample from the posterior is $(\epsilon_1 + \frac{\ln(\frac{1}{\delta})}{\nu - 1}, \delta)$ differentially private.*

By choosing ν such that $\frac{\ln(\frac{1}{\delta})}{\nu - 1} < \frac{\epsilon}{2}$ and then choosing n big enough such that $\epsilon_1 < \frac{\epsilon}{2}$, we get that the posterior is (ϵ, δ) differentially private.

B. Stochastic Gradient Langevin Dynamics Privacy

To complete the proof of Theorem IV.1, we need to show that given a *Bayesian linear regression problem* on domain $\mathcal{D}(n, \gamma_1, x_h, x_l, c)$, even if one sample from the posterior is (ϵ, δ) differentially private, it does not guarantee SGLD is private in the interim region. In order to do so, this section will first consider the loss of privacy when using SGLD for the *Bayesian linear regression problem* on domain $\mathcal{D}(n, \gamma_1, x_h, x_l, c)$ and then, together with the results of section IV-A, will prove Theorem IV.1.

In order to show that SGLD is not differentially private after initial steps and before convergence, it is enough to find two neighbouring datasets for which the loss in privacy is as big as desired after a certain number of steps. We define neighbouring datasets $D_1, D_2 \in \mathcal{D}(n, \gamma_1, x_h, x_l, c)$ in eq. 8 and consider the *Bayesian linear regression problem* on D_1 and D_2 with a learning rate: $\eta = \frac{2}{(\alpha + n x_h^2 \beta)^2}$.

$$D_1 = \{(x_i, y_i) : x_i = x_h, y_i = c \cdot x_h\}_{i=1}^n$$
$$D_2 = \{(x_i, y_i) : x_i = x_h, y_i = c \cdot x_h\}_{i=1}^{n-1} \cup \{(\frac{x_h}{2}, c \cdot \frac{x_h}{2})\} \tag{8}$$

A closed-form solution for the distribution at each step enables us to get a tight lower bound over the differential privacy loss when approximately sampling via SGLD at each step. For

dataset D_1, the solution is a Normal distribution. For dataset D_2, different shuffling of samples produces different Gaussian distributions, therefore giving a mixture of Gaussians.

We look at cyclic-SGLD with a batch size of 1 and mark by $\theta_j, \hat{\theta}_j$ the samples on the j'th SGLD step when using datasets D_1 and D_2 accordingly. Since D_1 samples are all equal, the update step of the cyclic-SGLD is the same for every step (with different noise generated for each step). This update-step contains only multiplication by a scalar, addition of a scalar, and addition of Gaussian noise, therefore, together with a conjugate prior results in Normal distribution for θ_j: $\mathcal{N}(\theta_j; \mu_j, \sigma_j^2)$, where $\mu_j, \sigma_j \in \mathbb{R}$.

For D_2, there is only one sample different from the rest. We mark by r the index in which this sample is used in the cyclic-SGLD and call this order r-order. Note that there are only n (n is the dataset size, defined in 5) different values for r and, as such, effectively only n different samples orders. Since every order of samples is chosen with the same probability, r is distributed uniformly in $\{1, .., n\}$. We mark by $\hat{\theta}_j^r$ the sample on the j'th SGLD step when using r-order. Since, for a given order, $\hat{\theta}_j^r$ is formed by a series of multiplications by a scalar, addition of scalar, and addition of Gaussian noise, and since the prior is also Gaussian, then $\hat{\theta}_j^r$ is distributed Normally, $\mathcal{N}(\hat{\theta}_j^r; \hat{\mu}_j^r, (\hat{\sigma}_j^r)^2)$, where $\hat{\mu}_j^r, \hat{\sigma}_j^r \in \mathbb{R}$. As r is distributed uniformly, $\hat{\theta}_j$ distribution mass is distributed evenly between all $\hat{\theta}_j^r$, resulting in a mixture of Gaussians.

Intuitively what will happen is that each Gaussian component, $\hat{\theta}_j$ as well as θ_j, will move towards a similar Gaussian posterior. However, at each epoch, $\hat{\theta}_j$ will drag a bit behind because a single gradient in one of the batches will be smaller. While this gap can be quite small, for large n, the Gaussians are very peaked with very small standard deviations; thus, they are separate enough that we can easily distinguish between the two distributions.

According to the approximate differential privacy definition (Definition III.1), it is enough to find one set, S, such that $p(\theta_j \in S) > e^{\epsilon} p(\hat{\theta}_j \in S) + \delta$, to prove that releasing θ_j is not (ϵ, δ) private. We choose $S = \{s | s > \mu_j\}$ at some step j that we will define later on.

To show that $p(\theta_j \in S) > e^{\epsilon} p(\hat{\theta}_j \in S) + \delta$, we first note that as the Gaussian θ_j is symmetric, it is clear that $p(\theta_j > \mu_j) = 1/2$. Now we turn our focus to upper bounding $p(\hat{\theta}_j > \mu_j)$. This can be done using Chernoff bound, as stated in Lemma IV.5.

Lemma IV.5. $p(\hat{\theta}_j > \mu_j) \leq \frac{1}{n} \sum_{r=1}^{n} \exp(-\frac{(\mu_j - \hat{\mu}_j^r)^2}{2(\hat{\sigma}_j^r)^2})$.

To bound $p(\hat{\theta}_j > \mu_j)$ using Lemma IV.5, we first need to lower bound $\frac{(\mu_j - \hat{\mu}_j^r)^2}{(\hat{\sigma}_j^r)^2}$ for a certain step. This is done in Lemma IV.6.

Lemma IV.6. $\exists k \in \mathbb{Z}_{>0}$ *such that* $\frac{(\mu_{(k+1)n} - \hat{\mu}_{(k+1)n}^r)^2}{(\hat{\sigma}_{(k+1)n}^r)^2} = \Omega(\frac{c^2}{n^2})$, *for big enough* n.

To prove Lemma IV.6, we first find closed-form solutions for $\hat{\theta}_{(k+1)n}^r$, $\theta_{(k+1)n}$ distributions (Lemma A.2). Using the closed-form solutions, we find a lower bound over $(\mu_{(k+1)n} - \hat{\mu}_{(k+1)n}^r)^2$ as a function of k, which applies for all k (Lemma A.4). To upper bound $(\hat{\sigma}_{(k+1)n}^r)^2$, we find an approximation to the epoch in which the data and prior effect on the variance is approximately equal, marked $\dot{k}$. We choose $(\lceil \dot{k} \rceil + 1)n$ as the step in which we will consider the privacy loss and show that $(\hat{\sigma}_{(\lceil \dot{k} \rceil + 1)n}^r)^2$ is upper bounded at this step (Lemma A.6). Using the lower bound on the difference in means and the upper bound on the variance, Lemma IV.6 is proved.

By using the lower bound from Lemma IV.6 in Lemma IV.5, we get Lemma IV.7.

Lemma IV.7. *For the Bayesian linear regression problem over dataset D_1 and n big enough, $\exists T \in \mathbb{Z}_{>0}$ such that approximate sampling by running SGLD for T steps will not be (ϵ, δ) private for $\epsilon = \Omega(\frac{c^2}{n^2}), \delta < 0.5$.*

From Lemma IV.4, we see that sampling from the posterior is (ϵ, δ) differentially private for $\epsilon = \mathcal{O}(\frac{c^2}{n^3})$. From Lemma IV.7, we see that for SGLD, there exists a step in which releasing a sample will not be (ϵ', δ) differentially private for $\epsilon' = \Omega(\frac{c^2}{n^2})$. Therefore, for problem instances where $c = \mathcal{O}(n^{\frac{3}{2}} \sqrt{\epsilon})$, sampling from the posterior will be (ϵ, δ) differentially private. However, there will be an SGLD step in which releasing a sample will not be (ϵ', δ) differentially private for $\epsilon' = \Omega(n\epsilon)$. Since we can choose n to be big as desired, we can make the lower bound over ϵ' as big as we desire it to be. This completes the proof of Theorem IV.1.

C. Propose Test Sample

Our analysis of the posterior and SGLD is done on a restricted domain - $\mathcal{D}(n, \gamma_1, x_h, x_l, c)$. These restrictions over the dataset simplify the proof but are a bit unnatural as they assume we approximately know c, the parameter we are trying to estimate. This section shows that these restrictions could be replaced with a Propose-Test-Release phase [32] and common practices in data science.

When training a statistical model, it is common to first preprocess the data by restricting it to a bounded region and removing outliers. After the data is cleaned, the training process is performed. This is especially important in DP, as outliers can significantly increase the algorithm's sensitivity to a single data point and thus hamper privacy.

Informally, Algorithm 1 starts by clipping the input to the accepted range. It then estimates a weighted average of the ratio $\frac{y_i}{x_i}$ (line 16) and throws away outliers that deviate too much from it. The actual implementation of this notion is a bit more complicated because of the requirement to do so privately. Once the dataset is cleaned, Algorithm 1 privately verifies that the number of samples is big enough, so the sensitivity of $p(\theta|W)$ (where W is the cleaned dataset) to a single change in the dataset will be small, therefore making sampling from $p(\theta|W)$ (ϵ, δ) differentially private. This method is regarded as Propose-Test-Release, where we first propose a bound over the sensitivity, then test if the dataset holds this bound, and finally release the result if so.

Algorithm 1 Propose Test Sample

1: **Input:** $D = \{x_i, y_i\}_{i=1}^{n_1}$
2: **Parameters:** $\epsilon, \delta < 0.5, x_l > 0, x_h > x_l, \alpha > 0, \beta \geq \frac{3}{x_h^2}, \rho_1 \in (1, \frac{3}{2}), \rho_2 \in (0, \frac{1}{2}), \gamma_1 \in (\rho_2, \frac{1}{2}), n_1 \in \mathbb{Z}_{>0}$
3: **for** $i = 1, 2, \ldots, N$ **do**
4: $\quad x_i \leftarrow \max\{x_i, x_l\}$
5: $\quad x_i \leftarrow \min\{x_i, x_h\}$
6: $\quad y_i \leftarrow \max\{y_i, 0\}$
7: **end for**
8: $l_1 \leftarrow$ sample from Laplace$(0, \frac{1}{\epsilon})$
9: $\breve{n}_1 \leftarrow n_1 - \frac{1}{\epsilon} \log \frac{1}{2\delta} + l_1$
10: $V = \{x_i, y_i | \frac{y_i}{x_i} \leq \breve{n}_1^{\rho_1}\}$
11: $l_2 \leftarrow$ sample from Laplace$(0, \frac{1}{\epsilon})$
12: $n_2 \leftarrow |V| - \frac{1}{\epsilon} \log \frac{1}{2\delta} + l_2$
13: **if** $n_2 \leq 1$ **then**
14: $\quad$ return null
15: **end if**
16: $m \leftarrow \frac{\sum_{(x_i,y_i) \in V} x_i y_i}{\sum_{(x_i,y_i) \in V} x_i^2}$
17: $l_3 \leftarrow$ sample from Laplace$(0, \frac{1}{\epsilon} \breve{n}_1^{\rho_1} \frac{2(n_2-1)x_h^2 x_l^2 + x_h^4}{n_2(n_2-1)x_l^4})$
18: $\breve{m} \leftarrow m + l_3$
19: $W \leftarrow \{(x_i, y_i) : |\frac{y_i}{x_i} - \breve{m}| \leq n_2^{\rho_2}\}$
20: $l_4 \leftarrow$ sample from Laplace$(0, \frac{1}{\epsilon})$
21: $n_W \leftarrow |W| - \frac{1}{\epsilon} \log(\frac{1}{2\delta}) + l_4$
22: **if** $n_W < n_{min}$ **then**
23: $\quad$ return null
24: **end if**
25: return sample from $p(\theta|W)$

In eq. 33 in the appendix, we define n_{min} as the minimum size of W for which the algorithm will sample from $p(\theta|W)$ with high probability. We will show later on that this limit ensures that sampling from $p(\theta|W)$ is (ϵ, δ) differentially private.

We define $p(\theta|W)$ as the posterior of the 1D linear regression model defined in eq. 4 over dataset W. From Lemma IV.2, it follows that $p(\theta|W)$ has the form of

$$p(\theta|W) = \mathcal{N}(\theta; \mu, \sigma^2);$$
$$\mu = \frac{\sum_{(x_i,y_i) \in W} x_i y_i \beta}{\alpha + \sum_{(x_i,y_i) \in W} x_i^2 \beta}; \ \sigma^2 = \frac{1}{\alpha + \sum_{(x_i,y_i) \in W} x_i^2 \beta}.$$

Claim IV.8. *Algorithm 1 is $(5\epsilon, 2\delta)$ differentially private.*

By Claim C.12, lines 8-18 are $(3\epsilon, \delta)$ differentially private. By Corollary C.17, lines 19-25 are $(2\epsilon, \delta)$ differentially private given $\breve{m}$ and n_2. Therefore by the sequential composition theorem, the composition is $(5\epsilon, 2\delta)$ differentially private. The claim is proved by noticing that if lines 8-25 are private with respect to the updated dataset (after line 7), then they are also private for the original dataset.

Claim IV.9. *When replacing line 25 with approximate sampling via SGLD with step size $\eta = \frac{1}{(\alpha + n_1 x_h^2 \beta)^2}$, there exists $T(n_1) : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ such that the updated algorithm is not (ϵ, δ) differentially private $\forall \epsilon \in \mathbb{R}_{>0}, \delta < \frac{1}{6}$ if ran for $T(n_1)$ steps.*

Proof sketch (See appendix for full proof). We analyze a run of Algorithm 1 on the neighbouring datasets, D_3 and D_4, defined in eq. 9. First, note that when choosing $1 + \rho_2 > \rho_1$, the sensitivity of $\breve{m}$ grows slower than the bound over the distance $|\frac{y_i}{x_i} - \breve{m}|$ in n_1 for both of the datasets. Therefore, with high probability, for n_1 big enough, W will contain all the samples that meet the condition $\frac{y_i}{x_i} = m$. Consequently, with high probability, the algorithm will reach line 25, which, from our previous analysis over SGLD (see subsection IV-B) will cause an unbounded loss of privacy.

$$\begin{aligned} &\rho_1 > \rho_3 > 1 \\ &D_3 = \{(x_i, y_i) : x_i = x_h, y_i = n_1^{\rho_3} \cdot x_h\}_{i=1}^{n_1} \\ &D_4 = \{(x_i, y_i) : x_i = x_h, y_i = n_1^{\rho_3} \cdot x_h\}_{i=1}^{n_1 - 1} \cup \\ &\{(\frac{x_h}{2}, n_1^{\rho_3} \cdot \frac{x_h}{2})\} \end{aligned} \tag{9}$$

V. Empirical Evidence

We augment our theoretical analysis with an empirical study on privacy loss when training a deep neural network via SGLD. This study strengthens our claim that one should use SGLD with great care for private learning.

To empirically estimate SGLD's privacy, we attack it using a version of the *adversary instantiation* method described in [33], with some modifications to the method's details. In broad strokes, we train with SGLD a set of models on each of two neighbouring datasets, D and D'. Then we try to predict for each model on which dataset it was trained. If the algorithm is DP, it will be hard to distinguish which dataset was used to train the model, and the accuracy will be low. Concretely, by analyzing the prediction's false positive and false negative rates, we can deduce a lower bound over the training DP parameters - (ϵ, δ).

To create the neighboring dataset D', we replace one of the samples from D with a novel data point - (x^*, y^*). To show SGLD is not private, we need a sample, (x^*, y^*), such that the models that were not trained on it will misclassify it, but models trained on it will classify it correctly after a small number of epochs.

To create x^*, we first train M models, $\{m\}_{i=1}^M$, on dataset D. Then, we search for a sample in D, marked (x^0, y^0), such that $\{m\}_{i=1}^M$ agree on it's label: $\forall 1 \leq i, j \leq M : m_i(x^0) = m_j(x^0)$. We then use DeepFool [34] to alter the sample x^0 into x^* such that all the models will misclassify it with regard to their original prediction: $\forall 1 \leq i \leq M : m_i(x^0) \neq m_i(x^*)$. We set $y^* = y^0$ and $D' = D \setminus \{(x^0, y^0)\} \cup \{(x^*, y^*)\}$.

Given D and D', we generate a dataset, A_1, of models trained on D and D' with equal probability. We represent a model with parameters θ by four features, $p(y = y^*|x^*, \theta)$, $p(y = y^0|x^*, \theta)$, $p(y = y^*|x^0, \theta)$, $p(y = y^0|x^0, \theta)$, and train a simple linear classifier. Finally, we create a second independent test set, A_2, of models trained on D and D' with equal probability and estimate our classifier's false negative

(FN) and false positive (FP) rates using the examples from A_2.

A. Deducing lower bound over ϵ

To translate the attack results into DP parameters, we follow the analysis approach suggested by [35], [36] and extended by [33]. Without loss of generality, we define false positive as predicting dataset D', when dataset D was used for training a model, and false negative as vice versa. The probability for FP and FN are marked as P_{FP} and P_{FN}, respectively. According to [37], if an algorithm is (ϵ, δ)-DP, then the following inequalities hold:

$$\begin{aligned} P_{FP} + e^{\epsilon} P_{FN} &\geq 1 - \delta \\ P_{FN} + e^{\epsilon} P_{FP} &\geq 1 - \delta. \end{aligned} \tag{10}$$

These inequalities can easily be translated into a lower bound over ϵ,

$$\epsilon_{lb} \geq \max\left(\log \frac{1-\delta-P_{FP}}{P_{FN}}, \log \frac{1-\delta-P_{FN}}{P_{FP}}\right). \tag{11}$$

Since we can only estimate P_{FP} and P_{FN} empirically, we use confidence intervals to upper-bound them. The confidence intervals are calculated using the Clopper-Pearson method [38] on the attack's false positive and false negative rates. The resulting upper bounds, P_{FP}^{high} and P_{FN}^{high}, are then used to provide an empirical lower bound on ϵ with high probability:

$$\epsilon_{lb}^{emp} \geq \max\left(\log \frac{1-\delta-P_{FP}^{high}}{P_{FN}^{high}}, \log \frac{1-\delta-P_{FN}^{high}}{P_{FP}^{high}}\right). \tag{12}$$

It is important to note that this method can only prove that a model is not private. A low value for ϵ_{lb}^{emp} does not show the model is private, only that our attack failed to prove a lack of privacy.

B. Results

We performed our attack on the SGLD-based training process of a LeNet5 [39], trained on the MNIST dataset [39]. We tested a learning rate of 0.001^2 with a batch size of 4. We trained a different classifier for each epoch to find a lower bound on the DP at each epoch.

When using Clopper-Pearson [38] confidence intervals, the resulting upper bounds (P_{FP}^{high} and P_{FN}^{high}) are limited by the number of experiments conducted, which limits the maximum ϵ_{lb}^{emp}. We used 500 models to train the classifier (i.e., $|A_1| = 500$) and evaluated the attack on 500 models (i.e., $|A_2| = 500$), which limits ϵ_{lb}^{emp} to a maximum of 4.89.

Fig. 2 depicts lower bounds over ϵ given $\delta = 10^{-5}$, with a confidence value of 90%, i.e., $P(\epsilon \geq \epsilon_{lb}^{emp}) \geq 0.9$, as well as the accuracy of the network, as a function of the number of epochs.

It should be emphasized that we show a lower bound over ϵ. As such, even a small value is sufficient to show a significant breach of privacy. For example, a lower bound of $(\delta = 10^{-5}, \epsilon = 0.95)$ allows a classifier to identify on which dataset a model was trained with an accuracy of 90% ($P_{FP}^{high} = P_{FN}^{high} = 0.1$).

[2]Effective learning rate after multiplication by SGLD's normalization factor, i.e. $\eta \frac{n}{2b}$. See SGLD step in eq. 3 for details.

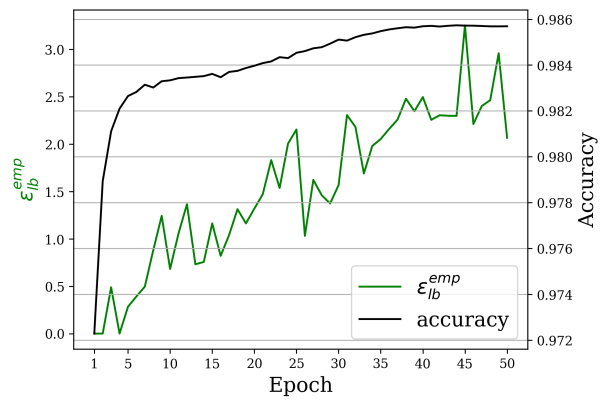

Fig. 2. A lower bound over the Differential Privacy of the LeNet5, SGLD-based training process over MNIST for a given $\delta = 10^{-5}$, a learning rate of 0.001, and a batch size of 4.

In appendix E, we show the results for SGLD with clipped gradients. We see that clipping the gradients protects the algorithm from our attack. As mentioned, an implementation that clips the gradients diverges from SGLD and, as such, has different sampling properties. Indeed, the results show that this version's accuracy degrades by 8% compared to SGLD.

References

[1] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Springer International Publishing, 2020, pp. 213–229.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423

[3] E. Balevi and J. G. Andrews, "Wideband channel estimation with a generative adversarial network," *IEEE Transactions on Wireless Communications*, vol. 20, no. 5, pp. 3049–3060, 2021.

[4] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1322–1333. [Online]. Available: https://doi.org/10.1145/2810103.2813677

[5] N. Carlini, F. Tramèr, E. Wallace, M. Jagielski, A. Herbert-Voss, K. Lee, A. Roberts, T. B. Brown, D. Song, Ú. Erlingsson, A. Oprea, and C. Raffel, "Extracting training data from large language models," in *USENIX Security Symposium*, 2021.

[6] Y.-X. Wang, S. Fienberg, and A. Smola, "Privacy for free: Posterior sampling and stochastic gradient monte carlo," in *Proceedings of the 32nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 37. Lille, France: PMLR, 07–09 Jul 2015, pp. 2493–2502. [Online]. Available: https://proceedings.mlr.press/v37/wangg15.html

[7] J. R. Foulds, J. Geumlek, M. Welling, and K. Chaudhuri, "On the theory and practice of privacy-preserving bayesian data analysis," in *Uncertainty in Artificial Intelligence, UAI*, 2016.

[8] C. Dimitrakakis, B. Nelson, Z. Zhang, A. Mitrokotsa, and B. I. P. Rubinstein, "Differential privacy for bayesian inference through posterior sampling," *Journal of Machine Learning Research*, vol. 18, no. 11, pp. 1–39, 2017. [Online]. Available: http://jmlr.org/papers/v18/15-257.html

[9] Y.-A. Ma, Y. Chen, C. Jin, N. Flammarion, and M. I. Jordan, "Sampling can be faster than optimization," *Proceedings of the National Academy of Sciences*, vol. 116, no. 42, pp. 20 881–20 885, 2019. [Online]. Available: https://www.pnas.org/content/116/42/20881

[10] M. Welling and Y. Teh, "Bayesian learning via stochastic gradient langevin dynamics," in *ICML*, 2011.

[11] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3–4, p. 211–407, Aug. 2014. [Online]. Available: https://doi.org/10.1561/0400000042

[12] Z. Zhang, B. I. P. Rubinstein, and C. Dimitrakakis, "On the differential privacy of bayesian inference," in *AAAI Conference on Artificial Intelligence*, 2016.

[13] J. Geumlek, S. Song, and K. Chaudhuri, "Renyi differential privacy mechanisms for posterior sampling," in *Advances in Neural Information Processing NeurIPS*, 2017.

[14] A. Ganesh and K. Talwar, "Faster differentially private samplers via rényi divergence analysis of discretized langevin mcmc," *ArXiv*, vol. abs/2010.14658, 2020.

[15] B. Li, C. Chen, H. Liu, and L. Carin, "On connecting stochastic gradient mcmc and differential privacy," in *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, vol. 89. PMLR, 16–18 Apr 2019, pp. 557–566. [Online]. Available: https://proceedings.mlr.press/v89/li19a.html

[16] J. M. Altschuler and K. Talwar, "Privacy of noisy stochastic gradient descent: More iterations without more privacy loss," 2022, unpublished. [Online]. Available: https://arxiv.org/abs/2205.13710

[17] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 308–318. [Online]. Available: https://doi.org/10.1145/2976749.2978318

[18] B. Wu, C. Chen, S. Zhao, C. Chen, Y. Yao, G. Sun, L. Wang, X. Zhang, and J. Zhou, "Characterizing membership privacy in stochastic gradient langevin dynamics," in *AAAI Conference on Artificial Intelligence*, 2020.

[19] A. Dalalyan, "Theoretical guarantees for approximate sampling from smooth and log-concave densities," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 79, 12 2014.

[20] X. Cheng and P. Bartlett, "Convergence of langevin mcmc in kl-divergence," in *ALT*, 2018.

[21] M. Raginsky, A. Rakhlin, and M. Telgarsky, "Non-convex learning via stochastic gradient langevin dynamics: a nonasymptotic analysis," in *Proceedings of the 2017 Conference on Learning Theory*, ser. Proceedings of Machine Learning Research, S. Kale and O. Shamir, Eds., vol. 65. PMLR, 07–10 Jul 2017, pp. 1674–1703. [Online]. Available: https://proceedings.mlr.press/v65/raginsky17a.html

[22] A. B. Tsybakov, *Introduction to Nonparametric Estimation*, 1st ed. Springer Publishing Company, Incorporated, 2008.

[23] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography*, S. Halevi and T. Rabin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 265–284.

[24] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Advances in Cryptology - EUROCRYPT 2006, 25th Annual International Conference on the Theory and Applications of Cryptographic Techniques*, ser. Lecture Notes in Computer Science, vol. 4004. Springer, 2006, pp. 486–503. [Online]. Available: https://iacr.org/archive/eurocrypt2006/40040493/40040493.pdf

[25] C. Dwork, "A firm foundation for private data analysis," *Commun. ACM*, vol. 54, no. 1, p. 86–95, Jan. 2011. [Online]. Available: https://doi.org/10.1145/1866739.1866758

[26] A. Rényi, "On measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. University of California Press, 1961, pp. 547–561.

[27] I. Mironov, "Renyi differential privacy," *CoRR*, vol. abs/1702.07476, 2017. [Online]. Available: http://arxiv.org/abs/1702.07476

[28] M. Gil, F. Alajaji, and T. Linder, "Rényi divergence measures for commonly used univariate continuous distributions," *Information Sciences*, vol. 249, pp. 124–131, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025513004441

[29] C. Yun, S. Sra, and A. Jadbabaie, "Open problem: Can single-shuffle SGD be better than reshuffling SGD and gd?" in *Conference on Learning Theory, COLT*, 2021.

[30] R. Zhang, C. Li, J. Zhang, C. Chen, and A. G. Wilson, "Cyclical stochastic gradient MCMC for bayesian deep learning," *CoRR*, vol. abs/1902.03932, 2019. [Online]. Available: http://arxiv.org/abs/1902.03932

[31] C. Bishop, *Pattern Recognition and Machine Learning*, ser. Information Science and Statistics. Springer-Verlag New York, 2006.

[32] C. Dwork and J. Lei, "Differential privacy and robust statistics," in *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, ser. STOC '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 371–380. [Online]. Available: https://doi.org/10.1145/1536414.1536466

[33] M. Nasr, S. Song, A. Thakurta, N. Papernot, and N. Carlini, "Adversary instantiation: Lower bounds for differentially private machine learning," *CoRR*, vol. abs/2101.04535, 2021. [Online]. Available: https://arxiv.org/abs/2101.04535

[34] S.-M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, "Deepfool: A simple and accurate method to fool deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[35] Z. Ding, Y. Wang, G. Wang, D. Zhang, and D. Kifer, "Detecting violations of differential privacy," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 475–489. [Online]. Available: https://doi.org/10.1145/3243734.3243818

[36] M. Jagielski, J. R. Ullman, and A. Oprea, "Auditing differentially private machine learning: How private is private sgd?" in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., 2020.

[37] P. Kairouz, S. Oh, and P. Viswanath, "The composition theorem for differential privacy," in *Proceedings of the 32nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, F. Bach and D. Blei, Eds., vol. 37. Lille, France: PMLR, 07–09 Jul 2015, pp. 1376–1385. [Online]. Available: https://proceedings.mlr.press/v37/kairouz15.html

[38] C. J. Clopper and E. S. Pearson, "The use of confidence or fiducial limits illustrated in the case of the binomial," *Biometrika*, vol. 26, pp. 404–413, 1934.

[39] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[40] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.

[41] A. Yousefpour, I. Shilov, A. Sablayrolles, D. Testuggine, K. Prasad, M. Malek, J. Nguyen, S. Ghosh, A. Bharadwaj, J. Zhao, G. Cormode, and I. Mironov, "Opacus: User-friendly differential privacy library in PyTorch," *arXiv preprint arXiv:2109.12298*, 2021.

APPENDIX A
SGLD AND POSTERIOR PRIVACY

Appendix A provides proofs for theorem IV.1 and the lemmas in subsections IV-B and IV-A. As such, it uses the notations defined in section IV and subsections IV-B and IV-A. To ease the proof's reading, we repeat these notations here.

α, β and θ are parameters of the linear model defined in eq. 13 (originally defined in 4), and $p(y|x)$ is the model likelihood.

$$\begin{aligned} &y = \theta x + \xi \\ &\xi \sim \mathcal{N}(0, \beta^{-1}) \\ &\theta \sim \mathcal{N}(0, \alpha^{-1}) \\ &p(y|x) \sim \mathcal{N}(\theta x, \beta^{-1}) \end{aligned} \tag{13}$$

x_h, x_l, c, n, and γ_1 are defined as part of domain $\mathcal{D}(n, \gamma_1, x_h, x_l, c)$ definition (originally defined in eq. 5):

$$\begin{aligned} &\mathcal{D}(n, \gamma_1, x_h, x_l, c) = \\ &\{(x_i, y_i) || \frac{y_i}{x_i} - c| \leq n^{\gamma_1};\ x_i, y_i, c, \gamma_1 \in \mathbb{R}_{>0}; \\ &n \in \mathbb{Z}_{>0}; x_l \leq x_i \leq x_h\}_{i=1}^{n} \end{aligned} \tag{14}$$

where $x_h^2 \beta > 3$ and $\gamma_1 < \frac{1}{2}$. The datasets $D_1, D_2 \in \mathcal{D}(n, \gamma_1, x_h, x_l, c)$ (originally defined in eq. 8) are defined in eq. 15.

$$\begin{aligned} &D_1 = \{x_i, y_i : x_i = x_h, y_i = c \cdot x_h\}_{i=1}^{n} \\ &D_2 = \{x_i, y_i : x_i = x_h, y_i = c \cdot x_h\}_{i=1}^{n-1} \cup \{\frac{x_h}{2}, c \cdot \frac{x_h}{2}\} \end{aligned} \tag{15}$$

The *Bayesian Linear Regression Problem* (originally defined in section IV) refers to the problem of sampling (or approximately sampling via SGLD) from the posterior for the model described in eq. 13.

We look at cyclic-SGLD with a batch size of 1 and mark by $\theta_j, \hat{\theta}_j$ the samples on the j'th SGLD step when using datasets D_1 and D_2 accordingly. $\mu_j, \sigma_j^2 \in \mathbb{R}$ are the mean and variance of θ_j. For D_2, there is only one sample different from the rest. We mark by r the index in which this sample is used in the cyclic-SGLD and call this order r-order. We mark by $\hat{\theta}_j^r$ the sample on the j'th SGLD step when using dataset D_2 and r-order. $\hat{\mu}_j^r, (\hat{\sigma}_j^r)^2 \in \mathbb{R}$ are the mean and variance of $\hat{\theta}_j^r$. $\eta = \frac{2}{(\alpha + n x_h^2 \beta)^2}$ is the SGLD learning rate. $\eta, \theta_j, \hat{\theta}_j, \mu_j, \hat{\mu}_j^r, \sigma_j, \hat{\sigma}_j^r$ were originally defined in subsection IV-B.

A. Theorem IV.1 Proof

Proof of Theorem IV.1. We first define several parameters used to configure the *Bayesian linear regression problem* on domain $\mathcal{D}(n, \gamma_1, x_h, x_l, c)$.

$$
\begin{aligned}
&\frac{1}{2} > \gamma_1 > 0;\ \frac{3}{2} > \gamma_2 > 1 + \gamma_1;\ x_l = \frac{x_h}{2} \\
&\nu_1 = \frac{2\ln(\frac{1}{\delta})}{\epsilon} + 1 \\
&n_1 = \max\{\frac{1}{2\alpha x_h^2\beta} - \frac{1}{x_h^2\beta}, \frac{\alpha}{x_h^2\beta}, \frac{\alpha}{x_h^2\beta}(e^{\frac{2}{x_h^2\beta}} - 2) + \frac{1}{2x^2\beta}\} + 1 \\
&n_2 = \max\{1 + \frac{x_h^2}{x_l^2}\frac{8}{\epsilon}, 1 + \nu\frac{x_h^2}{x_l^2}\left(1 + 8\frac{(\nu-1)}{\epsilon}\right), \left(\frac{16\nu\beta x_h^4}{\frac{9}{10}\epsilon x_l^2}\right)^{\frac{1}{1-2\gamma_1}}, \\
&\left(\frac{16\nu}{\epsilon}\cdot\frac{x_h^4\left(\alpha + x_h^2\beta\right)}{\frac{9}{10}x_l^4}\left(1 + \frac{1}{\left(1 + 10\frac{x_h^2}{x_l^2}\frac{\nu}{\beta}\right)^{\gamma_2-\gamma_1}}\right)\right)^{\frac{1}{2-\gamma_1-\gamma_2}}, \\
&\left(\frac{4\nu}{\epsilon}\cdot\frac{\left(x_h^2\alpha + x_h^4\beta\right)^2}{\frac{9}{10}x_l^6\beta}\left(1 + \frac{1}{\left(1 + 10\frac{x_h^2}{x_l^2}\frac{\nu}{\beta}\right)^{\gamma_2-\gamma_1}}\right)^2\right)^{\frac{1}{3-2\gamma_2}}\} \\
&n_3 = \max\{1 + 10\frac{x_h^2}{x_l^2}\frac{\nu_1}{\beta}, 1 + \nu_1\frac{x_h^2}{x_l^2}\} + 1 \\
&n_p = \max\{n_1, n_2, n_3, \left(\frac{2v_1}{\alpha}\left(\frac{32x_h^2\beta}{3}\right)^2(\epsilon' - \ln(0.5-\delta))\,e^{\frac{2}{x_h^2\beta}}\right)^{\frac{1}{2(\gamma_2-1)}}\} \\
&v_1 = \max\{6, 1 + 2e^{\frac{1}{x_h^2\beta}}\} \\
&c_p = n_p^{\gamma_2}.
\end{aligned}
$$

By looking at the *Bayesian linear regression problem* on domain $\mathcal{D}(n_p, \gamma_1, x_h, x_l, c_p)$, we next show that sampling from the posterior is (ϵ, δ) differentially private, although there is an SGLD step for which approximate sampling from the posterior using SGLD is not (ϵ', δ) differentially private.

Given dataset $D_1 \in \mathcal{D}(n_p, \gamma_1, x_h, x_l, c_p)$ (as defined in eq. 15, with $n = n_p$ and $c = c_p$), as $n_p \geq n_1$, the problem holds the constraints of Lemma $A.9$. Consequently, there exists an SGLD step that is not (ϵ'', δ) private for all $\epsilon'' \leq e^{-\frac{2}{x_h^2\beta}}\frac{\alpha}{2v_1}(\frac{3}{32x_h^2\beta})^2(\frac{c_p}{n_p})^2 + \ln(0.5-\delta)$. From eq. 16, the choice of n_p promises that $\epsilon' \leq e^{-\frac{2}{x_h^2\beta}}\frac{\alpha}{2v_1}(\frac{3}{32x_h^2\beta})^2(\frac{c_p}{n_p})^2 + \ln(0.5-\delta)$. Therefore, approximate sampling from the posterior using SGLD is not (ϵ', δ) differentially private.

Since $n_p \geq n_2$ and $n_p \geq n_3$, the problem holds the constraints of Claim D.28; therefore, one sample from the posterior is (ϵ, δ) differentially private.

$$
\begin{aligned}
&e^{-\frac{2}{x_h^2\beta}}\frac{\alpha}{2v_1}\left(\frac{3}{32x_h^2\beta}\right)^2\left(\frac{c_p}{n_p}\right)^2 + \ln(0.5-\delta) \geq \epsilon' \\
&\left(\frac{c_p}{n_p}\right)^2 \geq (\epsilon' - \ln(0.5-\delta))\,e^{\frac{2}{x_h^2\beta}}\left(\frac{32x_h^2\beta}{3}\right)^2\frac{2v_1}{\alpha} \\
&n_p^{2(\gamma_2-1)} \geq (\epsilon' - \ln(0.5-\delta))\,e^{\frac{2}{x_h^2\beta}}\left(\frac{32x_h^2\beta}{3}\right)^2\frac{2v_1}{\alpha} \\
&n_p \geq \left((\epsilon' - \ln(0.5-\delta))\,e^{\frac{2}{x_h^2\beta}}\left(\frac{32x_h^2\beta}{3}\right)^2\frac{2v_1}{\alpha}\right)^{\frac{1}{2(\gamma_2-1)}}
\end{aligned}
\tag{16}
$$

□

B. Posterior Sampling Privacy

This subsection provides proofs for the lemmas provided in subsection IV-A, along with a supporting lemma.

Proof of Lemma IV.2. Eq. 17 is a known result for the Bayesian inference problem for a linear 1D model with Gaussian noise with a known precision parameter (β) and a conjugate prior (see [31] - 3.49-4.51. for details). By choosing the basis function to be $\phi(x) = x$, working in one dimension, and choosing $\boldsymbol{m}_0 = 0, \boldsymbol{S}_0 = \alpha^{-1}$, we get the linear model defined in eq. 4 and the matching posterior described in Lemma IV.2.

$$p(\theta|D) = \mathcal{N}(\theta; m_N, S_N);\ m_N = S_N(S_0^{-1}m_0 + \beta\Phi^T t);\ S_N^{-1} = S_0^{-1} + \beta\Phi^T\Phi \tag{17}$$

□

Lemma A.1. *For a Bayesian linear regression problem on domain $\mathcal{D}(n, \gamma_1, x_h, x_l, c)$, such that $n > \max\{1+10\frac{x_h^2}{x_l^2}\frac{\nu}{\beta}, 1+\nu\frac{x_h^2}{x_l^2}\}$, one sample from the posterior is (ν, ϵ_1)-Rényi differentially private, and ϵ_1 is*

$$\begin{aligned}\epsilon_1 = &\frac{x_h^2}{2(n-1)x_l^2} + \frac{(\nu-1)\,\nu x_h^2}{2\left((n-1)\,x_l^2 - \nu x_h^2\right)} + \\ &\frac{20\nu\beta x_h^4}{9n^{1-2\gamma_1}x_l^2} + \frac{20\nu x_h^4(\alpha + x_h^2\beta)}{9x_l^4}\cdot\frac{(c+n^{\gamma_1})}{n^{2-\gamma_1}} \\ &+ \frac{5\nu x_h^4(\alpha + x_h^2\beta)^2}{9x_l^6\beta}\cdot\frac{(c+n^{\gamma_1})^2}{n^3}.\end{aligned}$$

Proof of Lemma A.1. By Definition III.3, for a single sample from the posterior to be (ν, ϵ') RDP, the Rényi divergence of order ν between any adjacent datasets needs to be bounded. Therefore, we consider two adjacent datasets, $D, \hat{D} \in \mathcal{D}(n, \gamma_1, x_h, x_l, c)$, and w.l.o.g, define that they differ in the last sample (where it is also allowed to be $(0, 0)$ for one of them, which saves us the need to consider also a neighbouring dataset with a size smaller by 1). To ease the already complex and detailed calculations, we use definitions in eq. 18.

$$\begin{aligned}&D = \{x_i, y_i\}_{i=1}^{n-1} \cup \{x_n, y_n\},\ \hat{D} = \{x_i, y_i\}_{i=1}^{n-1} \cup \{\hat{x}_n, \hat{y}_n\} \\ &z = \sum_{i=1}^{n-1} x_i^2,\ q = \sum_{i=1}^{n-1} y_i x_i\end{aligned} \tag{18}$$

According to Lemma IV.2 and with the definitions of eq. 18, the posterior distributions are

$$\begin{aligned}&p(\theta|D) = \mathcal{N}(\theta; \mu, \sigma^2);\ \mu = \frac{\beta(q + x_n y_n)}{\alpha + (z + x_n^2)\beta};\ \sigma^2 = \frac{1}{\alpha + (z + x_n)\beta} \\ &p(\theta|\hat{D}) = \mathcal{N}(\theta; \hat{\mu}, \hat{\sigma}^2);\ \hat{\mu} = \frac{\beta(q + \hat{x}_n \hat{y}_n)}{\alpha + (z + \hat{x}_n^2)\beta};\ \hat{\sigma}^2 = \frac{1}{\alpha + (z + \hat{x}_n)\beta}.\end{aligned} \tag{19}$$

Mark by $\mathrm{D}_\nu(f_1||f_2)$ the Réyni divergence of order ν between f_1 and f_2 - uni-variate normal distributions with means μ_1, μ_2 and variances σ_1, σ_2 accordingly. By [28], $\mathrm{D}_\nu(f_1||f_2)$ is

$$\begin{aligned}&\mathrm{D}_\nu\left(f_1||f_2\right) = \ln\frac{\sigma_1}{\sigma_2} + \frac{1}{2}\left(\nu-1\right)\ln\frac{\sigma_2^2}{\left(\sigma_{f_1,f_2}^2\right)_\nu^*} + \frac{1}{2}\frac{\nu\left(\mu_1-\mu_2\right)^2}{\left(\sigma^2\right)_\nu^*} \\ &\left(\sigma_{f_1,f_2}^2\right)_\nu^* = \nu\sigma_2^2 + \left(1-\nu\right)\sigma_1^2 > 0.\end{aligned}$$

Therefore, for $p(\theta|D)$ and $p(\theta|\hat{D})$, the Rényi divergence of order ν is given in eq. 20, where we omit the subscript for $(\sigma^2)_\nu^*$ since it is clear from context to which distributions it applies.

$$\begin{aligned}&\mathrm{D}_\nu\left(p\left(\theta|D\right)||p(\theta|\hat{D})\right) = \ln\frac{\sigma}{\hat{\sigma}} + \frac{1}{2}\left(\nu-1\right)\ln\frac{\hat{\sigma}^2}{\left(\sigma^2\right)_\nu^*} + \frac{1}{2}\frac{\nu\left(\mu-\hat{\mu}\right)^2}{\left(\sigma^2\right)_\nu^*} \\ &\left(\sigma^2\right)_\nu^* = \nu\hat{\sigma}^2 + \left(1-\nu\right)\sigma^2\end{aligned} \tag{20}$$

According to Claim D.23, $(\sigma^2)_\nu^* > 0$; therefore, the value $\mathrm{D}_\nu(p(\theta|D), p(\theta|\hat{D}))$ exists. In order to prove Rényi differential privacy, each of the terms of $\mathrm{D}_\nu(p(\theta|D), p(\theta|\hat{D}))$ is bounded separately, so their sum will be equal to ϵ_1. The bounds for each of the terms are proved at Claims D.24, D.25, and D.26. □

Proof of Lemma IV.3. The Lemma is a direct corollary of Lemma A.1 □

Proof of Lemma IV.4. By Lemma A.1, sampling from the posterior is (ν, ϵ_1)-RDP; therefore, by Lemma III.4, sampling from the posterior is also $(\epsilon_1 + \frac{\ln(\frac{1}{\delta})}{\nu-1}, \delta)$ differentially private. □

C. Stochastic Gradient Langevin Dynamics Privacy

This subsection provides proofs for the lemmas presented in subsection IV-B. The proofs in this section rely heavily on the analysis of the SGLD behaviour for the *Bayesian Linear Regression Problem* on datasets D_1, D_2. This analysis is provided in subsection A-D.

Proof of Lemma IV.5.

$$p(\hat{\theta}_j > \mu_j | D_2) = \sum_{r=1}^{n} p(\hat{\theta}_j^r > \mu_j | D_2) p(\hat{\theta}_j = \hat{\theta}_j^r | D_2) = \sum_{r=1}^{n} p(\hat{\theta}_j - \hat{\mu}_j^r > \mu_j - \hat{\mu}_j^r | D_2) p(\hat{\theta}_j = \hat{\theta}_j^r | D_2) =$$
$$\frac{1}{n} \sum_{r=1}^{n} p(\hat{\theta}_j - \hat{\mu}_j^r > \mu_j - \hat{\mu}_j^r | D_2) \leq \frac{1}{n} \sum_{r=1}^{n} \exp\left(-\frac{(\mu_j - \hat{\mu}_j^r)^2}{2(\hat{\sigma}_j^r)^2} \right)$$

where the inequality holds due to Chernoff bound (For further details about Chernoff bound, see [40]). □

Proof of Lemma IV.6. By Lemma A.7, for $n > \max\{\frac{\alpha}{x_h^2 \beta}, \frac{\alpha}{x_h^2 \beta}(e^{\frac{2}{x^2 \beta}} - 2) + \frac{1}{2x_h^2 \beta}, \frac{1}{2\alpha x_h^2 \beta} - \frac{1}{x_h^2 \beta}\}$ and $\dot{k} \in \mathbb{R}_{>0}$, eq. 21 holds. We can see that the lower bound described in eq. 21 is dominated by $\frac{c^2}{n^2}$, thus proving Lemma IV.6.

$$\begin{aligned} &\frac{(\mu_{(\lceil \dot{k} \rceil + 1)n} - \hat{\mu}^r_{(\lceil \dot{k} \rceil + 1)n})^2}{(\hat{\sigma}^r_{(\lceil \dot{k} \rceil + 1)n})^2} \geq e^{-\frac{2}{x_h^2 \beta}} \frac{\alpha}{v_1} (\frac{3}{32 x_h^2 \beta})^2 (\frac{c}{n})^2 \\ &v_1 = \max\{6, 1 + 2e^{\frac{1}{x_h^2 \beta}}\} \end{aligned} \tag{21}$$

□

Proof of Lemma IV.7. Define ϵ' as

$$\begin{aligned} &\epsilon' = e^{-\frac{2}{x_h^2 \beta}} \frac{\alpha}{2v_1} \left(\frac{3}{32 x_h^2 \beta} \right)^2 \left(\frac{c}{n} \right)^2 + \ln(0.5 - \delta) \\ &v_1 = \max\{6, 1 + 2e^{\frac{1}{x_h^2 \beta}}\}. \end{aligned}$$

By Lemma A.9, for $n > \max\{\frac{\alpha}{x_h^2 \beta}, \frac{\alpha}{x_h^2 \beta}(e^{\frac{2}{x_h^2 \beta}} - 2) + \frac{1}{2x_h^2 \beta}, \frac{1}{2\alpha x_h^2 \beta} - \frac{1}{x_h^2 \beta}\}$, there exists $T \in \mathbb{Z}_{>0}$ (marked in Lemma A.9 as $(\lceil \dot{k} \rceil + 1)n$) such that running SGLD for the *Bayesian linear regression problem* over D_1 for T steps will not be (ϵ, δ) differentially private for $\epsilon < \epsilon'$. As ϵ' is dominated by $\frac{c^2}{n^2}$, Lemma IV.7 is proved. □

D. Stochastic Gradient Langevin Dynamics Detailed Analysis

This subsection provides an analysis of SGLD behaviour for the *Bayesian Linear Regression Problem* on datasets D_1, D_2. We advise the reader to read the lemmas by order and provide here a summary of the analysis: Lemma A.2 provides an expression for the sample at the $(k+1)n$ SGLD step when using datasets D_1, D_2. Lemmas A.4 and A.5 use this expression to get a lower bound on the difference in means and an upper bounds on the variance, respectively. In turn, Lemma A.7 uses these lower and upper bounds to find a lower bound over $\left(\mu_{\lceil \dot{k} \rceil n+n} - \hat{\mu}^r_{\lceil \dot{k} \rceil n+n}\right)^2 / \left(\hat{\sigma}^r_{\lceil \dot{k} \rceil n+n}\right)^2$. This lower bound is used by Lemma A.8 to upper bound the probability mass of the SGLD process running on dataset D_2 in $S = \{s | s > \mu_j\}$. The difference in probability masses in S between the weights of an SGLD running on datasets $D1$ and D_2 leads to a breach of privacy, as shown in Lemma A.9.

In order to ease the analysis of the SGLD process for the *Bayesian linear regression problem* on domain $\mathcal{D}(n, \gamma_1, x_h, x_l, c)$, we use the markings in eq. 22.

$$\lambda = \left(1 - \frac{\eta}{2}\left(\alpha + n x_h^2 \beta\right)\right), \hat{\lambda} = \left(1 - \frac{\eta}{2}\left(\alpha + n\left(\frac{x_h}{2}\right)^2 \beta\right)\right), \rho = \frac{\eta}{2} n c x_h^2 \beta, \hat{\rho} = \frac{\eta}{2} n c \left(\frac{x_h}{2}\right)^2 \beta \tag{22}$$

Lemma A.2. *$\forall k \in \mathbb{Z}_{>0}$, $\hat{\theta}^r_{(k+1)n}$ has the following forms:*

$$\hat{\theta}^1_{(k+1)n} = \theta_0 \hat{\lambda}^{k+1} \lambda^{(n-1)(k+1)} + \sum_{j=0}^{k} \left(\hat{\lambda}\lambda^{n-1}\right)^j \left(\hat{\rho}\lambda^{n-1} + \rho \sum_{i=0}^{n-2} \lambda^i + \sqrt{\eta} \sum_{i=0}^{n-1} \lambda^i \xi_i\right)$$

$$\hat{\theta}^{r>1}_{(k+1)n} = \theta_0 \left(\hat{\lambda}\lambda^{n-1}\right)^{k+1} + \left(\sum_{i=1}^{r-1} \left(\rho + \sqrt{\eta}\xi\right) \hat{\lambda}\lambda^{n-i-1} + \left(\hat{\rho} + \sqrt{\eta}\xi\right) \lambda^{n-r} + \sum_{j=r+1}^{n} \left(\rho + \sqrt{\xi}\eta\right) \lambda^{n-j}\right) \sum_{l=0}^{k} \left(\hat{\lambda}\lambda^{n-1}\right)^l.$$

Proof of Lemma A.2. We can apply the SGLD update rule, defined in eq. 3, to the *Bayesian linear regression problem* over datasets D_1 and D_2 as follows: First, $p(\theta_j) = \mathbb{N}(\theta_j; 0, \alpha^{-1})$, and therefore

$$\nabla_{\theta_j} \ln p(\theta_j) = \nabla_{\theta_j} \ln\left(\frac{1}{\sqrt{2\pi\alpha^{-1}}}\right) - \nabla_{\theta_j} \frac{1}{2}\theta_j^2 \alpha = -\theta_j \alpha.$$

In a similar manner,

$$\nabla_{\theta_j} \ln p(y_i | \theta_j) = \nabla_{\theta_j} \ln\left(\frac{1}{\sqrt{2\pi\beta^{-1}}}\right) - \nabla_{\theta_j} \frac{1}{2}(y_i - \theta_j x_i)^2 \beta = (y_i - \theta_j x_i) x_i \beta.$$

Inserting these expressions to the SGLD update rule yields

$$\begin{aligned} &\theta_{j+1} = \theta_j + \frac{\eta}{2}\left(-\theta_j \alpha + n\left(y_i - \theta_j x_j\right) x_i \beta\right) + \sqrt{\eta}_j \xi_i = \theta_j \left(1 - \frac{\eta}{2}\left(\alpha + n x_j^2 \beta\right)\right) + \frac{\eta}{2} n y_i x_i \beta + \sqrt{\eta}\xi_j = \\ &\theta_j \left(1 - \frac{\eta}{2}\left(\alpha + n x_j^2 \beta\right)\right) + \frac{\eta}{2} n c x_i^2 \beta + \sqrt{\eta}\xi_j. \end{aligned} \tag{23}$$

By using standard tools for solving first-order non-homogeneous recurrence relations with variable coefficients, the value of $\hat{\theta}^1_n$ can be found:

$$\begin{aligned} &\hat{\theta}^1_n = \hat{\lambda}\lambda^{n-1}\left(\frac{\theta_0 \hat{\lambda} + \hat{\rho} + \sqrt{\eta}\xi}{\hat{\lambda}} + \sum_{i=2}^{n} \frac{\rho + \sqrt{\eta}\xi}{\hat{\lambda}\lambda^{i-1}}\right) = \theta_0 \hat{\lambda}\lambda^{n-1} + \left(\hat{\rho} + \sqrt{\eta}\xi\right)\lambda^{n-1} + \left(\rho + \sqrt{\eta}\xi\right) \sum_{i=2}^{n} \lambda^{n-1-(i-1)} = \\ &\theta_0 \hat{\lambda}\lambda^{n-1} + \left(\hat{\rho} + \sqrt{\eta}\xi\right)\lambda^{n-1} + \left(\rho + \sqrt{\eta}\xi\right) \sum_{i=2}^{n} \lambda^{n-1-(i-1)} = \theta_0 \hat{\lambda}\lambda^{n-1} + \left(\hat{\rho} + \sqrt{\eta}\xi\right)\lambda^{n-1} + \left(\rho + \sqrt{\eta}\xi\right) \sum_{i=0}^{n-2} \lambda^i = \\ &\theta_0 \hat{\lambda}\lambda^{n-1} + \hat{\rho}\lambda^{n-1} + \rho \sum_{i=0}^{n-2} \lambda^i + \sqrt{\eta}\xi \sum_{i=0}^{n-1} \lambda^i. \end{aligned}$$

Thus, we can define a new series, $\hat{\theta}^1_{(k+1)n} = c_1 \hat{\theta}^1_{kn} + c_2$ where $c_1 = \hat{\lambda}\lambda^{n-1}, c_2 = \hat{\rho}\lambda^{n-1} + \rho \sum_{i=0}^{n-2} \lambda^i + \sqrt{\eta}\xi \sum_{i=0}^{n-1} \lambda^i$. Using tools for solving first order non-homogeneous recurrence relations with constant coefficients, the value of $\hat{\theta}^1_{kn}$ can be found:

$$\begin{aligned} &\hat{\theta}^1_{kn} = c_1^k \left(\frac{\hat{\theta}^1_n}{c_1} + \sum_{i=2}^{k} \frac{c_2}{c_1^i}\right) = \theta^1_n c_1^{k-1} + \sum_{i=2}^{k} c_2 c_1^{k-i} = \theta^1_n c_1^{k-1} + c_2 \sum_{i=0}^{k-2} c_1^i = \left(\theta_0 c_1 + c_2\right) c_1^{k-1} + c_2 \sum_{i=0}^{k-2} c_1^i = \\ &\theta_0 \left(\hat{\lambda}\lambda^{n-1}\right)^k + \left(\hat{\rho}\lambda^{n-1} + \rho \sum_{i=0}^{n-2} \lambda^i + \sqrt{\eta}\xi \sum_{i=0}^{n-1} \lambda^i\right) \sum_{j=0}^{k-1} \left(\hat{\lambda}\lambda^{n-1}\right)^j. \end{aligned}$$

The proof for $\hat{\theta}^r_{kn}$ is done in a similar manner:

$$\hat{\theta}^{r>1}_n = \hat{\lambda}\lambda^{n-1}\left(\theta_0 + \sum_{i=1}^{r-1}\frac{\rho+\sqrt{\eta}\xi}{\lambda^i} + \frac{\hat{\rho}+\sqrt{\eta}\xi}{\lambda^{r-1}\hat{\lambda}} + \sum_{j=r+1}^{n}\frac{\rho+\sqrt{\eta}\xi}{\lambda^{j-1}\hat{\lambda}}\right) =$$
$$\hat{\lambda}\lambda^{n-1}\theta_0 + \sum_{i=1}^{r-1}\left(\rho+\sqrt{\eta}\xi\right)\hat{\lambda}\lambda^{n-i-1} + \lambda^{n-r}\left(\hat{\rho}+\sqrt{\eta}\xi\right) + \sum_{j=r+1}^{n}\left(\rho+\sqrt{\eta}\xi\right)\lambda^{n-j}$$

Thus, we can define a new series, $\hat{\theta}^{r>1}_{(k+1)n} = c_3\hat{\theta}^{r>1}_{kn} + c_4$, where $c_3 = \hat{\lambda}\lambda^{n-1}, c_4 = \sum_{i=1}^{r-1}\left(\rho+\sqrt{\eta}\xi\right)\hat{\lambda}\lambda^{n-i-1} + \lambda^{n-r}\left(\hat{\rho}+\sqrt{\eta}\xi\right) + \sum_{j=r+1}^{n}\left(\rho+\sqrt{\eta}\xi\right)\lambda^{n-j}$. Similarly to $\hat{\theta}^1_{kn}$ derivation, we can use first order non-homogeneous recurrence relations with constant coefficients to get:

$$\hat{\theta}^{r>1}_{kn} = \theta_0\left(\hat{\lambda}\lambda^{n-1}\right)^k + \left(\sum_{i=1}^{r-1}\left(\rho+\sqrt{\eta}\xi\right)\hat{\lambda}\lambda^{n-i-1} + \left(\hat{\rho}+\sqrt{\eta}\xi\right)\lambda^{n-r} + \sum_{j=r+1}^{n}\left(\rho+\sqrt{\xi}\eta\right)\lambda^{n-j}\right)\sum_{l=0}^{k-1}\left(\hat{\lambda}\lambda^{n-1}\right)^l.$$

□

Lemma A.3. *$\forall k \in \mathbb{Z}_{>0}$, $\theta_{(k+1)n}$ has the following form:*

$$\theta_{(k+1)n} = \theta_0\lambda^{(k+1)n} + \sum_{j=0}^{k}\lambda^{jn}\left(\rho\lambda^{n-1} + \rho\sum_{i=0}^{n-2}\lambda^i + \sqrt{\eta}\sum_{i=0}^{n-1}\lambda^i\xi_i\right)$$

Proof of Lemma A.3. First, notice that equation 23 applies for SGLD using dataset D_1. By using standard tools for solving first-order non-homogeneous recurrence relations, an expression for θ_n can be found:

$$\theta_n = \lambda^n\left(\frac{\theta_0\lambda+\rho+\sqrt{\eta}\xi}{\lambda} + \sum_{i=2}^{n}\frac{\rho+\sqrt{\eta}\xi}{\lambda^i}\right) = \theta_0\lambda^n + \rho\lambda^{n-1} + \rho\sum_{i=0}^{n-2}\lambda^i + \sqrt{\rho}\xi\sum_{i=0}^{n-1}\lambda^i$$

By defining a new series $\theta_{(k+1)n} = c_5\theta_{kn} + c_6$ and using tools for solving first-order non-homogeneous recurrence relations with constant coefficients, the value of θ_{kn} can be found:

$$\theta_{kn} = c_5^k\left(\frac{\theta_n}{c_5} + \sum_{i=2}^{k}\frac{c_6}{c_5}\right) = \theta_0\lambda^{nk} + \left(\rho\lambda^{n-1} + \rho\sum_{i=0}^{n-2}\lambda^i + \sqrt{\eta}\xi\sum_{i=0}^{n-1}\lambda^i\right)\sum_{j=0}^{k-1}\lambda^{jn}$$

□

Lemma A.4. *$\forall k \in \mathbb{Z}_{>0}$, the value of $\mu_{kn+n} - \hat{\mu}^r_{kn+n}$ can be lower bounded:*

$$\mu_{kn+n} - \hat{\mu}^r_{kn+n} \geq \lambda^{n-1}\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\lambda^{k(n-1)}\left(\hat{\lambda}^{k+1} - \lambda^{k+1}\right).$$

Proof of Lemma A.4. The proof of this lemma is separated into two cases, for $r = 1$ and for $r > 1$. For $r = 1$, using $\mathbb{E}[\theta_0] = 0$ and $\mathbb{E}[\xi] = 0$, it is easy to derive eq. 24 from Lemmas A.2 and A.3.

$$\begin{aligned}\hat{\mu}^1_{(k+1)n} &= \rho\sum_{i=0}^{n-2}\lambda^i\sum_{j=0}^{k}\left(\hat{\lambda}\lambda^{(n-1)}\right)^j + \hat{\rho}\lambda^{n-1}\sum_{j=0}^{k}\left(\hat{\lambda}\lambda^{(n-1)}\right)^j \\ \mu_{kn+n} &= \rho\sum_{i=0}^{n-2}\lambda^i\sum_{j=0}^{k}\lambda^{jn} + \rho\lambda^{n-1}\sum_{r=0}^{k}\lambda^{rn}\end{aligned} \tag{24}$$

We use the sum of a geometric sequence to get

$$\hat{\mu}^1_{(k+1)n} = \rho\sum_{i=0}^{n-2}\lambda^i\sum_{j=0}^{k}\left(\hat{\lambda}\lambda^{(n-1)}\right)^j + \hat{\rho}\lambda^{n-1}\sum_{j=0}^{k}\left(\hat{\lambda}\lambda^{(n-1)}\right)^j = \left(\rho\left(\frac{1-\lambda^{n-1}}{1-\lambda}\right) + \hat{\rho}\lambda^{n-1}\right)\frac{1-\left(\hat{\lambda}\lambda^{n-1}\right)^{k+1}}{1-\hat{\lambda}\lambda^{n-1}}$$

and

$$\mu_{(k+1)n} = \rho\sum_{i=0}^{n-2}\lambda^i\sum_{j=0}^{k}\lambda^{nj} + \rho\lambda^{n-1}\sum_{j=0}^{k}\lambda^{nj} = \left(\rho\left(\frac{1-\lambda^{n-1}}{1-\lambda}\right) + \rho\lambda^{n-1}\right)\frac{1-\lambda^{(k+1)n}}{1-\lambda^n}.$$

Therefore the difference between the means can be lower bounded:

$$
\begin{aligned}
&\mu_{kn+n} - \hat{\mu}^1_{kn+n} = \\
&\frac{1-\lambda^{(k+1)n}}{1-\lambda^n}\left(\rho\left(\frac{1-\lambda^{n-1}}{1-\lambda}\right)+\rho\lambda^{n-1}\right) - \frac{1-\left(\hat{\lambda}\lambda^{(n-1)}\right)^{k+1}}{1-\hat{\lambda}\lambda^{n-1}}\left(\rho\left(\frac{1-\lambda^{n-1}}{1-\lambda}\right)+\hat{\rho}\lambda^{n-1}\right) =^* \\
&\frac{1-\lambda^{(k+1)n}}{1-\lambda^n}\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(1-\lambda^n\right) - \frac{1-\left(\hat{\lambda}\lambda^{(n-1)}\right)^{k+1}}{1-\hat{\lambda}\lambda^{n-1}}\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(1-\lambda^{n-1}\left(\frac{1}{4}\lambda+\frac{3}{4}\right)\right) = \\
&\left(1-\lambda^{(k+1)n}\right)\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta} - \frac{1-\left(\hat{\lambda}\lambda^{(n-1)}\right)^{k+1}}{1-\hat{\lambda}\lambda^{n-1}}\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(1-\lambda^{n-1}\left(\frac{1}{4}\lambda+\frac{3}{4}\right)\right) = \\
&\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(\left(1-\lambda^{(k+1)n}\right) - \frac{1-\left(\hat{\lambda}\lambda^{(n-1)}\right)^{k+1}}{1-\hat{\lambda}\lambda^{n-1}}\left(1-\lambda^{n-1}\left(\frac{1}{4}\lambda+\frac{3}{4}\right)\right)\right) =^{**} \\
&\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(\frac{\lambda^{n-1}\frac{3}{4}\frac{\eta}{2}\alpha\left(1-\hat{\lambda}^{k+1}\lambda^{(k+1)(n-1)}\right)+\lambda^{(k+1)(n-1)}\left(\hat{\lambda}^{k+1}-\lambda^{k+1}\right)\left(1-\lambda^{n-1}\hat{\lambda}\right)}{1-\lambda^{n-1}\hat{\lambda}}\right) \geq \\
&\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(\frac{\lambda^{(k+1)(n-1)}\left(\hat{\lambda}^{k+1}-\lambda^{k+1}\right)\left(1-\lambda^{n-1}\hat{\lambda}\right)}{1-\lambda^{n-1}\hat{\lambda}}\right) = \\
&\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\lambda^{(k+1)(n-1)}\left(\hat{\lambda}^{k+1}-\lambda^{k+1}\right) = \\
&\lambda^{n-1}\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\lambda^{k(n-1)}\left(\hat{\lambda}^{k+1}-\lambda^{k+1}\right)
\end{aligned}
$$

where equality * holds from Claims D.1, D.2, and D.3, equality ** holds from Claim D.5, and the inequality holds because $\lambda < \hat{\lambda} < 1$. This proves Lemma A.4 for $r = 1$.

For $r > 1$, from Lemma A.2, it is easy to see that

$$
\begin{aligned}
&\hat{\theta}^{r>1}_{(k+1)n} = \left(\left(\theta_0\lambda^{r-1}+\rho\sum_{i=0}^{r-2}\lambda^i+\sqrt{\eta}\sum_{i=0}^{r-2}\lambda^i\xi_i\right)\hat{\lambda}^k\lambda^{k(n-1)}+\sum_{j=0}^{k-1}(\hat{\lambda}\lambda^{n-1})^j\left(\hat{\rho}\lambda^{n-1}+\rho\sum_{i=0}^{n-2}\lambda^i+\sqrt{\eta}\sum_{i=0}^{n-1}\lambda^i\xi_i\right)\right)\hat{\lambda}\lambda^{n-r}+ \\
&\hat{\rho}\lambda^{n-r}+\rho\sum_{j=0}^{n-r-1}\lambda^j+\sqrt{\eta}\sum_{j=0}^{n-r}\xi\lambda^j.
\end{aligned}
$$

Therefore,

$$
\hat{\mu}^{r>1}_{(k+1)n} = \left(\left(\rho\sum_{i=0}^{r-2}\lambda^i\right)\hat{\lambda}^k\lambda^{k(n-1)}+\sum_{j=0}^{k-1}(\hat{\lambda}\lambda^{n-1})^j\left(\hat{\rho}\lambda^{n-1}+\rho\sum_{i=0}^{n-2}\lambda^i\right)\right)\hat{\lambda}\lambda^{n-r}+\hat{\rho}\lambda^{n-r}+\rho\sum_{j=0}^{n-r-1}\lambda^j.
$$

Consequently, the difference between the means, for $r > 1$, can be lower bounded:

$$
\begin{aligned}
&\mu_{kn+n} - \hat{\mu}^r_{kn+n} = \\
&\lambda^{n-r}\big(\lambda\rho\lambda^k\lambda^{k(n-1)}\sum_{i=0}^{r-2}\lambda^i + \lambda\sum_{j=0}^{k-1}(\lambda\lambda^{n-1})^j(\rho\lambda^{n-1} + \rho\sum_{i=0}^{n-2}\lambda^i) - \\
&\hat{\lambda}\rho\hat{\lambda}^k\lambda^{k(n-1)}\sum_{i=0}^{r-2}\lambda^i - \hat{\lambda}\sum_{j=0}^{k-1}(\hat{\lambda}\lambda^{n-1})^j(\hat{\rho}\lambda^{n-1} + \rho\sum_{i=0}^{n-2}\lambda^i)\big) + \lambda^{n-r}(\rho - \hat{\rho}) = \\
&\lambda^{n-r}\lambda^{k(n-1)}\rho(\lambda^{k+1} - \hat{\lambda}^{k+1})\sum_{i=0}^{r-2}\lambda^i + \lambda^{n-r}\sum_{j=0}^{k-1}\lambda^{(n-1)j}(\lambda^{n-1}(\rho\lambda^{j+1} - \\
&\hat{\rho}\hat{\lambda}^{j+1}) + (\lambda^{j+1} - \hat{\lambda}^{j+1})\rho\sum_{i=0}^{n-2}\lambda^i) + \lambda^{n-r}(\rho - \hat{\rho}) =^* \\
&\lambda^{n-r}\lambda^{k(n-1)}\rho(\lambda^{k+1} - \hat{\lambda}^{k+1})\frac{1-\lambda^{r-1}}{1-\lambda} + \lambda^{n-r}\frac{ncx_h^2\beta}{\alpha + nx_h^2\beta}(\lambda(1-\lambda^{kn}) - \\
&\hat{\lambda}\frac{1-(\lambda^{n-1}\hat{\lambda})^k}{1-\lambda^{n-1}\hat{\lambda}}(1-\lambda^n(\frac{3}{4}\lambda^{-1} + \frac{1}{4}))) + \lambda^{n-r}(\rho - \hat{\rho}) =^{**} \\
&\lambda^{n-r}\lambda^{k(n-1)}\rho(\lambda^{k+1} - \hat{\lambda}^{k+1})\frac{1-\lambda^{r-1}}{1-\lambda} + \lambda^{n-r}\frac{ncx_h^2\beta}{\alpha + nx_h^2\beta}((\lambda - \hat{\lambda}) + \\
&\frac{\lambda^{n-1}(\frac{3}{4}\frac{\eta}{2}\alpha(1-\hat{\lambda}^k\lambda^{k(n-1)}))}{1-\hat{\lambda}\lambda^{n-1}} + \lambda^{k(n-1)}(\hat{\lambda}^{k+1} - \lambda^{k+1})) + \lambda^{n-r}(\rho - \hat{\rho}) =^{***} \\
&\lambda^{n-r}\frac{ncx_h^2\beta}{\alpha + nx_h^2\beta}\lambda^{k(n-1)}(\lambda^{k+1} - \hat{\lambda}^{k+1})(1-\lambda^{r-1}) + \lambda^{n-r}\frac{ncx_h^2\beta}{\alpha + nx_h^2\beta}((\lambda - \hat{\lambda}) \\
&+ \frac{\lambda^{n-1}(\frac{3}{4}\frac{\eta}{2}\alpha(1-\hat{\lambda}^k\lambda^{k(n-1)}))}{1-\hat{\lambda}\lambda^{n-1}} + \lambda^{k(n-1)}(\hat{\lambda}^{k+1} - \lambda^{k+1})) + \lambda^{n-r}(\rho - \hat{\rho}) = \\
&\lambda^{n-r}\frac{ncx_h^2\beta}{\alpha + nx_h^2\beta}\lambda^{k(n-1)}(\lambda^{k+1} - \hat{\lambda}^{k+1})(1-\lambda^{r-1} - 1) + \\
&\lambda^{n-r}\frac{ncx_h^2\beta}{\alpha + nx_h^2\beta}(\lambda - \hat{\lambda} + \frac{\lambda^{n-1}(\frac{3}{4}\frac{\eta}{2}\alpha(1-\hat{\lambda}^k\lambda^{k(n-1)}))}{1-\hat{\lambda}\lambda^{n-1}}) + \lambda^{n-r}(\rho - \hat{\rho}) = \\
&\lambda^{n-r}\frac{ncx_h^2\beta}{\alpha + nx_h^2\beta}\lambda^{k(n-1)}(\hat{\lambda}^{k+1} - \lambda^{k+1})\lambda^{r-1} + \\
&\lambda^{n-r}\frac{ncx_h^2\beta}{\alpha + nx_h^2\beta}(\lambda - \hat{\lambda} + \frac{\lambda^{n-1}(\frac{3}{4}\frac{\eta}{2}\alpha(1-\hat{\lambda}^k\lambda^{k(n-1)}))}{1-\hat{\lambda}\lambda^{n-1}}) + \lambda^{n-r}(\rho - \hat{\rho}) = \\
&\lambda^{n-1}\frac{ncx_h^2\beta}{\alpha + nx_h^2\beta}\lambda^{k(n-1)}(\hat{\lambda}^{k+1} - \lambda^{k+1}) + \\
&\lambda^{n-r}\frac{ncx_h^2\beta}{\alpha + nx_h^2\beta}(\lambda - \hat{\lambda} + \frac{\lambda^{n-1}(\frac{3}{4}\frac{\eta}{2}\alpha(1-\hat{\lambda}^k\lambda^{k(n-1)}))}{1-\hat{\lambda}\lambda^{n-1}}) + \lambda^{n-r}(\rho - \hat{\rho}) >^{****} \\
&\lambda^{n-1}\frac{ncx_h^2\beta}{\alpha + nx_h^2\beta}\lambda^{k(n-1)}(\hat{\lambda}^{k+1} - \lambda^{k+1})
\end{aligned}
$$

where equality * holds from Claims D.6 and D.7, equality ** holds from Claim D.10, equality *** holds from Claim D.1, and equality **** holds from Claim D.11 and because $\hat{\lambda} > \lambda$. □

Lemma A.5. *For all $k \in \mathbb{Z}_{>0}$ such that $0 < k \leq \frac{1}{2n}\log_\lambda(\frac{1}{1+\frac{1}{\alpha\eta}(1-\lambda^2)}) - 1$, and $x_h^2\beta > 3, n > \max\{\frac{1}{2\alpha x_h^2\beta} - \frac{1}{x_h^2\beta}, \frac{\alpha}{x_h^2\beta}(e^{\frac{2}{x_h^2\beta}} - 2) + \frac{1}{2x_h^2\beta}\}$, the values of $(\hat{\sigma}^r_{(k+1)n})^2$ can be upper bounded as following:*

$$
\begin{aligned}
(\hat{\sigma}^1_{(k+1)n})^2 &\leq 2(\hat{\lambda}\lambda^{n-1})^2\frac{1}{\alpha}(\hat{\lambda}\lambda^{n-1})^{2k} \\
(\hat{\sigma}^{r>1}_{(k+1)n})^2 &\leq 6(\hat{\lambda}\lambda^{n-r})^2\frac{1}{\alpha}(\hat{\lambda}\lambda^{n-1})^{2k}.
\end{aligned} \tag{25}
$$

Proof of Lemma A.5. The proof will be separated into two cases: $r = 1$ and $r > 1$. Starting from the case of $r = 1$, since the noise and the prior have Normal distributions, $(\hat{\sigma}^1_{kn+n})^2$ could be easily computed from Lemma A.2. Eq. 26 yields a first general upper bound on $(\hat{\sigma}^1_{kn+n})^2$ applicable for all $k \in \mathbb{Z}_{>0}$.

$$
\begin{aligned}
&(\hat{\sigma}^1_{kn+n})^2 = \frac{1}{\alpha}\left(\hat{\lambda}\lambda^{(n-1)}\right)^{2(k+1)} + \eta\sum_{i=0}^{n-1}\lambda^{2i}\sum_{j=0}^{k}\left(\hat{\lambda}^2\lambda^{2(n-1)}\right)^j = \left(\hat{\lambda}\lambda^{(n-1)}\right)^{2(k+1)}\left(\frac{1}{\alpha} + \eta\sum_{i=0}^{n-1}\lambda^{2i}\sum_{j=0}^{k}\frac{\left(\hat{\lambda}\lambda^{(n-1)}\right)^{2j}}{\left(\hat{\lambda}\lambda^{(n-1)}\right)^{2(k+1)}}\right) = \\
&\left(\hat{\lambda}\lambda^{(n-1)}\right)^{2(k+1)}\left(\frac{1}{\alpha} + \eta\sum_{i=0}^{n-1}\lambda^{2i}\sum_{j=0}^{k}\left(\hat{\lambda}\lambda^{(n-1)}\right)^{2(j-(k+1))}\right) \leq \left(\hat{\lambda}\lambda^{(n-1)}\right)^{2(k+1)}\left(\frac{1}{\alpha} + \eta\sum_{i=0}^{n-1}\lambda^{2i}\sum_{j=0}^{k}\lambda^{2n(j-(k+1))}\right) = \\
&\left(\hat{\lambda}\lambda^{(n-1)}\right)^{2(k+1)}\left(\frac{1}{\alpha} + \eta\sum_{i=1}^{(k+1)n}\lambda^{-2i}\right) = \left(\hat{\lambda}\lambda^{(n-1)}\right)^{2(k+1)}\left(\frac{1}{\alpha} + \eta\lambda^{-2}\frac{1-\lambda^{-2(k+1)n}}{1-\lambda^{-2}}\right)
\end{aligned}
\tag{26}
$$

where the inequality holds because $\lambda < \hat{\lambda}$.

By Claim D.12, this upper bound can be further refined for $0 < k \leq \frac{1}{2n}\log_\lambda(\frac{1}{1+\frac{1}{\alpha\eta}(1-\lambda^2)}) - 1$:

$$
\left(\hat{\lambda}\lambda^{(n-1)}\right)^{2(k+1)}\left(\frac{1}{\alpha} + \eta\lambda^{-2}\frac{1-\lambda^{-2(k+1)n}}{1-\lambda^{-2}}\right) \leq 2\left(\hat{\lambda}\lambda^{(n-1)}\right)^{2(k+1)}\frac{1}{\alpha}. \tag{27}
$$

This proves the lemma for $r = 1$.

For $r > 1$, $(\hat{\sigma}^{r>1}_{kn+n})^2$ can be bounded as follows:

$$
\begin{aligned}
&(\hat{\sigma}^{r>1}_{kn+n})^2 = \\
&\left(\hat{\lambda}\lambda^{n-r}\right)^2\eta\left(\left(\hat{\lambda}^k\lambda^{k(n-1)}\right)^2\sum_{i=0}^{r-2}\lambda^{2i} + \sum_{j=0}^{k-1}\left(\hat{\lambda}\lambda^{n-1}\right)^{2j}\sum_{i=0}^{n-1}\lambda^{2i}\right) + \eta\sum_{i=0}^{n-r}\lambda^{2i} + \frac{1}{\alpha}\left(\hat{\lambda}\lambda^{n-1}\right)^{2k}\left(\hat{\lambda}\lambda^{n-1}\right)^2 \leq^* \\
&\left(\hat{\lambda}\lambda^{n-r}\right)^2\eta\left(\left(\hat{\lambda}^k\lambda^{k(n-1)}\right)^2\sum_{i=0}^{r-2}\lambda^{2i} + \sum_{j=0}^{k-1}\left(\hat{\lambda}\lambda^{n-1}\right)^{2j}\sum_{i=0}^{n-1}\lambda^{2i}\right) + \eta\sum_{i=0}^{n-r}\lambda^{2i} + \frac{1}{\alpha}\left(\hat{\lambda}\lambda^{n-1}\right)^{2k}\left(\hat{\lambda}\lambda^{n-r}\right)^2 = \\
&\left(\hat{\lambda}\lambda^{n-r}\right)^2\left(\frac{1}{\alpha}\left(\hat{\lambda}\lambda^{n-1}\right)^{2k} + \eta\left(\hat{\lambda}\lambda^{n-1}\right)^{2k}\sum_{i=0}^{r-2}\lambda^{2i} + \eta\sum_{j=0}^{k-1}\left(\hat{\lambda}\lambda^{n-1}\right)^{2j}\sum_{i=0}^{n-1}\lambda^{2i}\right) + \eta\sum_{i=0}^{n-r}\lambda^{2i} \leq^{**} \\
&\left(\hat{\lambda}\lambda^{n-r}\right)^2\left(\frac{1}{\alpha}\left(\hat{\lambda}\lambda^{n-1}\right)^{2k} + \eta\left(\hat{\lambda}\lambda^{n-1}\right)^{2k}\sum_{i=0}^{n-1}\lambda^{2i} + \eta\sum_{j=0}^{k-1}\left(\hat{\lambda}\lambda^{n-1}\right)^{2j}\sum_{i=0}^{n-1}\lambda^{2i}\right) + \eta\sum_{i=0}^{n-r}\lambda^{2i} \leq^{***} \\
&2\left(\hat{\lambda}\lambda^{n-r}\right)^2\left(\frac{1}{\alpha}\left(\hat{\lambda}\lambda^{n-1}\right)^{2\dot{k}} + \eta\left(\hat{\lambda}\lambda^{n-1}\right)^{2k}\sum_{i=0}^{n-1}\lambda^{2i} + \eta\sum_{j=0}^{k-1}\left(\hat{\lambda}\lambda^{n-1}\right)^{2j}\sum_{i=0}^{n-1}\lambda^{2i}\right)
\end{aligned}
\tag{28}
$$

where inequality * is true because $\lambda < 1$ and $r > 1$, inequality ** is true because $r \leq n$, and inequality *** is true because of Claim D.15.

For $0 < k \leq \frac{1}{2n}\log_\lambda(\frac{1}{1+\frac{1}{\alpha\eta}(1-\lambda^2)}) - 1$, this bound can be further refined:

$$
\begin{aligned}
&2(\hat{\lambda}\lambda^{n-r})^2\left(\frac{1}{\alpha}\left(\hat{\lambda}\lambda^{n-1}\right)^{2k} + \eta\left(\hat{\lambda}\lambda^{n-1}\right)^{2k}\sum_{i=0}^{n-1}\lambda^{2i} + \eta\sum_{j=0}^{k-1}\left(\hat{\lambda}\lambda^{n-1}\right)^{2j}\sum_{i=0}^{n-1}\lambda^{2i}\right) \leq \\
&6(\hat{\lambda}\lambda^{n-r})^2\frac{1}{\alpha}(\hat{\lambda}\lambda^{n-1})^{2k}
\end{aligned}
\tag{29}
$$

where the inequality is true because of Claims D.12 and D.14, which provide the bound for $r > 1$. □

Lemma A.6. *Mark* $\dot{k} = \frac{1}{2n}\log_\lambda(\frac{1}{1+\frac{1}{\alpha\eta}(1-\lambda^2)}) - 1$. *For the conditions of Lemma A.5,* $\dot{k} > 0$ *and the values of* $\hat{\sigma}^r_{\lceil\dot{k}\rceil n+n}$ *can be upper bounded as following:*

$$(\hat{\sigma}^1_{\lceil\dot{k}\rceil n+n})^2 \leq \left(1 + 2e^{\frac{1}{x_h^2\beta}}\right)\left(\hat{\lambda}\lambda^{n-1}\right)^2 \frac{1}{\alpha}\left(\hat{\lambda}\lambda^{(n-1)}\right)^{2\lceil\dot{k}\rceil}$$
$$(\hat{\sigma}^{r>1}_{\lceil\dot{k}\rceil n+n})^2 \leq 6\left(\hat{\lambda}\lambda^{n-r}\right)^2 \frac{1}{\alpha}\left(\hat{\lambda}\lambda^{n-1}\right)^{2\lceil\dot{k}\rceil}.$$

Proof of Lemma A.6. We prove this lemma by augmenting the proof of Lemma A.5. We begin with $r > 1$. Lemma A.5 first proved, in eq. 28, a bound over $(\hat{\sigma}^{r>1}_{(k+1)n})^2$ applicable for $k > 0$. As this bound is applicable for all $k > 0$, it also applies for $\lceil\dot{k}\rceil$. Then, in eq. 29, Lemma A.5 refined the bound for $0 < k \leq \dot{k}$ using Claims D.14, D.12 and, D.15. Therefore, if these claims also hold for $k = \lceil\dot{k}\rceil$, then the result of Lemma A.5 also applies to $k = \lceil\dot{k}\rceil$. Claims D.14 and D.12 apply to all $k \leq \frac{1}{2n}\log_\lambda(\frac{1}{1+\frac{1}{\alpha\eta}(1-\lambda^2)})$. Since $\lceil\dot{k}\rceil \leq \dot{k} + 1 = \frac{1}{2n}\log_\lambda(\frac{1}{1+\frac{1}{\alpha\eta}(1-\lambda^2)})$, the claims also apply to $\lceil\dot{k}\rceil$. Claim D.15 was proved for all k, thus also applies to $\lceil\dot{k}\rceil$.

For $r = 1$, the bound found at eq. 26 is applicable for all k, hence

$$(\hat{\sigma}^1_{(\lceil\dot{k}\rceil+1)n})^2 \leq \left(\hat{\lambda}\lambda^{n-1}\right)^{2(\lceil\dot{k}\rceil+1)}\left(\frac{1}{\alpha} + \eta\lambda^{-2}\frac{1-\lambda^{-2(\lceil\dot{k}\rceil+1)n}}{1-\lambda^{-2}}\right) \leq \left(\hat{\lambda}\lambda^{n-1}\right)^{2(\lceil\dot{k}\rceil+1)}\frac{1}{\alpha}\left(1 + 2e^{\frac{1}{x_h^2\beta}}\right)$$

where the last inequality is true because of Claim D.17.

All that is left is to prove that $\dot{k} = \frac{1}{2n}\log_\lambda(\frac{1}{1+\frac{1}{\alpha\eta}(1-\lambda^2)}) - 1 > 0$, which is done in Claim D.21. □

Lemma A.7. *For* $\dot{k}$ *defined in Lemma A.6, the conditions of Lemma A.5, and* $n > \frac{\alpha}{x_h^2\beta}$*:*

$$\frac{\left(\mu_{\lceil\dot{k}\rceil n+n} - \hat{\mu}^r_{\lceil\dot{k}\rceil n+n}\right)^2}{\left(\hat{\sigma}^r_{\lceil\dot{k}\rceil n+n}\right)^2} \geq e^{-\frac{2}{x_h^2\beta}}\frac{\alpha}{v_1}\left(\frac{3}{32x_h^2\beta}\right)^2\left(\frac{c}{n}\right)^2$$
$$v_1 = \max\{6, 1 + 2e^{\frac{1}{x_h^2\beta}}\}.$$

Proof of Lemma A.7.

$$\frac{\left(\mu_{\lceil\dot{k}\rceil n+n} - \hat{\mu}^r_{\lceil\dot{k}\rceil n+n}\right)^2}{\hat{\sigma}^r_{\lceil\dot{k}\rceil n+n}} \geq \frac{\left(\lambda^{n-1}\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\lambda^{\lceil\dot{k}\rceil(n-1)}\left(\hat{\lambda}^{\lceil\dot{k}\rceil+1} - \lambda^{\lceil\dot{k}\rceil+1}\right)\right)^2}{v_1\left(\hat{\lambda}\lambda^{n-r}\right)^2\frac{1}{\alpha}\left(\hat{\lambda}\lambda^{n-1}\right)^{2\lceil\dot{k}\rceil}} =$$
$$\frac{\lambda^{2\lceil\dot{k}\rceil(n-1)}\left(\lambda^{n-1}\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(\hat{\lambda}^{\lceil\dot{k}\rceil+1} - \lambda^{\lceil\dot{k}\rceil+1}\right)\right)^2}{v_1\left(\hat{\lambda}\lambda^{n-r}\right)^2\frac{1}{\alpha}\left(\hat{\lambda}\lambda^{n-1}\right)^{2\lceil\dot{k}\rceil}} = \frac{\left(\lambda^{n-1}\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(\hat{\lambda}^{\lceil\dot{k}\rceil+1} - \lambda^{\lceil\dot{k}\rceil+1}\right)\right)^2}{v_1\left(\hat{\lambda}\lambda^{n-r}\right)^2\frac{1}{\alpha}\hat{\lambda}^{2\lceil\dot{k}\rceil}} =$$
$$\frac{\alpha\lambda^{2(r-1)}}{v_1}\left(\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\right)^2\frac{\left(\hat{\lambda}^{\lceil\dot{k}\rceil+1} - \lambda^{\lceil\dot{k}\rceil+1}\right)^2}{\hat{\lambda}^{2\lceil\dot{k}\rceil+1}} = \frac{\alpha\lambda^{2(r-1)}}{v_1}\left(\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\right)^2\left(1 - \frac{\lambda^{\lceil\dot{k}\rceil+1}}{\hat{\lambda}^{\lceil\dot{k}\rceil+1}}\right)^2 \geq$$
$$\frac{\alpha\lambda^{2(r-1)}}{v_1}\left(\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\right)^2\left(1 - \left(1 - \frac{\frac{3}{4}nx^2\beta}{(\alpha+nx_h^2\beta)^2 - (\alpha+\frac{1}{4}nx^2\beta)}\right)\right)^2 =$$
$$\frac{\alpha\lambda^{2(r-1)}}{v_1}\left(\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\right)^2\left(\frac{\frac{3}{4}nx_h^2\beta}{(\alpha+nx_h^2\beta)^2 - (\alpha+\frac{1}{4}nx^2\beta)}\right)^2 \geq$$
$$\frac{\alpha\lambda^{2(r-1)}}{v_1}\left(\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\right)^2\left(\frac{\frac{3}{4}nx^2\beta}{(\alpha+nx_h^2\beta)^2}\right)^2 \geq \frac{\alpha\lambda^{2(r-1)}}{v_1}\left(\frac{ncx^2\beta}{2nx_h^2\beta}\right)^2\left(\frac{\frac{3}{4}nx^2\beta}{(2nx^2\beta)^2}\right)^2 =$$
$$\frac{\alpha\lambda^{2(r-1)}}{v_1}\left(\frac{c}{2}\right)^2\left(\frac{\frac{3}{4}}{4nx_h^2\beta}\right)^2 = \frac{\alpha\lambda^{2(r-1)}}{v_1}\left(\frac{3}{32x_h^2\beta}\right)^2\left(\frac{c}{n}\right)^2 \geq$$
$$\frac{\alpha\lambda^{2(n-1)}}{v_1}\left(\frac{3}{32x_h^2\beta}\right)^2\left(\frac{c}{n}\right)^2 \geq e^{-\frac{2}{x_h^2\beta}}\frac{\alpha}{v_1}\left(\frac{3}{32x_h^2\beta}\right)^2\left(\frac{c}{n}\right)^2$$

where first inequality holds from Lemmas A.4 and A.6, and the definition of v_1, the second inequality is true because of Claim D.17 and Claim D.21, fourth inequality holds under the assumption of $nx_h^2\beta > \alpha$, and the last inequality holds from Claim D.18. □

Claim A.8. *For $\dot{k}$, defined in Lemma A.6, and the conditions of Lemma A.7:*

$$p\left(\hat{\theta}_{(\lceil \dot{k} \rceil+1)n} > \mu_{(\lceil \dot{k} \rceil+1)n} | D_2\right) \leq e^{-e^{-\frac{2}{x_h^2\beta}} \frac{\alpha}{2v_1}\left(\frac{3}{32x_h^2\beta}\right)^2 \left(\frac{c}{n}\right)^2}.$$

Proof of Claim A.8.

$$p\left(\hat{\theta}_{(\lceil \dot{k} \rceil+1)n} > \mu_{(\lceil \dot{k} \rceil+1)n} | \hat{D}\right) \leq \frac{1}{n}\sum_{r=1}^{n} \exp\left(-\frac{\left(\mu_{(\lceil \dot{k} \rceil+1)n} - \hat{\mu}^r_{(\lceil \dot{k} \rceil+1)n}\right)^2}{2\left(\sigma^r_{(\lceil \dot{k} \rceil+1)n}\right)^2}\right) \leq$$

$$\frac{1}{n}\sum_{r=1}^{n} \exp\left(-e^{-\frac{2}{x_h^2\beta}} \frac{\alpha}{2v_1}\left(\frac{3}{32x_h^2\beta}\right)^2 \left(\frac{c}{n}\right)^2\right) = \exp\left(-e^{-\frac{2}{x_h^2\beta}} \frac{\alpha}{2v_1}\left(\frac{3}{32x_h^2\beta}\right)^2 \left(\frac{c}{n}\right)^2\right)$$

where the first inequality holds due to Lemma IV.5 and second inequality holds due to Lemma A.7. □

Lemma A.9. *For the Bayesian linear regression problem over $\mathcal{D}(n, \gamma_1, x_h, x_l, c)$, the conditions of Lemma A.7, and $\dot{k}$ defined in Lemma A.6, approximate sampling by running SGLD for $(\lceil \dot{k} \rceil + 1)n$ steps will not be (ϵ, δ) differentially private for*

$$\delta < 0.5, \epsilon < e^{-\frac{2}{x_h^2\beta}} \frac{\alpha}{2v_1}\left(\frac{3}{32x_h^2\beta}\right)^2 \left(\frac{c}{n}\right)^2 + \ln(0.5 - \delta)$$
$$v_1 = \max\{6, 1 + 2e^{\frac{1}{x_h^2\beta}}\}.$$

Proof of Lemma A.9. According to Definition III.1, if there exists a group, S, such that

$$p\left(\theta_{(\lceil \dot{k} \rceil+1)n} \in S | D_1\right) > e^{\epsilon} p\left(\hat{\theta}_{(\lceil \dot{k} \rceil+1)n} \in S | D_2\right) + \delta \tag{30}$$

then releasing $\theta_{(\lceil \dot{k} \rceil+1)n}$ is not (ϵ, δ) differentially private. We will show that eq. 30 is true for $S = \{s | s > \mu_{(\lceil \dot{k} \rceil+1)n}\}$ and the conditions of the lemma. First, notice that eq. 30 can be rearranged as

$$e^{\epsilon} p\left(\hat{\theta}_{(\lceil \dot{k} \rceil+1)n} \in S | D_2\right) + \delta - p\left(\hat{\theta}_{(\lceil \dot{k} \rceil+1)n} \in S | D_1\right) \leq 0$$

By Claim A.8, and since $\theta_{(\lceil \dot{k} \rceil+1)n} \sim \mathbb{N}(\theta_{(\lceil \dot{k} \rceil+1)n}; \mu_{(\lceil \dot{k} \rceil+1)n}, \sigma^2_{(\lceil \dot{k} \rceil+1)n})$,

$$e^{\epsilon} p\left(\hat{\theta}_{(\lceil \dot{k} \rceil+1)n} \in S | D_2\right) + \delta - p\left(\hat{\theta}_{(\lceil \dot{k} \rceil+1)n} \in S | D_1\right) \leq e^{\epsilon} e^{-e^{-\frac{2}{x_h^2\beta}} \frac{\alpha}{2v_1}\left(\frac{3}{32x_h^2\beta}\right)^2 \left(\frac{c}{n}\right)^2} + \delta - 0.5. \tag{31}$$

Therefore, if

$$e^{\epsilon} e^{-e^{-\frac{2}{x_h^2\beta}} \frac{\alpha}{2v_1}\left(\frac{3}{32x_h^2\beta}\right)^2 \left(\frac{c}{n}\right)^2} + \delta - 0.5 < 0$$

then eq. 30 is true and the lemma is proved. As shown in eq. 32, this inequality holds under the conditions of Lemma A.9.

$$\begin{aligned}
&e^{\epsilon - e^{-\frac{2}{x_h^2\beta}} \frac{\alpha}{2v_1}(\frac{3}{32x_h^2\beta})^2 (\frac{c}{n})^2} + \delta - 0.5 < 0 \iff \\
&e^{\epsilon - e^{-\frac{2}{x_h^2\beta}} \frac{\alpha}{2v_1}(\frac{3}{32x_h^2\beta})^2 (\frac{c}{n})^2} < 0.5 - \delta \iff \\
&\epsilon - e^{-\frac{2}{x_h^2\beta}} \frac{\alpha}{2v_1}(\frac{3}{32x_h^2\beta})^2 (\frac{c}{n})^2 < \ln(0.5 - \delta) \iff \\
&\epsilon < e^{-\frac{2}{x_h^2\beta}} \frac{\alpha}{2v_1}(\frac{3}{32x_h^2\beta})^2 (\frac{c}{n})^2 + \ln(0.5 - \delta)
\end{aligned} \tag{32}$$

□

E. Figure 1 derivation

The lower bound depicted in figure 1 is derived by analysing the distributions of SGLD running on datasets D_1 and D_2 (defined in 8). Given these distributions, and using notations $\theta_j, \mu_j, \hat{\theta}_j, \hat{\mu}_j, \hat{\sigma}_j^r$ (defined in subsection IV-B), the lower bound over ϵ for a given δ can be deduced as follows:

By definition III.1, for SGLD running on datasets D_1, D_2 to be (ϵ, δ)-DP, it must hold

$$p\left(\theta_j > \mu_j\right) \leq \exp\left(\epsilon\right) p(\hat{\theta}_j > \mu_j) + \delta.$$

This condition can be easily translated to a lower bound over ϵ:

$$\ln\left(p\left(\theta_j > \mu_j\right) - \delta\right) - \ln(p(\hat{\theta}_j > \mu_j)) \leq \epsilon.$$

By Lemma IV.5,

$$p(\hat{\theta}_j > \mu_j) \leq \frac{1}{n} \sum_{r=1}^{n} \exp\left(-\frac{\left(\mu_j - \hat{\mu}_j^r\right)^2}{2\left(\hat{\sigma}_j^r\right)^2}\right).$$

As ln is monotonically increasing, this induces a necessary condition for SGLD to be (ϵ, δ)-DP:

$$\ln\left(p\left(\theta_j > \mu_j\right) - \delta\right) - \ln\left(\frac{1}{n} \sum_{r=1}^{n} \exp\left(-\frac{\left(\mu_j - \hat{\mu}_j^r\right)^2}{2\left(\hat{\sigma}_j^r\right)^2}\right)\right) \leq \epsilon.$$

In figure 1 we plot the value

$$\max\left\{0, \ln\left(p\left(\theta_j > \mu_j\right) - \delta\right) - \ln\left(\frac{1}{n} \sum_{r=1}^{n} \exp\left(-\frac{\left(\mu_j - \hat{\mu}_j^r\right)^2}{2\left(\hat{\sigma}_j^r\right)^2}\right)\right)\right\}.$$

APPENDIX B
PROPOSE TEST SAMPLE SUPPLEMENTARY

n_{min}, which is used in Algorithm 1, is defined as follows:

$$
\begin{aligned}
&\nu = \frac{2\ln(\frac{1}{\delta})}{\epsilon} + 1 \\
&n_{b1} = \max\{1 + \frac{x_h^2}{x_l^2}\frac{8}{\epsilon}, 1 + \nu\frac{x_h^2}{x_l^2}\left(1 + 8\frac{\nu - 1}{\epsilon}\right), \left(\frac{16\nu\beta x_h^4}{\frac{9}{10}\epsilon x_l^2}\right)^{\frac{1}{1-2\gamma_1}}, \\
&\left(\frac{32\nu}{\epsilon}\cdot\frac{x_h^4\left(\alpha + x_h^2\beta\right)}{\frac{9}{10}x_l^4}\cdot\breve{m}\right)^{\frac{1}{2-\gamma_1}}, \left(\frac{32\nu}{\epsilon}\cdot\frac{x_h^4\left(\alpha + x_h^2\beta\right)}{\frac{9}{10}x_l^4}\right)^{\frac{1}{2-2\gamma_1}}, \\
&\left(\frac{4\nu}{\epsilon}\cdot\frac{x_h^4\left(\alpha + x_h^2\beta\right)^2}{\frac{9}{10}x_l^6\beta}\right)^{\frac{1}{3}}\cdot(2\breve{m})^{\frac{2}{3}}, \left(\frac{16\nu}{\epsilon}\cdot\frac{x_h^4\left(\alpha + x_h^2\beta\right)^2}{\frac{9}{10}x_l^6\beta}\right)^{\frac{1}{3-2\gamma_1}}\} \\
&n_{b2} = \max\{1 + \frac{x_h^2}{x_l^2}\frac{10\nu}{\beta}, 1 + \nu\frac{x_h^2}{x_l^2}\} \\
&n_{min} = \max\{n_{b1}, n_{b2}, n_1^{\frac{\rho_2}{\gamma_1}}\}
\end{aligned}
\tag{33}
$$

where $\epsilon, \delta, x_l, x_h, \alpha, \beta, \gamma_1, \rho_1, \rho_2, n_1$ are parameters of Algorithm 1.

APPENDIX C
PROPOSE TEST SAMPLE PRIVACY

Appendix C provides auxiliary claims and proof for Claim IV.9, along with auxiliary claims which support the proof of Claim IV.8. This subsection uses definitions and notations defined in subsection IV-C, specifically: D_3, D_4 (defined in eq. 9), $\epsilon, \delta, x_l, x_h, \alpha, \beta, \gamma_1, \rho_1, \rho_2$ - parameters of algorithm 1, D input dataset to algorithm 1 of size n_1, $V, n_2, m, \breve{m}, n_W, W, p(\theta|W), l_1, l_2, l_3, l_4$ (defined in algorithm 1), and n_{min} (defined in eq. 33).

Proof of Claim IV.9. We analyze Algorithm 1, running on datasets D_3 and D_4 defined in eq. 9, with parameters values:

$$
\begin{aligned}
&\rho_3 = 1.15; \rho_2 = 0.45; \rho_1 = 1.25, \gamma_1 = 0.49; \\
&\beta = 3; x_h = 1; x_l = 0.5; \alpha = 1 \\
&n_1 > \max\{\frac{4}{\epsilon}\log\frac{1}{2\delta}, 2^{10\rho_1}, 2^{9+10\rho_2}\}.
\end{aligned}
\tag{34}
$$

Note that we only define a lower bound over n_1, which will be updated later on.

Mark the return value of the algorithm as r, the event of the algorithm running on dataset D_3 and $W = D_3$ as A_{D_3}, the event of the algorithm running on dataset D_4 and $W = D_4$ as A_{D_4}, and $S = \{s|s > \mu_i\}$, where μ_i is the mean of the sample distribution at the SGLD i'th step given dataset D_3 (similarly to the definition of S in subsection IV-B). We will show that $\forall\epsilon \in \mathbb{R}_{>0}, \delta < \frac{1}{6}$ there exists n_1 such that

$$
p(r \in S|D_3) > e^{\epsilon}p(r \in S|D_4) + \delta. \tag{35}
$$

We first show that

$$
\begin{aligned}
&p(r \in S \wedge A_{D_3}^c|D_3) = 0 \\
&p(r \in S \wedge A_{D_4}^c|D_4) = 0.
\end{aligned}
\tag{36}
$$

Notice that $r \in S$ only if the algorithm reached line 25. Consider an event where the algorithm reached line 25 and $A_{D_3}^c$. Because $A_{D_3}^c$, $\exists(x_i, y_i) \in D_3$ such that $|\frac{y_i}{x_i} - \breve{m}| \geq n_2^{\rho_2}$. However, since $\forall(x_i, y_i) \in D_3 : \frac{y_i}{x_i} = n_1^{\rho_3}$ then $\forall(x_i, y_i) \in D_3 : |\frac{y_i}{x_i} - \breve{m}| > n_2^{\rho_2}$ and therefore $|W| = 0$. Under the assumption that a sample from $p(\theta|\{\})$ returns $null$, the algorithm, in this case, also returns $null$ and therefore $P(r \in S \wedge A_{D_3}^c|D_3) = 0$. Same arguments hold for D_4.

Because we showed eq. 36 is true, then to prove eq. 35, it is enough to show that eq. 37 is true.

$$
\begin{aligned}
&p(r \in S|D_3, A_{D_3})p(A_{D_3}|D_3) \geq^{*} p(r \in S|D_3, A_{D_3}) - 5\delta >^{**} e^{\epsilon}p(r \in S|D_4, A_{D_4}) + \delta \geq \\
&e^{\epsilon}p(r \in S|D_4, A_{D_4})p(A_{D_4}|D_4) + \delta = e^{\epsilon}p(r \in S \wedge A_{D_4}|D_4) + \delta
\end{aligned}
\tag{37}
$$

From Claim C.1, $\exists n_{bound_1}$ such that $\forall n_1 > n_{bound_1}$ inequality * holds. From Lemma IV.7, for n_1 big enough, $\exists T \in \mathbb{Z}_{>0}$ such that eq. 38 hold (Where $6\delta < 0.5$ according to the claim's conditions). Therefore, $\exists k, n_{bound2} \in \mathbb{R}_{>0}$ such that $\forall n_1 > n_{bound_2} : \epsilon' > kn_1^{2(1-\rho_3)}$ and eq. 38 hold. As $\rho_3 > 1$, by choosing $n_1 > \max\{n_{bound2}, (\frac{\epsilon}{k})^{\frac{1}{2(\rho_3-1)}}\}$ we get that $\epsilon' > \epsilon$. Consequently, by choosing $n_1 > \max\{n_{bound_1}, n_{bound_2}, (\frac{\epsilon}{k})^{\frac{1}{2(\rho_3-1)}}\}$, inequalities * and ** hold, and the claim is proved.

$$\begin{aligned}&\epsilon' = \Omega(n_1^{2(\rho_3-1)})\\&p(r \in S|D_3, A_{D_3}) > e^{\epsilon'} p(r \in S|D_4, A_{D_4}) + 6\delta\end{aligned} \tag{38}$$

□

Claim C.1. *Given a run of Algorithm 1 on dataset D_3, mark by A the event of the algorithm reaching line 25 with $W = D_3$; the following holds:*

$$\exists n_{bound_1} \in \mathbb{Z}_{>0} \ s.t. \ \forall n_1 > n_{bound_1} : p(A) \geq 1 - 5\delta.$$

Proof. For abbreviation, mark the event of $n_W > n_{min} \wedge \breve{m} \in [m - n_2^{\rho_2}, m + n_2^{\rho_2}] \wedge n_2^{1+\rho_2-0.1} > \breve{n}_1^{\rho_1} \wedge \breve{n}_1 \leq n_1 \wedge V = D$ as B. Since $P(A|D_3, B) = 1$, then $P(A|D_3) \geq P(A \wedge B|D_3) = P(A|B, D_3)P(B|D_3) = P(B|D_3)$. Therefore, we can prove the claim by showing the existence of n_{lb} such that $\forall n_1 > n_{lb} : P(B|D_3) \geq 1 - 5\delta$. We do so in eq. 39:

$$\begin{aligned}&p(B|D_3) =\\&p(\breve{m} \in [m - n_2^{\rho_2}, m + n_2^{\rho_2}] \wedge n_W > n_{min}|D_3, V = D, n_2^{1+\rho_2-0.1} > \breve{n}_1^{\rho_1}, \breve{n}_1 \leq n_1)\cdot\\&p(n_2^{1+\rho_2-0.1} > \breve{n}_1^{\rho_1}|V = D, \breve{n}_1 \leq n_1, D_3)P(V = D, \breve{n}_1 \leq n_1|D_3) \geq\\&p(\breve{m} \in [m - n_2^{\rho_2}, m + n_2^{\rho_2}] \wedge n_W > n_{min}|D_3, V = D, n_2^{1+\rho_2-0.1} > \breve{n}_1^{\rho_1}, \breve{n}_1 \leq n_1) - 3\delta =\\&p(n_W > n_{min}|D_3, V = D, n_2^{1+\rho_2-0.1} > \breve{n}_1^{\rho_1}, \breve{n}_1 \leq n_1, \breve{m} \in [m - n_2^{\rho_2}, m + n_2^{\rho_2}])\\&p(\breve{m} \in [m - n_2^{\rho_2}, m + n_2^{\rho_2}]|D_3, V = D, n_2^{1+\rho_2-0.1} > \breve{n}_1^{\rho_1}, \breve{n}_1 \leq n_1) - 3\delta \geq\\&p(n_W > n_{min}|D_3, V = D, n_2^{1+\rho_2-0.1} > \breve{n}_1^{\rho_1}, \breve{n}_1 \leq n_1, \breve{m} \in [m - n_2^{\rho_2}, m + n_2^{\rho_2}]) - 4\delta \geq\\&1 - 5\delta\end{aligned} \tag{39}$$

where by Corollary C.3 and Claim C.4, for n_1 big enough, first inequality holds. By Claim C.5, for n_1 big enough, second inequality holds. Lastly, by Claim C.6, for n_1 big enough, third inequality holds. □

Claim C.2. *For $n_1 > \max\{2^{10\rho_1}, \frac{4}{\epsilon}\log\frac{1}{2\delta}\}$:*

$$p(\breve{n}_1^{\rho_1} \geq n_1^{\rho_3} \wedge \breve{n}_1 \leq n_1|D_3) \geq 1 - 2\delta.$$

Proof of Claim C.2.

$$\begin{aligned}&p(\breve{n}_1^{\rho_1} \geq n_1^{\rho_3} \wedge \breve{n}_1 \leq n_1|D_3) =\\&p((n_1 + l_1 - \frac{1}{\epsilon}\log\frac{1}{2\delta})^{\rho_1} \geq n_1^{\rho_3} \wedge n_1 + l_1 - \frac{1}{\epsilon}\log\frac{1}{2\delta} \leq n_1|D_3) =\\&p((n_1 + l_1 - \frac{1}{\epsilon}\log\frac{1}{2\delta})^{\rho_1} \geq n_1^{\rho_3} \wedge l_1 \leq \frac{1}{\epsilon}\log\frac{1}{2\delta}|D_3) =\\&p((n_1 + l_1 - \frac{1}{\epsilon}\log\frac{1}{2\delta})^{\rho_1} \geq n_1^{\rho_3} \wedge |l_1| \leq \frac{1}{\epsilon}\log\frac{1}{2\delta}|D_3) +\\&p((n_1 + l_1 - \frac{1}{\epsilon}\log\frac{1}{2\delta})^{\rho_1} \geq n_1^{\rho_3} \wedge l_1 \leq -\frac{1}{\epsilon}\log\frac{1}{2\delta}|D_3) \geq\\&p((n_1 + l_1 - \frac{1}{\epsilon}\log\frac{1}{2\delta})^{\rho_1} \geq n_1^{\rho_3} \wedge |l_1| \leq \frac{1}{\epsilon}\log\frac{1}{2\delta}|D_3) =\\&p((n_1 + l_1 - \frac{1}{\epsilon}\log\frac{1}{2\delta})^{\rho_1} \geq n_1^{\rho_3}||l_1| \leq \frac{1}{\epsilon}\log\frac{1}{2\delta}, D_3)p(|l_1| \leq \frac{1}{\epsilon}\log\frac{1}{2\delta}|D_3)\end{aligned} \tag{40}$$

As

$$p(|l_1| \leq \frac{1}{\epsilon}\log\frac{1}{2\delta}) = 1 - 2p(l_1 \leq -\frac{1}{\epsilon}\log\frac{1}{2\delta}) = 1 - \exp(-\frac{\frac{1}{\epsilon}\log\frac{1}{2\delta}}{\frac{1}{\epsilon}}) = 1 - 2\delta,$$

we can further develop eq. 40:

$$\begin{aligned}&p((n_1 + l_1 - \frac{1}{\epsilon}\log\frac{1}{2\delta})^{\rho_1} \geq n_1^{\rho_3}||l_1| \leq \frac{1}{\epsilon}\log\frac{1}{2\delta}, D_3)p(|l_1| \leq \frac{1}{\epsilon}\log\frac{1}{2\delta}|D_3) \geq\\&p((n_1 - 2\frac{1}{\epsilon}\log\frac{1}{2\delta})^{\rho_1} \geq n_1^{\rho_3}|D_3) - 2\delta\end{aligned}$$

Finally, because $n_1 > 2^{10\rho_1}$, then $n_1^{-0.1} < (\frac{1}{2})^{\rho_1}$. Therefore, $(n_1 - 2\frac{1}{\epsilon}\log\frac{1}{2\delta})^{\rho_1} > (\frac{1}{2}n_1)^{\rho_1} > n_1^{\rho_1-0.1} = n_1^{\rho_3}$. Which leads to

$$p((n_1 - 2\frac{1}{\epsilon}\log\frac{1}{2\delta})^{\rho_1} \geq n_1^{\rho_3}|D_3) - 2\delta = 1 - 2\delta$$

□

Noticing that the conditions of corollary C.3 includes the conditions of Claim C.2, and that given that the algorithm runs on dataset D_3 and $\breve{n}_1^{\rho_1} \geq n_1^{\rho_3}$ then $V = D_3$, we get that Corollary C.3 is a direct result of Claim C.2.

Corollary C.3. $\forall n_1 > \max\{2^{10\rho_1}, 4\frac{1}{\epsilon}\log\frac{1}{2\delta}\}$, *when running Algorithm 1 on dataset D_3, the following inequality holds:*

$$P(V = D_3 \wedge \breve{n}_1 \leq n_1) \geq 1 - 2\delta.$$

Claim C.4. $\forall n_1 > \max\{\frac{1}{2}^{-(9+10\rho_2)}, 4\frac{1}{\epsilon}\log\frac{1}{2\delta}\}$, *when running Algorithm 1 on dataset D_3 the following inequality holds:*

$$p(n_2^{1+\rho_2-0.1} > \breve{n}_1^{\rho_1} | D_3, V = D_3, \breve{n}_1 \leq n_1) \geq 1 - \delta.$$

Proof of Claim C.4.

$$\begin{aligned}
&p(n_2^{0.9+\rho_2} > \breve{n}_1^{\rho_1} | D_3, V = D_3, \breve{n}_1 \leq n_1) \geq p(n_2^{0.9+\rho_2} > n_1^{\rho_1} | D_3, V = D_3) \geq \\
&p((n_1 - 2\frac{1}{\epsilon}\log\frac{1}{2\delta})^{0.9+\rho_2} > n_1^{\rho_1} | D_3) p(n_2 \geq |V| - 2\frac{1}{\epsilon}\log\frac{1}{2\delta}) \geq \\
&p((n_1 - 2\frac{1}{\epsilon}\log\frac{1}{2\delta})^{0.9+\rho_2} > n_1^{\rho_1} | D_3) - \delta = 1 - \delta
\end{aligned}$$

where the second inequality holds since $p(\text{Lap}(\frac{1}{\epsilon}) < -\frac{1}{\epsilon}\log\frac{1}{2\delta}) < \delta$, and last equality holds since $(n_1 - 2\frac{1}{\epsilon}\log\frac{1}{2\delta})^{0.9+\rho_2} > (\frac{1}{2}n_1)^{0.9+\rho_2} > n_1^{1+\rho_2-0.2} = n_1^{\rho_1}$ □

Claim C.5. $\exists n_{lb_1} \in \mathbb{Z}_{>0}$ *such that* $\forall n_1 > n_{lb_1}$, *when running Algorithm 1 on dataset D_3, the following inequality holds:*

$$p(\breve{m} \in [m - n_2^{\rho_2}, m + n_2^{\rho_2}] | D_3, n_2^{0.9+\rho_2} > \breve{n}_1^{\rho_1}) \geq 1 - \delta.$$

Proof of Claim C.5.

$$\begin{aligned}
&p(\breve{m} \in [m - n_2^{\rho_2}, m + n_2^{\rho_2}] | D_3, n_2^{0.9+\rho_2} > \breve{n}_1^{\rho_1}) = p(l_3 \in [-n_2^{\rho_2}, n_2^{\rho_2}] | D_3, n_2^{0.9+\rho_2} > \breve{n}_1^{\rho_1}) \geq \\
&1 - 2(\frac{1}{2}\exp(-n_2^{\rho_2}\frac{1}{\frac{1}{\epsilon}\breve{n}_1^{\rho_1}\frac{2(n_2-1)x_h^2x_l^2+x_h^4}{n_2(n_2-1)x_l^4}}) = 1 - \exp(-\frac{n_2^{1+\rho_2}\epsilon(n_2-1)x_l^4}{\breve{n}_1^{\rho_1}(2(n_2-1)x_h^2x_l^2 + x_h^4)})
\end{aligned}$$

Because $n_2^{0.9+\rho_2} > \breve{n}_1^{\rho_1}$, then $\breve{n}_1^{\rho_1} = o(n_2^{1+\rho_2})$, and therefore for n_1 big enough, the exponent is smaller than δ. □

Claim C.6. $\exists n_{lb_2} \in \mathbb{Z}_{>0}$ *such that* $\forall n_1 > n_{lb_2}$, *when running Algorithm 1 on dataset D_3, the following inequality holds:*

$$p(n_W > n_{min} | |V| = D_3 \wedge n_2^{0.9+\rho_2} > \breve{n}_1^{\rho_1} \wedge \breve{m} \in [m - n_2^{\rho_2}, m + n_2^{\rho_2}], D_3) \geq 1 - \delta.$$

Proof of Claim C.6. For abbreviation, mark event B as when the following apply $|V| = D_3 \wedge n_2^{0.9+\rho_2} > \breve{n}_1^{\rho_1} \wedge \breve{m} \in [m - n_2^{\rho_2}, m + n_2^{\rho_2}]$.

$$\begin{aligned}
&p(n_W > n_{min} | B, D_3) = p(n_1 - \frac{1}{\epsilon}\log\frac{1}{2\delta} + l_4 > n_{min} | B, D_3) > \\
&p(n_1 - \frac{1}{\epsilon}\log\frac{1}{2\delta} - \frac{1}{\epsilon}\log\frac{1}{\delta} > n_{min} | B, D_3) p(l_4 > -\frac{1}{\epsilon}\log\frac{1}{\delta}) + \\
&p(n_1 - \frac{1}{\epsilon}\log\frac{1}{2\delta} + l_4 > n_{min} \wedge l_4 < -\frac{1}{\epsilon}\log\frac{1}{\delta} | B, D_3) > \\
&p(n_1 - \frac{2}{\epsilon}\log\frac{1}{2\delta} - \frac{1}{\epsilon}\log\frac{1}{\delta} > n_{min} | B, D_3)(1 - \frac{\delta}{2}) > p(n_1 - \frac{2}{\epsilon}\log\frac{1}{2\delta} - \frac{1}{\epsilon}\log\frac{1}{\delta} > n_{min} | B, D_3) - \frac{\delta}{2} = \\
&p(n_1 - \frac{2}{\epsilon}\log\frac{1}{2\delta} - \frac{1}{\epsilon}\log\frac{1}{\delta} > n_{min} | l_2 < \frac{1}{\epsilon}\log\frac{1}{\delta}, B, D_3) p(l_2 < \frac{1}{\epsilon}\log\frac{1}{\delta} | B, D_3) + \\
&p(n_1 - \frac{2}{\epsilon}\log\frac{1}{2\delta} - \frac{1}{\epsilon}\log\frac{1}{\delta} > n_{min} \wedge l_2 \geq \frac{1}{\epsilon}\log\frac{1}{\delta} | B, D_3) - \frac{\delta}{2} \geq \\
&p(n_1 - \frac{2}{\epsilon}\log\frac{1}{2\delta} - \frac{1}{\epsilon}\log\frac{1}{\delta} > n_{min} | l_2 < \frac{1}{\epsilon}\log\frac{1}{\delta}, B, D_3) p(l_2 < \frac{1}{\epsilon}\log\frac{1}{\delta} | B, D_3) - \frac{\delta}{2} \geq \\
&p(n_1 - \frac{2}{\epsilon}\log\frac{1}{2\delta} - \frac{1}{\epsilon}\log\frac{1}{\delta} > n_{min} | l_2 < \frac{1}{\epsilon}\log\frac{1}{\delta}, B, D_3) - \delta
\end{aligned}$$

From B, $|m - \breve{m}| < n_2^{\rho_2}$, and therefore $\breve{m} < m + n_2^{\rho_2}$, and for the case of $l_2 < \frac{1}{\epsilon}\log\frac{1}{\delta}$ it holds that

$$n_2 < n_1 - \frac{1}{\epsilon}\log\frac{1}{2\delta} + \frac{1}{\epsilon}\log\frac{1}{\delta} < n_1 + \frac{1}{\epsilon}\log\frac{1}{\delta}.$$

Therefore $\breve{m} \leq m + (n_1 + \frac{1}{\epsilon}\log\frac{1}{\delta})^{\rho_2}$. As $n_{min} = \mathcal{O}(\max\{\breve{m}^{\frac{2}{3}}, n_1^{\frac{\rho_2}{\gamma_1}}\})$ then for the case of $l_2 < \frac{1}{\epsilon}\log\frac{1}{\delta}$ and B, it holds that $n_{min} = \mathcal{O}(\max\{(m + n_1^{\rho_2})^{\frac{2}{3}}, n_1^{\frac{\rho_2}{\gamma_1}}\}) = \mathcal{O}(\max\{n_1^{\frac{2\rho_3}{3}}, n_1^{\frac{\rho_2}{\gamma_1}}\}) < o(n_1)$; therefore $\exists n_{lb_2}$ such that

$$\forall n_1 > n_{lb_2} : n_1 - \frac{2}{\epsilon}\log\frac{1}{2\delta} - \frac{1}{\epsilon}\log\frac{1}{\delta} > n_{min}.$$

Consequently, $\forall n_1 > n_{lb_2} : p(n_1 - \frac{2}{\epsilon}\log\frac{1}{2\delta} - \frac{1}{\epsilon}\log\frac{1}{\delta} > n_{min} | l_2 < \frac{1}{\epsilon}\log\frac{1}{\delta}, B, D_3) = 1$. □

Definition C.7. *A randomized function $f(X, y) : \chi^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}$, is (ϵ, δ)-differentially private with respect to X if $\forall S \subseteq \mathbb{R}$, and $\forall X, \hat{X} \in \chi^n : \|X - \hat{X}\| \leq 1$, eq. 41 holds.*

$$p(f(X, y) \in S) \leq \exp(\epsilon)p(f(\hat{X}, y) \in S) + \delta \tag{41}$$

Definition C.8 (l_1-sensitivity, [11])**.** *The l_1-sensitivity of a function $f : \mathbb{N}^{|\chi|} \to \mathbb{R}^k$ is:*

$$\Delta f = \max_{x,y \in \mathbb{N}^{|\chi|}, \|x-y\|_1 = 1} \|f(x) - f(y)\|_1$$

Claim C.9. *For Algorithm 1, calculating $\breve{n}_1, n_2$ is $(2\epsilon, 0)$ differentially private.*

Proof of Claim C.9. n_1 l_1-sensitivity is 1; therefore, calculating $\breve{n}$ is $(\epsilon, 0)$ DP by the Laplace mechanism's privacy guarantees. For a given $\breve{n}_1$ value, the l_1-sensitivity of $|V|$ is 1. Therefore, given $\breve{n}_1$, calculating n_2 is $(\epsilon, 0)$ DP by the Laplace mechanism's privacy guarantees. Consequently, by the sequential composition theorem, the composition is $(2\epsilon, 0)$ differentially private. □

Claim C.10. *For Algorithm 1, $p(n_2 \leq |V| | D, \breve{n}_1) = 1 - \delta$.*

Proof of Claim C.10.

$$p(n_2 \leq |V| | D, \breve{n}_1) = p(|V| - \frac{1}{\epsilon}\log\frac{1}{2\delta} + l_2 \leq |V| | D, \breve{n}_1) = p(l_2 \leq \frac{1}{\epsilon}\log\frac{1}{2\delta} | D, \breve{n}_1) = 1 - \frac{1}{2}\exp(-\frac{\frac{1}{\epsilon}\log\frac{1}{2\delta}}{\frac{1}{\epsilon}}) = 1 - \delta$$

□

Claim C.11. *Given $\breve{n}_1, n_2$ and $n_2 < |V|$, calculating $\breve{m}$ in Algorithm 1 is $(\epsilon, 0)$ differentially private with respect to D.*

Proof of Claim C.11. Given two neighbouring datasets, D_5 and $\hat{D}_5$, mark by V_5 and $\hat{V}_5$ the realizations of V (calculated at line 10) when Algorithm 1 runs on each of the datasets, respectively. If $V_5 = \hat{V}_5$ then the claim follows trivially. In case they differ, assume w.l.o.g that $|V_5| \geq |\hat{V}_5|$, and that if $|V_5| = |\hat{V}_5|$ then they differ in their last sample. Define $q = \sum_{(x_i,y_i) \in V_5/\{x_{|V_5|}, y_{|V_5|}\}} x_i y_i$, $z = \sum_{(x_i,y_i) \in V_5/\{x_{|V_5|}, y_{|V_5|}\}} x_i^2$.

$$\begin{aligned}
&|\frac{q + x_{|V_5|}y_{|V_5|}}{z + x_{|V_5|}^2} - \frac{q + \hat{x}_{|V_5|}\hat{y}_{|V_5|}}{z + \hat{x}_{|V_5|}^2}| = \\
&|\frac{q\hat{x}_{|V_5|}^2 + x_{|V_5|}y_{|V_5|}\hat{x}_{|V_5|}^2 + x_{|V_5|}y_{|V_5|}z - qx_{|V_5|}^2 - \hat{x}_{|V_5|}\hat{y}_{|V_5|}x_{|V_5|}^2 - \hat{x}_{|V_5|}\hat{y}_{|V_5|}z}{(z + x_{|V_5|}^2)(z + \hat{x}_{|V_5|}^2)}| \leq \\
&\frac{qx_h^2 + \breve{n}_1^{\rho_1}x_h^2 z + \breve{n}_1^{\rho_1}x_h^4}{(z + x_l^2)z} \leq \breve{n}_1^{\rho_1}\frac{2zx_h^2 + x_h^4}{(z + x_l^2)z} = \breve{n}_1^{\rho_1}(\frac{2x_h^2}{z + x_l} + \frac{x_h^4}{(z + x_l^2)z}) \leq \\
&\breve{n}_1^{\rho_1}(\frac{2x_h^2}{|V_5|x_l^2} + \frac{x_h^4}{|V_5|(|V_5| - 1)x_l^4}) \leq \breve{n}_1^{\rho_1}(\frac{2x_h^2}{n_2 x_l^2} + \frac{x_h^4}{n_2(n_2 - 1)x_l^4}) = \breve{n}_1^{\rho_1}\frac{2(n_2 - 1)x_h^2 x_l^2 + x_h^4}{n_2(n_2 - 1)x_l^4}
\end{aligned}$$

Therefore, by the Laplace mechanism's privacy guarantees, calculating $\breve{m}$ is $(\epsilon, 0)$ differentially private. □

Claim C.12. *Lines 8-15 of Algorithm 1 are $(3\epsilon, \delta)$ differentially private.*

Proof of Claim C.12. Define $\hat{D}$ as a neighbouring dataset to D.

$$
\begin{aligned}
&p(\breve{m} \in S|D) = \\
&\int_{r_1,r_2\in\mathbb{R}_{>0}\times\mathbb{R}_{>0}} p(\breve{m} \in S|D, \breve{n}_1 = r_1, n_2 = r_2)p(\breve{n}_1 = r_1, n_2 = r_2|D)dr_1dr_2 = \\
&\int_{r_1,r_2\in\mathbb{R}_{>0}\times[1,|V|]} p(\breve{m} \in S|D, \breve{n}_1 = r_1, n_2 = r_2)p(\breve{n}_1 = r_1, n_2 = r_2|D)dr_1dr_2 + \\
&\int_{r_1,r_2\in\mathbb{R}_{>0}\times(|V|,\infty]} p(\breve{m} \in S|D, \breve{n}_1 = r_1, n_2 = r_2)p(\breve{n}_1 = r_1, n_2 = r_2|D)dr_1dr_2 \leq^{*} \\
&\int_{r_1,r_2\in\mathbb{R}_{>0}\times[1,|V|]} p(\breve{m} \in S|D, \breve{n}_1 = r_1, n_2 = r_2)p(\breve{n}_1 = r_1, n_2 = r_2|D)dr_1dr_2 + \delta \leq^{**} \\
&\int_{r_1,r_2\in\mathbb{R}_{>0}\times[1,|V|]} e^{2\epsilon}p(\breve{m} \in S|\hat{D}, \breve{n}_1 = r_1, n_2 = r_2)p(\breve{n}_1 = r_1, n_2 = r_2|\hat{D})dr_1dr_2 + \delta \leq \\
&\int_{r_1,r_2\in\mathbb{R}_{>0}\times\mathbb{R}_{>0}} e^{2\epsilon}p(\breve{m} \in S|\hat{D}, \breve{n}_1 = r_1, n_2 = r_2)p(\breve{n}_1 = r_1, n_2 = r_2|\hat{D})dr_1dr_2 + \delta = \\
&e^{2\epsilon}p(\breve{m} \in S|\hat{D}) + \delta
\end{aligned}
$$

where inequality * follows Claim C.10 and inequality ** follows Claims C.11 and C.9. □

Claim C.13. *Given $n_2, \breve{m}$ and $|W| < n_{min}$, lines 19-25 of Algorithm 1 are (ϵ, δ) differentially private with respect to D.*

Proof of Claim C.13. Mark $l \sim \text{Lap}(\frac{1}{\epsilon})$, and $\hat{D}$ as a neighbouring dataset to D. Eq. 42 proves the claim.

$$
\begin{aligned}
&p(S|D, |W| < n_{min}, \breve{m}, n_2) = p(S \cap \{null\}|D, |W| < n_{min}, \breve{m}, n_2) + p(S \cap \{null\}^c|D, |W| < n_{min}, \breve{m}, n_2) \leq \\
&e^{\epsilon}p(S \cap \{null\}|\hat{D}, |W| < n_{min}, \breve{m}, n_2) + \delta \leq e^{\epsilon}p(S|\hat{D}, |W| < d, \breve{m}, n_2) + \delta
\end{aligned} \tag{42}
$$

where the first inequality is true from eq. 43 and the Laplace mechanism's privacy guarantees for n_W.

$$
p(null|D, |W| < n_{min}, \breve{m}, n_2) = p(n_W < n_{min} + \frac{1}{\epsilon}\log(\frac{1}{2\delta})|D, |W| < n_{min}, \breve{m}, n_2) \geq p(l < \frac{1}{\epsilon}\log(\frac{1}{2\delta})) \geq 1 - \delta \tag{43}
$$

□

Claim C.14. *Line 25 of Algorithm 1 is (ϵ, δ) differentially private with respect to D for $|W| \geq n_{min}$ and given $n_2, \breve{m}$.*

Proof of Claim C.14. For a given n_2 and $\breve{m}$ the group W can change by up to one sample for a neighbouring dataset. Mark $n = |W|$ and $c = \breve{m}$. As $n \geq n_2^{\frac{\rho_2}{\gamma_1}}$, then $n^{\frac{1}{2}} > n^{\gamma_1} \geq n_2^{\rho_2}$, and therefore $W \in \mathcal{D}(n, \gamma_1, x_h, x_l, c)$, as defined in eq. 5.

Because $W \in \mathcal{D}(n, \gamma_1, x_h, x_l, c)$, $n \geq n_{b1}$, and $n \geq n_{b2}$, the problem of sampling from $p(\theta|W)$ for $|W| \geq n_{min}$ holds the constraints of Claim D.27. Therefore one sample from $p(\theta|W)$ is (ϵ, δ) differentially private. □

Claim C.15. *Lines 19-24 of Algorithm 1 are $(\epsilon, 0)$ differentially private with respect to D for $|W| \geq n_{min}$ and given $\breve{m}, n_2$.*

Proof of Claim C.15. The only data released in lines 19-24 is n_W. Since the l_1-sensitivity of $|W|$ given $\breve{m}, n_2$ is 1, then the Laplace mechanism ensures $(\epsilon, 0)$ differential privacy. □

Corollary C.16. *Lines 19-25 of Algorithm 1 are $(2\epsilon, \delta)$ differentially private with respect to D for $|W| \geq n_{min}$ and given $\breve{m}, n_2$.*

Corollary C.16 follows directly from Claims C.15 and C.14.

Corollary C.17. *Lines 19-25 of Algorithm 1 are $(2\epsilon, \delta)$ differentially private with respect to D given $\breve{m}, n_2$.*

Corollary C.17 follows directly from Claims C.16 and C.13.

APPENDIX D
AUXILIARY CLAIMS

Appendix D contains claims used to simplify the otherwise complex proofs throughout the paper.

A. Stochastic Gradient Langevin Dynamics Privacy

This subsection provides auxiliary claims for SGLD privacy analysis performed in subsection A-D. It uses the notations defined in section IV, subsection IV-B, and subection A-D, specifically: $\alpha, \beta, \theta, p(y|x,\theta)$ (defined in eq. 4), $\mathcal{D}(n, \gamma_1, x_h, x_l, c)$ (defined in eq. 5), D_1, D_2 (defined in eq. 8), $\eta, \theta_j, \hat{\theta}_j, \mu_j, \hat{\mu}_j^r, \sigma_j, \hat{\sigma}_j^r$ (defined in subsection IV-B), and $\lambda, \hat{\lambda}, \rho, \hat{\rho}$ (defined in eq. 22).

Claim D.1. $\rho\frac{1}{1-\lambda} = \frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}$.

Proof of Claim D.1.

$$\rho\frac{1}{1-\lambda} = \frac{\eta}{2}ncx_h^2\beta\frac{1}{1-\left(1-\frac{\eta}{2}\left(\alpha+n\left(\frac{x_h}{2}\right)^2\beta\right)\right)} = ncx_h^2\beta\frac{1}{\alpha+nx_h^2\beta} = \frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}.$$

□

Claim D.2. $\rho\frac{1-\lambda^{n-1}}{1-\lambda} + \rho\lambda^{n-1} = \frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(1-\lambda^n\right)$.

Proof of Claim D.2.

$$\rho\frac{1-\lambda^{n-1}}{1-\lambda} + \rho\lambda^{n-1} = \rho\left(\frac{1-\lambda^{n-1}+\lambda^{n-1}-\lambda^n}{1-\lambda}\right) = \rho\left(\frac{1-\lambda^n}{1-\lambda}\right) = \frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(1-\lambda^n\right)$$

where the last equality holds from Claim D.1. □

Claim D.3. $\rho\left(\frac{1-\lambda^{n-1}}{1-\lambda}\right) + \hat{\rho}\lambda^{n-1} = \frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(1-\lambda^n\left(\frac{3}{4}\lambda^{-1}+\frac{1}{4}\right)\right)$.

Proof of Claim D.3.

$$\rho\left(\frac{1-\lambda^{n-1}}{1-\lambda}\right) + \hat{\rho}\lambda^{n-1} = \rho\left(\frac{1-\lambda^{n-1}}{1-\lambda}\right) + \rho\frac{1}{4}\lambda^{n-1} = \rho\left(\frac{1-\frac{3}{4}\lambda^{n-1}-\frac{1}{4}\lambda^n}{1-\lambda}\right) =$$
$$\rho\left(\frac{1-\lambda^n\left(\frac{3}{4}\lambda^{-1}+\frac{1}{4}\right)}{1-\lambda}\right) = \frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(1-\lambda^n\left(\frac{3}{4}\lambda^{-1}+\frac{1}{4}\right)\right)$$

where the last equality holds from Claim D.1. □

Claim D.4. $\frac{1}{4}\lambda + \frac{3}{4} - \hat{\lambda} = \frac{3}{4}\frac{\eta}{2}\alpha$.

Proof of Claim D.4.

$$\frac{1}{4}\lambda + \frac{3}{4} - \hat{\lambda} = \frac{1}{4}\left(1-\frac{\eta}{2}\left(\alpha+nx^2\beta\right)\right) + \frac{3}{4} - \left(1-\frac{\eta}{2}\left(\alpha+\frac{1}{4}nx^2\beta\right)\right) =$$
$$\frac{\eta}{2}\left(\alpha+\frac{1}{4}nx^2\beta-\frac{1}{4}\left(\alpha+nx^2\beta\right)\right) = \frac{3}{4}\frac{\eta}{2}\alpha.$$

□

Claim D.5. $\forall k \in \mathbb{Z}_{>0}$:

$$\left(1-\lambda^{kn}\right)\left(1-\hat{\lambda}\lambda^{n-1}\right) - \left(1-\left(\hat{\lambda}\lambda^{(n-1)}\right)\right)^k\left(1-\lambda^{n-1}\left(\frac{1}{4}\lambda+\frac{3}{4}\right)\right) =$$
$$\lambda^{n-1}\frac{3}{4}\frac{\eta}{2}\alpha\left(1-\hat{\lambda}^k\lambda^{k(n-1)}\right) + \lambda^{k(n-1)}\left(\hat{\lambda}^k-\lambda^k\right)\left(1-\lambda^{n-1}\hat{\lambda}\right).$$

Proof of Claim D.5.

$$\left(1-\lambda^{kn}\right)\left(1-\hat{\lambda}\lambda^{n-1}\right)-\left(1-\left(\hat{\lambda}\lambda^{(n-1)}\right)\right)^{k}\left(1-\lambda^{n-1}\left(\frac{1}{4}\lambda+\frac{3}{4}\right)\right)=$$
$$\lambda^{n-1}\left(\frac{1}{4}\lambda+\frac{3}{4}-\hat{\lambda}\right)+\lambda^{kn}\left(\hat{\lambda}\lambda^{n-1}-1\right)+\left(\hat{\lambda}\lambda^{n-1}\right)^{k}\left(1-\lambda^{n-1}\left(\frac{1}{4}\lambda+\frac{3}{4}\right)\right)=$$
$$\lambda^{n-1}\left(\frac{1}{4}\lambda+\frac{3}{4}-\hat{\lambda}\right)+\lambda^{k(n-1)}\left(\lambda^{k}\left(\hat{\lambda}\lambda^{n-1}-1\right)+\hat{\lambda}^{k}\left(1-\lambda^{n-1}\left(\frac{1}{4}\lambda+\frac{3}{4}\right)\right)\right)=$$
$$\lambda^{n-1}\left(\frac{1}{4}\lambda+\frac{3}{4}-\hat{\lambda}\right)+\lambda^{k(n-1)}\left(\hat{\lambda}^{k}\left(1-\lambda^{n-1}\left(\frac{1}{4}\lambda+\frac{3}{4}\right)\right)-\lambda^{k}\left(1-\hat{\lambda}\lambda^{n-1}\right)\right)=$$
$$\lambda^{n-1}\left(\frac{1}{4}\lambda+\frac{3}{4}-\hat{\lambda}\right)+\lambda^{k(n-1)}\left(\hat{\lambda}^{k}\left(1-\lambda^{n-1}\left(\frac{1}{4}\lambda+\frac{3}{4}\right)\right)-\lambda^{k}\left(1-\lambda^{n-1}\hat{\lambda}\right)\right)=^{*}$$
$$\lambda^{n-1}\frac{\eta}{2}\frac{3}{4}\alpha+\lambda^{k(n-1)}\left(\hat{\lambda}^{k}\left(1-\lambda^{n-1}\left(\frac{1}{4}\lambda+\frac{3}{4}\right)\right)-\lambda^{k}\left(1-\lambda^{n-1}\hat{\lambda}\right)\right)=^{*}$$
$$\lambda^{n-1}\frac{\eta}{2}\frac{3}{4}\alpha+\lambda^{k(n-1)}\left(\hat{\lambda}^{k}\left(1-\lambda^{n-1}\left(\hat{\lambda}+\frac{3}{4}\frac{\eta}{2}\alpha\right)\right)-\lambda^{k}\left(1-\lambda^{n-1}\hat{\lambda}\right)\right)=$$
$$\lambda^{n-1}\frac{\eta}{2}\frac{3}{4}\alpha-\hat{\lambda}^{k}\lambda^{n-1}\lambda^{k(n-1)}\frac{3}{4}\frac{\eta}{2}\alpha+\lambda^{k(n-1)}\left(\hat{\lambda}^{k}\left(1-\lambda^{n-1}\hat{\lambda}\right)-\lambda^{k}\left(1-\lambda^{n-1}\hat{\lambda}\right)\right)=$$
$$\lambda^{n-1}\frac{3}{4}\frac{\eta}{2}\alpha\left(1-\hat{\lambda}^{k}\lambda^{k(n-1)}\right)+\lambda^{k(n-1)}\left(\hat{\lambda}^{k}-\lambda^{k}\right)\left(1-\lambda^{n-1}\hat{\lambda}\right).$$

where equality marked by * holds from Claim D.4. □

Claim D.6. $\forall k\in\mathbb{Z}_{>0}:\lambda\sum_{j=0}^{k-1}\lambda^{(n-1)j}\lambda^{j}\left(\lambda^{n-1}\rho+\rho\sum_{i=0}^{n-2}\lambda^{i}\right)=\lambda\left(1-\lambda^{kn}\right)\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}$.

Proof of Claim D.6.

$$\lambda\sum_{j=0}^{k-1}\lambda^{(n-1)j}\lambda^{j}\left(\lambda^{n-1}\rho+\rho\sum_{i=0}^{n-2}\lambda^{i}\right)=\rho\lambda\sum_{i=0}^{kn-1}\lambda^{i}=\rho\lambda\frac{1-\lambda^{kn}}{1-\lambda}=^{*}\lambda\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(1-\lambda^{kn}\right)$$

where equality * follows from Claim D.1. □

Claim D.7. $\forall k\in\mathbb{Z}_{>0}:\hat{\lambda}\sum_{j=0}^{k-1}\lambda^{(n-1)j}\hat{\lambda}^{j}\left(\lambda^{n-1}\hat{\rho}+\rho\sum_{i=0}^{n-2}\lambda^{i}\right)=\hat{\lambda}\frac{1-\left(\lambda^{n-1}\hat{\lambda}\right)^{k}}{1-\lambda^{n-1}\hat{\lambda}}\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(1-\lambda^{n}\left(\frac{3}{4}\lambda^{-1}+\frac{1}{4}\right)\right)$.

Proof of Claim D.7.

$$\hat{\lambda}\sum_{j=0}^{k-1}\lambda^{(n-1)j}\hat{\lambda}^{j}\left(\lambda^{n-1}\hat{\rho}+\rho\sum_{i=0}^{n-2}\lambda^{i}\right)=$$
$$\hat{\lambda}\frac{1-\left(\lambda^{n-1}\hat{\lambda}\right)^{k}}{1-\lambda^{n-1}\hat{\lambda}}\left(\lambda^{n-1}\hat{\rho}+\rho\frac{1-\lambda^{n-1}}{1-\lambda}\right)=^{*}\hat{\lambda}\frac{1-\left(\lambda^{n-1}\hat{\lambda}\right)^{k}}{1-\lambda^{n-1}\hat{\lambda}}\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(1-\lambda^{n}\left(\frac{3}{4}\lambda^{-1}+\frac{1}{4}\right)\right)$$

where equality * follows from Claims D.1 and D.3. □

Claim D.8. $\forall k\in\mathbb{R}_{>0}:\lambda\lambda^{k}-\lambda^{k}\lambda^{n}\hat{\lambda}-\hat{\lambda}\hat{\lambda}^{k}+\hat{\lambda}\hat{\lambda}^{k}\lambda^{n}\left(\frac{3}{4}\lambda^{-1}+\frac{1}{4}\right)=\left(1-\hat{\lambda}\lambda^{n-1}\right)\left(\lambda^{k+1}-\hat{\lambda}^{k+1}\right)+\hat{\lambda}^{k+1}\lambda^{n-1}\left(\frac{3}{4}\frac{\eta}{2}\alpha\right)$.

Proof of Claim D.8.

$$\lambda\lambda^{k}-\lambda^{k}\lambda^{n}\hat{\lambda}-\hat{\lambda}\hat{\lambda}^{k}+\hat{\lambda}\hat{\lambda}^{k}\lambda^{n}\left(\frac{3}{4}\lambda^{-1}+\frac{1}{4}\right)=\lambda^{k+1}\left(1-\hat{\lambda}\lambda^{n-1}\right)-\hat{\lambda}^{k+1}\left(1-\lambda^{n-1}\left(\frac{1}{4}\lambda+\frac{3}{4}\right)\right)=^{*}$$
$$\lambda^{k+1}\left(1-\hat{\lambda}\lambda^{n-1}\right)-\hat{\lambda}^{k+1}\left(1-\lambda^{n-1}\left(\hat{\lambda}+\frac{3}{4}\frac{\eta}{2}\alpha\right)\right)=\left(1-\hat{\lambda}\lambda^{n-1}\right)\left(\lambda^{k+1}-\hat{\lambda}^{k+1}\right)+\hat{\lambda}^{k+1}\lambda^{n-1}\left(\frac{3}{4}\frac{\eta}{2}\alpha\right)$$

where equality * holds from Claim D.4. □

Claim D.9. $\forall k\in\mathbb{Z}_{>0}:$

$$\lambda\left(1-\lambda^{kn}\right)\left(1-\lambda^{n-1}\hat{\lambda}\right)-\hat{\lambda}\left(1-\left(\lambda^{n-1}\hat{\lambda}\right)^{k}\right)\left(1-\lambda^{n}\left(\frac{3}{4}\lambda^{-1}+\frac{1}{4}\right)\right)=$$
$$\left(\lambda-\hat{\lambda}\right)\left(1-\hat{\lambda}\lambda^{n-1}\right)+\lambda^{n-1}\hat{\lambda}\left(\frac{3}{4}\frac{\eta}{2}\alpha\left(1-\hat{\lambda}^{k}\lambda^{k(n-1)}\right)\right)+\lambda^{k(n-1)}\left(\left(1-\hat{\lambda}\lambda^{n-1}\right)\left(\hat{\lambda}^{k+1}-\lambda^{k+1}\right)\right).$$

Proof of Claim D.9.

$$\lambda\left(1-\lambda^{kn}\right)\left(1-\lambda^{n-1}\hat{\lambda}\right)-\hat{\lambda}\left(1-\left(\lambda^{n-1}\hat{\lambda}\right)^{k}\right)\left(1-\lambda^{n}\left(\frac{3}{4}\lambda^{-1}+\frac{1}{4}\right)\right)=$$
$$\lambda-\hat{\lambda}-\lambda^{n}\hat{\lambda}\left(1-\left(\frac{3}{4}\lambda^{-1}+\frac{1}{4}\right)\right)-\lambda^{k(n-1)}\left(\lambda\lambda^{k}-\lambda^{k}\lambda^{n}\hat{\lambda}-\hat{\lambda}\hat{\lambda}^{k}+\hat{\lambda}\hat{\lambda}^{k}\lambda^{n}\left(\frac{3}{4}\lambda^{-1}+\frac{1}{4}\right)\right)=^{*}$$
$$\lambda-\hat{\lambda}-\lambda^{n}\hat{\lambda}\left(1-\left(\frac{3}{4}\lambda^{-1}+\frac{1}{4}\right)\right)-\lambda^{k(n-1)}\left(\left(1-\hat{\lambda}\lambda^{n-1}\right)\left(\lambda^{k+1}-\hat{\lambda}^{k+1}\right)+\hat{\lambda}^{k+1}\lambda^{n-1}\left(\frac{3}{4}\frac{\eta}{2}\alpha\right)\right)=$$
$$\lambda-\hat{\lambda}-\lambda^{n-1}\hat{\lambda}\left(\lambda-\left(\frac{3}{4}+\frac{1}{4}\lambda\right)\right)-\lambda^{k(n-1)}\left(\left(1-\hat{\lambda}\lambda^{n-1}\right)\left(\lambda^{k+1}-\hat{\lambda}^{k+1}\right)+\hat{\lambda}^{k+1}\lambda^{n-1}\left(\frac{3}{4}\frac{\eta}{2}\alpha\right)\right)=^{**}$$
$$\lambda-\hat{\lambda}-\lambda^{n-1}\hat{\lambda}\left(\lambda-\left(\hat{\lambda}+\frac{3}{4}\frac{\eta}{2}\alpha\right)\right)-\lambda^{k(n-1)}\left(\left(1-\hat{\lambda}\lambda^{n-1}\right)\left(\lambda^{k+1}-\hat{\lambda}^{k+1}\right)+\hat{\lambda}^{k+1}\lambda^{n-1}\left(\frac{3}{4}\frac{\eta}{2}\alpha\right)\right)=$$
$$\left(\lambda-\hat{\lambda}\right)\left(1-\hat{\lambda}\lambda^{n-1}\right)+\lambda^{n-1}\hat{\lambda}\left(\frac{3}{4}\frac{\eta}{2}\alpha\left(1-\hat{\lambda}^{k}\lambda^{k(n-1)}\right)\right)+\lambda^{k(n-1)}\left(\left(1-\hat{\lambda}\lambda^{n-1}\right)\left(\hat{\lambda}^{k+1}-\lambda^{k+1}\right)\right)$$

where equality * follows from Claim D.8 and equality ** follows from Claim D.4. □

Claim D.10. $\forall k\in\mathbb{Z}_{>0}$:

$$\lambda\left(1-\lambda^{kn}\right)-\hat{\lambda}\frac{1-\left(\lambda^{n-1}\hat{\lambda}\right)^{k}}{1-\lambda^{n-1}\hat{\lambda}}\left(1-\lambda^{n}\left(\frac{3}{4}\lambda^{-1}+\frac{1}{4}\right)\right)=$$
$$\left(\lambda-\hat{\lambda}\right)+\frac{\lambda^{n-1}\left(\frac{3}{4}\frac{\eta}{2}\alpha\left(1-\hat{\lambda}^{k}\lambda^{k(n-1)}\right)\right)}{1-\hat{\lambda}\lambda^{n-1}}+\lambda^{k(n-1)}\left(\hat{\lambda}^{k+1}-\lambda^{k+1}\right).$$

Proof of Claim D.10.

$$\lambda\left(1-\lambda^{kn}\right)-\hat{\lambda}\frac{1-\left(\lambda^{n-1}\hat{\lambda}\right)^{k}}{1-\lambda^{n-1}\hat{\lambda}}\left(1-\lambda^{n}\left(\frac{3}{4}\lambda^{-1}+\frac{1}{4}\right)\right)=$$
$$\frac{\left(\lambda-\hat{\lambda}\right)\left(1-\hat{\lambda}\lambda^{n-1}\right)+\lambda^{n-1}\hat{\lambda}\left(\frac{3}{4}\frac{\eta}{2}\alpha\left(1-\hat{\lambda}^{k}\lambda^{k(n-1)}\right)\right)+\lambda^{k(n-1)}\left(\left(1-\hat{\lambda}\lambda^{n-1}\right)\left(\hat{\lambda}^{k+1}-\lambda^{k+1}\right)\right)}{\left(1-\hat{\lambda}\lambda^{n-1}\right)}=$$
$$\left(\lambda-\hat{\lambda}\right)+\frac{\lambda^{n-1}\left(\frac{3}{4}\frac{\eta}{2}\alpha\left(1-\hat{\lambda}^{k}\lambda^{k(n-1)}\right)\right)}{1-\hat{\lambda}\lambda^{n-1}}+\lambda^{k(n-1)}\left(\hat{\lambda}^{k+1}-\lambda^{k+1}\right)$$

where first equality is true from Claim D.9. □

Claim D.11. $\forall k\in\mathbb{Z}_{>0}:\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(\lambda-\hat{\lambda}+\frac{\lambda^{n-1}\left(\frac{3}{4}\frac{\eta}{2}\alpha\left(1-\hat{\lambda}^{k}\lambda^{k(n-1)}\right)\right)}{1-\hat{\lambda}\lambda^{n-1}}\right)+(\rho-\hat{\rho})>0.$

Proof of Claim D.11.

$$\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(\lambda-\hat{\lambda}+\frac{\lambda^{n-1}\left(\frac{3}{4}\frac{\eta}{2}\alpha\left(1-\hat{\lambda}^k\lambda^{k(n-1)}\right)\right)}{1-\hat{\lambda}\lambda^{n-1}}\right)+(\rho-\hat{\rho})=$$
$$\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(\lambda-\hat{\lambda}+\frac{\lambda^{n-1}\left(\frac{3}{4}\frac{\eta}{2}\alpha\left(1-\hat{\lambda}^k\lambda^{k(n-1)}\right)\right)}{1-\hat{\lambda}\lambda^{n-1}}\right)+\frac{\eta}{2}ncx_h^2\beta\left(1-\frac{1}{4}\right)=$$
$$\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}(1-\frac{\eta}{2}\left(\alpha+nx^2\beta\right)-\left(1-\frac{\eta}{2}\left(\alpha+\frac{1}{4}nx^2\beta\right)\right)+$$
$$\frac{\lambda^{n-1}\left(\frac{3}{4}\frac{\eta}{2}\alpha\left(1-\hat{\lambda}^k\lambda^{k(n-1)}\right)\right)}{1-\hat{\lambda}\lambda^{n-1}})+\frac{3}{4}\frac{\eta}{2}ncx_h^2\beta=$$
$$\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\left(-\frac{3}{4}\frac{\eta}{2}\left(nx^2\beta\right)+\frac{\lambda^{n-1}\left(\frac{3}{4}\frac{\eta}{2}\alpha\left(1-\hat{\lambda}^k\lambda^{k(n-1)}\right)\right)}{1-\hat{\lambda}\lambda^{n-1}}\right)+\frac{3}{4}\frac{\eta}{2}ncx_h^2\beta=$$
$$\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\frac{\lambda^{n-1}\left(\frac{3}{4}\frac{\eta}{2}\alpha\left(1-\hat{\lambda}^k\lambda^{k(n-1)}\right)\right)}{1-\hat{\lambda}\lambda^{n-1}}+ncx_h^2\beta\left(\frac{3}{4}\frac{\eta}{2}-\frac{3}{4}\frac{\eta}{2}\frac{nx^2\beta}{\alpha+nx^2\beta}\right)=$$
$$\frac{ncx_h^2\beta}{\alpha+nx_h^2\beta}\frac{\lambda^{n-1}\left(\frac{3}{4}\frac{\eta}{2}\alpha(1-\hat{\lambda}^k\lambda^{k(n-1)})\right)}{1-\hat{\lambda}\lambda^{n-1}}+ncx_h^2\beta\frac{3}{4}\frac{\eta}{2}\left(1-\frac{nx^2\beta}{\alpha+nx^2\beta}\right)>0$$

where the last inequality holds because $\lambda,\hat{\lambda}<1$ and $\alpha>0$. □

Claim D.12. $\frac{1}{\alpha}>\lambda^{-2}\eta\frac{1-\lambda^{-2(k+1)n}}{1-\lambda^{-2}}$ *is true for all* k *such that* $0\le k\le\frac{1}{2n}\log_\lambda\left(\frac{1}{1+\frac{1}{\alpha\eta}(1-\lambda^2)}\right)-1$

Proof of Claim D.12.

$$\frac{1}{\alpha}\ge\lambda^{-2}\eta\frac{1-\lambda^{-2kn}}{1-\lambda^{-2}}\iff$$
$$\lambda^2\frac{1}{\alpha}\frac{1}{\eta}\left(1-\lambda^{-2}\right)\le 1-\lambda^{-2kn}\iff$$
$$\lambda^2\frac{1}{\alpha}\frac{1}{\eta}\left(\lambda^{-2}-1\right)\ge\lambda^{-2kn}-1\iff$$
$$1+\lambda^2\frac{1}{\alpha}\frac{1}{\eta}\left(\lambda^{-2}-1\right)\ge\lambda^{-2kn}\iff$$
$$-k\ge\frac{1}{2n}\log_\lambda\left(1+\frac{1}{\alpha\eta}\left(1-\lambda^2\right)\right)\iff$$
$$k\le\frac{1}{2n}\log_\lambda\left(\frac{1}{1+\frac{1}{\alpha\eta}\left(1-\lambda^2\right)}\right)$$

□

Claim D.13. $\frac{1}{\alpha}\left(\hat{\lambda}\lambda^{(n-1)}\right)^{2k}>\eta\sum_{i=0}^{n-1}\lambda^{2i}\sum_{j=0}^{k-1}\left(\hat{\lambda}^2\lambda^{2(n-1)}\right)^j$ *is true for all* $k\in\mathbb{Z}_{>0}:k\le\frac{1}{2n}\log_\lambda\left(\frac{1}{1+\frac{1}{\alpha\eta}(1-\lambda^2)}\right)$.

Proof of Claim D.13. First, note that the inequality can also be written as

$$\frac{1}{\alpha}>\eta\sum_{i=0}^{n-1}\lambda^{2i}\sum_{j=0}^{k-1}\left(\hat{\lambda}\lambda^{(n-1)}\right)^{2(j-k)}.$$

Second, the right-hand term of the inequality could be upper bound as in eq. 44. Therefore, for the claim's inequality to hold, it is enough that $\frac{1}{\alpha}\ge\eta\lambda^{-2}\frac{1-\lambda^{-2nk}}{1-\lambda^{-2}}$, which is proved by Claim D.12 to be true for $0<k\le\frac{1}{2n}\log_\lambda\left(\frac{1}{1+\frac{1}{\alpha\eta}(1-\lambda^2)}\right)$.

$$\eta\sum_{i=0}^{n-1}\lambda^{2i}\sum_{j=0}^{k-1}\left(\hat{\lambda}\lambda^{(n-1)}\right)^{2(j-k)}=\eta\sum_{i=0}^{n-1}\lambda^{2i}\sum_{j=0}^{k-1}\frac{1}{\left(\hat{\lambda}\lambda^{(n-1)}\right)^{2(k-j)}}<_{k>j}$$
$$\eta\sum_{i=0}^{n-1}\lambda^{2i}\sum_{j=0}^{k-1}\frac{1}{\left(\lambda\lambda^{(n-1)}\right)^{2(k-j)}}=\eta\sum_{i=0}^{n-1}\lambda^{2i}\sum_{j=0}^{k-1}\frac{1}{\lambda^{2n(k-j)}}=$$
$$\eta\sum_{i=0}^{n-1}\sum_{j=0}^{k-1}\frac{1}{\lambda^{2(nk-nj-i)}}=_{r=nj+i}\eta\sum_{r=0}^{nk-1}\frac{1}{\lambda^{2(nk-r)}}=_{r'=nk-r,1<r'<nk}\eta\sum_{r'=1}^{nk}\frac{1}{\lambda^{2(r')}}=$$
$$\eta\sum_{i=1}^{nk}\lambda^{-2i}=\eta\frac{\lambda^{-2}-\lambda^{-2(nk+1)}}{1-\lambda^{-2}}=\eta\lambda^{-2}\frac{1-\lambda^{-2nk}}{1-\lambda^{-2}} \tag{44}$$

□

Claim D.14. $\frac{1}{\alpha}\left(\hat{\lambda}\lambda^{n-1}\right)^{2k}\geq\eta\left(\hat{\lambda}\lambda^{n-1}\right)^{2k}\sum_{i=0}^{n-1}\lambda^{2i}$ *is true for all* $k\in\mathbb{Z}_{>0}:k\leq\frac{1}{2n}\log_{\lambda}\left(\frac{1}{1+\frac{1}{\alpha\eta}(1-\lambda^2)}\right)$.

Proof of Claim D.14. Eq. 45 holds because $\lambda,\hat{\lambda}<1$. By multiplying both sides with $\sum_{i=0}^{n-1}\lambda^{2i}$, we get eq. 46. Then, noticing that the right term equals to the right term of Claim D.13 inequality, and hence smaller than the left term of Claim D.13 inequality, Claim D.14 is proved.

$$\left(\hat{\lambda}\lambda^{n-1}\right)^{2k}<1<\sum_{i=0}^{k-1}\left(\hat{\lambda}\lambda^{n-1}\right)^{2j} \tag{45}$$

$$\eta\left(\hat{\lambda}\lambda^{n-1}\right)^{2k}\sum_{i=0}^{n-1}\lambda^{2i}<\eta\sum_{j=0}^{k-1}\left(\hat{\lambda}\lambda^{n-1}\right)^{2j}\sum_{i=0}^{n-1}\lambda^{2i} \tag{46}$$

□

Claim D.15. *The inequality*

$$\left(\hat{\lambda}\lambda^{n-r}\right)^{2}\left(\frac{1}{\alpha}\left(\hat{\lambda}\lambda^{n-1}\right)^{2k}+\eta\sum_{j=0}^{k-1}\left(\hat{\lambda}\lambda^{n-1}\right)^{2j}\sum_{i=0}^{n-1}\lambda^{2i}\right)>\eta\sum_{i=0}^{n-r}\lambda^{2i}$$

holds for all $k\in\mathbb{Z}_{>0}$ *and* $x_h^2\beta>3,n>\frac{1}{2\alpha x_h^2\beta}-\frac{1}{x_h^2\beta}$.

Proof of Claim D.15. We start with eq. 47, where the first inequality holds because $\lambda<1$ and $r>1$, and the second inequality holds because $\lambda<\hat{\lambda}$. Using eq. 47, we can lower-bound the left-hand side of the claim's inequality. We continue with eq. 48, where the inequality holds because $\lambda<\hat{\lambda}$ and $r>1$. This allows us to upper-bound the right side of the claim's inequality. Given these lower and upper bounds, it's enough to show that $\lambda^{2n}(\frac{1}{\alpha}\lambda^{2kn}+\eta\frac{1-\lambda^{2kn}}{1-\lambda^2})>\eta\frac{1-\lambda^{2n}}{1-\lambda^2}$, which according to eq. 49 is equivalent to showing that $(2nx_h^2\beta-1)\frac{1}{\alpha}\lambda^{2(k+1)n}+2(2\lambda^{2n}-1)>0$. Since $n>\frac{1}{2\alpha x_h^2\beta}-\frac{1}{x_h^2\beta}$, Claim D.18 applies, and therefore $\lambda^{2n}\geq e^{-\frac{2}{x_h^2\beta}}$. Consequently, it's enough to show that $(2nx_h^2\beta-1)\frac{1}{\alpha}\lambda^{2(k+1)n}+2(2e^{-\frac{2}{x_h^2\beta}}-1)>0$, which is true for $x_h^2\beta>3$ by Claim D.16.

$$\left(\hat{\lambda}\lambda^{n-r}\right)^{2}\left(\frac{1}{\alpha}\left(\hat{\lambda}\lambda^{n-1}\right)^{2k}+\eta\sum_{j=0}^{k-1}\left(\hat{\lambda}\lambda^{n-1}\right)^{2j}\sum_{i=0}^{n-1}\lambda^{2i}\right)>$$
$$\left(\hat{\lambda}\lambda^{n-1}\right)^{2}\left(\frac{1}{\alpha}\left(\hat{\lambda}\lambda^{n-1}\right)^{2k}+\eta\sum_{j=0}^{k-1}\left(\hat{\lambda}\lambda^{n-1}\right)^{2j}\sum_{i=0}^{n-1}\lambda^{2i}\right)> \tag{47}$$
$$\lambda^{2n}\left(\frac{1}{\alpha}\lambda^{2kn}+\eta\sum_{j=0}^{k-1}\lambda^{2jn}\sum_{i=0}^{n-1}\lambda^{2i}\right)=\lambda^{2n}\left(\frac{1}{\alpha}\lambda^{2kn}+\eta\frac{1-\lambda^{2kn}}{1-\lambda^2}\right)$$

$$\eta\sum_{i=0}^{n-r}\lambda^{2i}<\eta\sum_{i=0}^{n-1}\lambda^{2i}=\eta\frac{1-\lambda^{2n}}{1-\lambda^2} \tag{48}$$

$$
\begin{aligned}
&\lambda^{2n}\left(\frac{1}{\alpha}\lambda^{2kn}+\eta\frac{1-\lambda^{2kn}}{1-\lambda^2}\right)>\eta\frac{1-\lambda^{2n}}{1-\lambda^2}\\
&\lambda^{2n}\left(1-\lambda^2\right)\frac{1}{\alpha}\lambda^{2kn}+\eta\lambda^{2n}\left(1-\lambda^{2kn}\right)>\eta\left(1-\lambda^{2n}\right)\\
&\left(1-\lambda^2\right)\frac{1}{\alpha}\lambda^{2(k+1)n}+\eta\left(2\lambda^{2n}-\lambda^{2(k+1)n}-1\right)>0\\
&\left(\alpha+nx_h^2\beta\right)^2\left(1-\lambda^2\right)\frac{1}{\alpha}\lambda^{2(k+1)n}+2\left(2\lambda^{2n}-\lambda^{2(k+1)n}-1\right)>0\\
&\left(\alpha+nx_h^2\beta\right)^2\left(1-\left(1-\frac{1}{\alpha+nx_h^2\beta}\right)^2\right)\frac{1}{\alpha}\lambda^{2(k+1)n}+2\left(2\lambda^{2n}-\lambda^{2(k+1)n}-1\right)>0\\
&\left(2\left(\alpha+nx_h^2\beta\right)-1\right)\frac{1}{\alpha}\lambda^{2(k+1)n}+2\left(2\lambda^{2n}-\lambda^{2(k+1)n}-1\right)>0\\
&2\lambda^{2(k+1)n}+\left(2nx^2\beta-1\right)\frac{1}{\alpha}\lambda^{2(k+1)n}+2\left(2\lambda^{2n}-\lambda^{2(k+1)n}-1\right)>0\\
&\left(2nx_h^2\beta-1\right)\frac{1}{\alpha}\lambda^{2(k+1)n}+2\left(2\lambda^{2n}-1\right)>0
\end{aligned}
\tag{49}
$$

□

Claim D.16. *For $x^2\beta>3$, the inequality $(2e^{-\frac{2}{x^2\beta}}-1)>0$ holds.*

Proof of Claim D.16. It's easy to see that the inequality holds only if $x^2\beta\geq\frac{-2}{\ln\frac{1}{2}}$. Since $\frac{-2}{\ln\frac{1}{2}}<3$, the claim is proved. □

Claim D.17. *For $\dot{k}$ as defined in Lemma A.6, and the conditions of Claim D.18:*

$$
\frac{1}{\alpha}\left(e^{\frac{2}{x_h^2\beta}}+\alpha\frac{\left(e^{\frac{2}{x_h^2\beta}}-1\right)}{(\alpha+nx_h^2\beta)+\frac{1}{8}}\right)>\lambda^{-2}\eta\frac{1-\lambda^{-2\left(\lceil\dot{k}\rceil+1\right)n}}{1-\lambda^{-2}}.
$$

Proof of Claim D.17.

$$
\begin{aligned}
&\eta\frac{1-\lambda^{-2\left(\lceil\dot{k}\rceil+1\right)n}}{\lambda^2-1}\leq\eta\frac{\lambda^{-2\left(\dot{k}+2\right)n}-1}{1-\lambda^2}=\eta\frac{\lambda^{-2\left(\frac{1}{2n}\log_\lambda\left(\frac{1}{1+\frac{1}{\alpha\eta}\left(1-\lambda^2\right)}\right)-1+2\right)n}-1}{1-\lambda^2}=\\
&\eta\frac{\lambda^{-\log_\lambda\left(\frac{1}{1+\frac{1}{\alpha\eta}\left(1-\lambda^2\right)}\right)}\lambda^{-2n}-1}{1-\lambda^2}=\eta\frac{\left(1+\frac{1}{\alpha\eta}\left(1-\lambda^2\right)\right)\lambda^{-2n}-1}{1-\lambda^2}=\\
&\eta\frac{\left(1-\lambda^2\right)\lambda^{-2n}\frac{1}{\alpha\eta}}{1-\lambda^2}+\eta\frac{\lambda^{-2n}-1}{1-\lambda^2}=\frac{1}{\alpha}\lambda^{-2n}+\frac{\left(\lambda^{-2n}-1\right)}{(\alpha+nx_h^2\beta)+\frac{1}{8}}\leq\\
&e^{\frac{2}{x_h^2\beta}}\frac{1}{\alpha}+\frac{1}{\alpha}\alpha\frac{\left(e^{\frac{2}{x_h^2\beta}}-1\right)}{(\alpha+nx_h^2\beta)+\frac{1}{8}}=\frac{1}{\alpha}\left(e^{\frac{2}{x_h^2\beta}}+\alpha\frac{\left(e^{\frac{2}{x_h^2\beta}}-1\right)}{(\alpha+nx_h^2\beta)+\frac{1}{8}}\right)
\end{aligned}
$$

where the fourth equality holds from eq. 50 and the second inequality holds from D.18.

$$
\begin{aligned}
&\frac{\eta}{\lambda^2-1}=\eta\frac{1}{\left(1-\frac{\eta}{2}\left(\alpha+nx_h^2\beta\right)\right)^2-1}=\eta\frac{1}{\eta\left(\alpha+nx_h^2\beta\right)+\left(\frac{\eta}{2}\left(\alpha+nx_h^2\beta\right)\right)^2}=\\
&\frac{1}{(\alpha+nx_h^2\beta)+\frac{\eta}{4}\left(\alpha+nx_h^2\beta\right)^2}=\frac{1}{(\alpha+nx_h^2\beta)+\frac{1}{8}}
\end{aligned}
\tag{50}
$$

□

Claim D.18. *For the conditions of claim D.20:*

$$
(1-\frac{1}{\alpha+nx^2\beta})^{2n}\geq e^{-\frac{2}{x^2\beta}}.
$$

Proof of Claim D.18. The proof is easily deduced from Claims D.19 and D.20 □

Claim D.19.

$$\lim_{n\to\infty}(1-\frac{1}{\alpha+nx^2\beta})^{2n}=e^{-\frac{2}{x^2\beta}}.$$

Proof of Claim D.19. From eq. 51, it is enough to find $\lim_{n\to\infty}\frac{\ln(1-\frac{1}{\alpha+nx^2\beta})}{\frac{1}{2n}}$. Since $\lim_{n\to\infty}\frac{\ln(1-\frac{1}{\alpha+nx^2\beta})}{\frac{1}{2n}}=\frac{0}{0}$, and both the numerator and denominator are differentiable around ∞, the use of L'Hôpital's rule is possible as shown in eq. 52, with the result proving the claim.

$$\left(1-\frac{1}{\alpha+nx^2\beta}\right)^{2n}=e^{\ln\left(1-\frac{1}{\alpha+nx^2\beta}\right)^{2n}}=e^{2n\ln\left(1-\frac{1}{\alpha+nx^2\beta}\right)}=e^{\frac{\ln\left(1-\frac{1}{\alpha+nx^2\beta}\right)}{\frac{1}{2n}}} \tag{51}$$

$$\lim_{n\to\infty}\frac{\frac{d}{dn}\ln\left(1-\frac{1}{\alpha+nx^2\beta}\right)}{\frac{d}{dn}\frac{1}{2n}}=\lim\frac{\frac{x^2\beta}{(\alpha+nx^2\beta-1)(\alpha+nx^2\beta)}}{-\frac{1}{2n^2}}=-\lim\frac{2n^2x^2\beta}{(nx^2\beta)^2}=-\frac{2}{x^2\beta} \tag{52}$$

□

Claim D.20.

$$\forall n>\frac{1}{2\alpha x^2\beta}-\frac{1}{x^2\beta}:\ \frac{d}{dn}(1-\frac{1}{\alpha+nx^2\beta})^{2n}<0.$$

Proof of Claim D.20. First, we find a simplified term for the derivative:

$$\begin{aligned}&\frac{d}{dn}\left(1-\frac{1}{\alpha+nx^2\beta}\right)^{2n}=\frac{d}{dn}e^{2n\ln\left(1-\frac{1}{\alpha+nx^2\beta}\right)}=\\&\left(1-\frac{1}{\alpha+nx^2\beta}\right)^{2n}\left(2\ln\left(1-\frac{1}{\alpha+nx^2\beta}\right)+2n\frac{1}{1-\frac{1}{\alpha+nx^2\beta}}\cdot\frac{x^2\beta}{(\alpha+nx^2\beta)^2}\right)=\\&\left(1-\frac{1}{\alpha+nx^2\beta}\right)^{2n}\left(2\ln\left(1-\frac{1}{\alpha+nx^2\beta}\right)+\frac{2nx^2\beta}{(\alpha+nx^2\beta-1)(\alpha+nx^2\beta)}\right).\end{aligned} \tag{53}$$

A lower bound for the ln term can be found using Taylor's theorem as shown in eq. 54, where $0\le\xi\le\frac{1}{\alpha+nx^2\beta}$.

$$\ln\left(1-\frac{1}{\alpha+nx^2\beta}\right)=-\frac{1}{\alpha+nx^2\beta}-\frac{1}{2}\frac{1}{(1-\xi)^2}\left(\frac{1}{\alpha+nx^2\beta}\right)^2\le-\frac{1}{\alpha+nx^2\beta}-\frac{1}{2}\left(\frac{1}{\alpha+nx^2\beta}\right)^2 \tag{54}$$

From eq. 53 and 54, it is enough to find the conditions for which $\frac{nx^2\beta}{(\alpha+nx^2\beta-1)(\alpha+nx^2\beta)}<\frac{1}{\alpha+nx^2\beta}+\frac{1}{2}\frac{1}{(\alpha+nx^2\beta)^2}$. A simplified version of this inequality is found at eq. 55, and it can be easily seen that for $\alpha>\frac{1}{2}(\frac{1}{nx^2\beta}+1)$, and therefore also for $n>\frac{1}{2\alpha x^2\beta}-\frac{1}{x^2\beta}$, this inequality holds.

$$\begin{aligned}&\frac{nx^2\beta}{(\alpha+nx^2\beta-1)(\alpha+nx^2\beta)}<\frac{1}{\alpha+nx^2\beta}+\frac{1}{2}\frac{1}{(\alpha+nx^2\beta)^2}\\&0<2\alpha^2+2nx^2\beta\alpha-2\alpha-2nx^2\beta+\alpha+nx^2\beta-1\\&0<nx^2\beta(2\alpha-1)+\alpha(2\alpha-1)-1\end{aligned} \tag{55}$$

□

Claim D.21. *For $n>\frac{\alpha}{x_h^2\beta}(e^{\frac{2}{x_h^2\beta}}-2)+\frac{1}{2x_h^2\beta}$ and the conditions of Claim D.18, $\dot{k}$ defined in Lemma A.6 is positive.*

Proof of Claim D.21. The claim's inequality is simplified at eq. 56:

$$
\begin{aligned}
&\dot{k} > 0 \\
&\frac{1}{2n}\log_\lambda\left(\frac{1}{1+\frac{1}{\alpha\eta}\left(1-\lambda^2\right)}\right) - 1 > 0 \iff \\
&\log_\lambda\left(\frac{1}{1+\frac{1}{\alpha\eta}\left(1-\lambda^2\right)}\right) > 2n \iff \\
&\frac{\ln\left(\frac{1}{1+\frac{1}{\alpha\eta}\left(1-\lambda^2\right)}\right)}{\ln\lambda} > 2n \iff \\
&\ln\left(\frac{1}{1+\frac{1}{\alpha\eta}\left(1-\lambda^2\right)}\right) < 2n\ln\lambda \iff \\
&\ln\left(\frac{1}{1+\frac{1}{\alpha\eta}\left(1-\lambda^2\right)}\right) < \ln\lambda^{2n} \iff \\
&\frac{1}{1+\frac{1}{\alpha\eta}\left(1-\lambda^2\right)} < \lambda^{2n} \iff \\
&\lambda^{-2n} < 1 + \frac{1}{\alpha\eta}\left(1-\lambda^2\right) \iff \\
&\lambda^{-2n} - 1 < \frac{1}{\alpha\eta}\left(1-\lambda^2\right).
\end{aligned}
\tag{56}
$$

By Claim D.18, $\lambda^{-2n} - 1 < e^{\frac{2}{x_h^2\beta}} - 1$; therefore it is enough to find conditions for $e^{\frac{2}{x_h^2\beta}} - 1 < \frac{1}{\alpha\eta}(1-\lambda^2)$, which is done at eq. 57. As this condition matches the claim conditions, the claim is proved.

$$
\begin{aligned}
&e^{\frac{2}{x_h^2\beta}} - 1 < \frac{1}{\alpha\eta}\left(1-\lambda^2\right) \iff \\
&\alpha\eta\left(e^{\frac{2}{x_h^2\beta}} - 1\right) < \left(1-\lambda^2\right) \iff \\
&\alpha\eta\left(e^{\frac{2}{x_h^2\beta}} - 1\right) < 1 - \left(1 - \frac{\eta}{2}\left(\alpha + nx_h^2\beta\right)\right)^2 \iff \\
&\alpha\left(e^{\frac{2}{x_h^2\beta}} - 1\right) < \left(\alpha + nx_h^2\beta\right) - \frac{\eta}{4}\left(\alpha + nx^2\beta\right)^2 \iff \\
&\alpha\left(e^{\frac{2}{x_h^2\beta}} - 1\right) < \left(\alpha + nx_h^2\beta\right) - \frac{1}{2} \iff \\
&\alpha\left(e^{\frac{2}{x_h^2\beta}} - 2\right) + \frac{1}{2} < nx_h^2\beta \iff \\
&\frac{\alpha}{x_h^2\beta}\left(e^{\frac{2}{x_h^2\beta}} - 2\right) + \frac{1}{2x_h^2\beta} < n
\end{aligned}
\tag{57}
$$

□

B. Posterior Sampling Privacy

This subsection provides auxiliary claims for the posterior sampling privacy analysis performed in subsection IV-A. It uses the notations defined in section IV, subsection IV-A, and subection A-B, specifically: $\alpha, \beta, \theta, p(y|x,\theta)$ (defined in eq. 4), $\mathcal{D}(n, \gamma_1, x_h, x_l, c)$ (defined in eq. 5), $p(\theta|D), p(\theta|\hat{D}), \theta, \mu, \sigma, \hat{\mu}, \hat{\sigma}, (\sigma^2)^*_\nu, \nu$ (defined in subsection IV-A), $D, \hat{D}, x_n, y_n, \hat{x}_n, \hat{y}_n, z, q$ (defined in eq. 18).

Claim D.22. *For $n > 1 + 10\frac{x_h^2}{x_l^2}\frac{\nu}{\beta}$, the inequality $\frac{1}{10}(\alpha + (z + x_n^2)\beta) > \nu(\hat{x}_n^2 - x_n^2)$ holds.*

Proof Claim D.22. Notice that $\frac{1}{10}(\alpha + (z + x_n^2)\beta) > \frac{1}{10}z\beta > \frac{1}{10}(n-1)x_l^2\beta$ and $\nu x_h^2 > \nu(\hat{x}_n^2 - x_n^2)$. Therefore a sufficient condition will be that $\frac{1}{10}(n-1)x_l^2\beta > \nu x_h^2$, which is equivalent to $n > 1 + \frac{x_h^2}{x_l^2}\frac{10\nu}{\beta}$. □

Claim D.23. *$(\sigma^2)^*_\nu$ is positive.*

Proof Claim D.23.

$$\left(\sigma^2\right)^*_\nu = \nu\sigma^2 + (1-\nu)\,\hat{\sigma}^2 = \frac{\nu}{\alpha + (z + x_n^2)\,\beta} + \frac{1-\nu}{\alpha + (z + \hat{x}_n^2)\,\beta} =$$
$$\frac{\nu\left(\alpha + \left(z + \hat{x}_n^2\right)\beta\right) + (1-\nu)\left(\alpha + \left(z + x_n^2\right)\beta\right)}{(\alpha + (z + x_n^2)\,\beta)\,(\alpha + (z + \hat{x}_n^2)\,\beta)} = \frac{\alpha + \left(z + x_n^2\right)\beta + \nu\left(x_n^2 - \hat{x}_n^2\right)}{(\alpha + (z + x_n^2)\,\beta)\,(\alpha + (z + \hat{x}_n^2)\,\beta)} \tag{58}$$

therefore, a sufficient condition is that $\alpha + (z + x_n^2)\beta + \nu(x_n^2 - \hat{x}_n^2) > 0$. Since the condition of Lemma A.1 dictates that $n > 1 + 10\frac{x_h^2}{x_l^2}\frac{\nu}{\beta}$, then Claim D.22 holds, and therefore the condition is satisfied. □

Claim D.24. *The value $\ln\frac{\sigma}{\hat{\sigma}}$ can be bounded as following:*

$$\ln\frac{\sigma}{\hat{\sigma}} \leq \frac{x_h^2}{2(n-1)x_l^2}.$$

Proof of Claim D.24. For $\hat{x}_n \leq x_n$, the term $\ln\frac{\sigma}{\hat{\sigma}}$ is negative and the claim trivially holds. For $\hat{x}_n > x_n$, consider $c_1 = \frac{x_h^2}{(n-1)x_l^2}$:

$$c_1 = \frac{x_h^2}{(n-1)x_l^2} > \frac{\hat{x}_n^2 - x_n^2}{z + x_n^2} > \frac{\hat{x}_n^2\beta - x_n^2\beta}{\alpha + (z + x_n^2)\beta} = \frac{\alpha + (z + \hat{x}_n^2)\beta}{\alpha + (z + x_n^2)\beta} - 1. \tag{59}$$

From eq. 59, by Taylor theorem:

$$e^{c_1} = 1 + c_1 + \frac{e^\zeta}{2}(c_1)^2 > 1 + c_1 > \frac{\alpha + (z + \hat{x}_n^2)\beta}{\alpha + (z + x_n^2)\beta}$$

where $0 \leq \zeta \leq c_1$. Consequently, because the natural logarithm is monotonically increasing, the following equation also holds:

$$\frac{1}{2}c_1 > \frac{1}{2}\ln\frac{\alpha + (z + \hat{x}_n)\beta}{\alpha + (z + x_n)\beta} = \ln\frac{\sigma}{\hat{\sigma}}.$$

Therefore $\ln\frac{\sigma}{\hat{\sigma}} < \frac{1}{2}\frac{x_h^2}{(n-1)x_l^2}$. □

Claim D.25. *For the conditions of Lemma A.1, the value of $\frac{1}{2}(\nu-1)\ln\frac{\hat{\sigma}^2}{(\sigma^2)^*_\nu}$ can be upper bounded as following:*

$$\frac{1}{2}(\nu-1)\ln\frac{\hat{\sigma}^2}{(\sigma^2)^*_\nu} \leq \frac{1}{2}(\nu-1)\frac{\nu x_h^2}{2\left((n-1)\,x_l^2 - \nu x_h^2\right)}.$$

Proof of Claim D.25. Consider $c_1 = \frac{\nu x_h^2}{((n-1)x_l^2 - \nu x_h^2)}$:

$$c_1 = \frac{\nu x_h^2}{(n-1)\,x_l^2 - \nu x_h^2} \geq^* \frac{\nu\beta x_h^2}{\alpha + (n-1)\,x_l^2\beta - \nu\beta x_h^2} \geq^* \frac{\nu\beta\hat{x}_n^2}{\alpha + (z + x_n^2)\,\beta - \nu\beta x_n^2} \geq \frac{\nu\beta\left(\hat{x}_n^2 - x_n^2\right)}{\alpha + (z + x_n^2)\,\beta - \nu\beta\,(x_n^2 - \hat{x}_n^2)} =$$
$$\frac{\alpha + \left(z + x_n^2\right)\beta}{\alpha + (z + x_n^2)\,\beta + \nu\beta\,(x_n^2 - \hat{x}_n^2)} - 1 = \frac{1}{\alpha + (z + \hat{x}_n^2)\,\beta}\cdot\frac{\left(\alpha + \left(z + x_n^2\right)\beta\right)\left(\alpha + \left(z + \hat{x}_n^2\right)\beta\right)}{\alpha + (z + x_n^2)\,\beta + \nu\beta\,(x_n^2 - \hat{x}_n^2)} - 1 = \frac{\hat{\sigma}^2}{(\sigma^2)^*_\nu} - 1$$

where inequalities * holds under the assumption that $n > 1 + \nu\frac{x_h^2}{x_l^2}$, and last equality holds from eq. 58. Therefore, by using Taylor theorem:

$$e^{c_1} = 1 + c_1 + \frac{e^\zeta}{2}(c_1)^2 > 1 + c_1 \geq \frac{\hat{\sigma}^2}{(\sigma^2)^*_\nu}.$$

where $0 \leq \zeta \leq c_1$. From this inequality, and because the natural logarithm is monotonically increasing, it is true that $\ln\frac{\hat{\sigma}^2}{(\sigma^2)^*_\nu} \leq c_1$. Therefore

$$\frac{1}{2}(\nu-1)\ln\frac{\hat{\sigma}^2}{(\sigma^2)^*_\nu} \leq \frac{1}{2}(\nu-1)c_1 = \frac{1}{2}(\nu-1)\frac{\nu x_h^2}{((n-1)x_l^2 - \nu x_h^2)}.$$

□

Claim D.26. *For the conditions of Lemma A.1, the value $\frac{\nu}{2}\frac{(\mu-\hat{\mu})^2}{(\sigma^2)^*_\nu}$ is bounded by*

$$2\nu\beta\cdot\frac{x_h^4}{\frac{9}{10}n^{1-2\gamma_1}x_l^2} + 2\nu\cdot\frac{x_h^4\left(\alpha + x_h^2\beta\right)}{\frac{9}{10}x_l^4}\cdot\frac{(c + n^{\gamma_1})}{n^{2-\gamma_1}} + \frac{\nu}{2}\cdot\frac{x_h^4\left(\alpha + x_h^2\beta\right)^2}{\frac{9}{10}x_l^6\beta}\cdot\frac{(c + n^{\gamma_1})^2}{n^3}.$$

Proof of Claim D.26. First, we bound $|\mu - \hat{\mu}|$:

$$
\begin{aligned}
&|\mu - \hat{\mu}| = \beta|\frac{q + x_n y_n}{\alpha + (z + x_n^2)\beta} - \frac{q + \hat{x}_n \hat{y}_n}{\alpha + (z + \hat{x}_n^2)\beta}| = \\
&|\frac{(q + x_n y_n)\left(\alpha + \left(z + \hat{x}_n^2\right)\beta\right) - (q + \hat{x}_n \hat{y}_n)\left(\alpha + \left(z + x_n^2\right)\beta\right)}{(\alpha + (z + x_n^2)\beta)(\alpha + (z + \hat{x}_n^2)\beta)}| = \\
&\beta|\frac{q\hat{x}_n^2\beta + x_n y_n \alpha + x_n y_n z\beta + x_n y_n \hat{x}_n^2 \beta - q x_n^2 \beta - \hat{x}_n \hat{y}_n \alpha - \hat{x}_n \hat{y}_n z\beta - \hat{x}_n \hat{y}_n x_n^2 \beta}{(\alpha + (z + x_n^2)\beta)(\alpha + (z + \hat{x}_n^2)\beta)}| = \\
&\beta|\frac{\hat{x}_n^2 z\left(\frac{q}{z} - \frac{\hat{y}_n}{\hat{x}_n}\right)\beta - x_n^2 z\left(\frac{q}{z} - \frac{y_n}{x_n}\right)\beta + \alpha(x_n y_n - \hat{x}_n \hat{y}_n) + x_n \hat{x}_n \beta(y_n \hat{x}_n - \hat{y}_n x_n)}{(\alpha + (z + x_n^2)\beta)(\alpha + (z + \hat{x}_n^2)\beta)}| < \\
&\beta|\frac{\hat{x}_h^2 z(2n^{\gamma_1})\beta + \alpha x_h^2(c + n^{\gamma_1}) + x_h^4 \beta(c + n^{\gamma_1})}{(\alpha + (z + x_n^2)\beta)(\alpha + (z + \hat{x}_n^2)\beta)}| = \beta|\frac{2\hat{x}_h^2 \beta z n^{\gamma_1} + \left(x_h^2 \alpha + x_h^4 \beta\right)(c + n^{\gamma_1})}{(\alpha + (z + x_n^2)\beta)(\alpha + (z + \hat{x}_n^2)\beta)}|.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\frac{\nu}{2}\frac{(\mu - \hat{\mu})^2}{(\sigma^2)_\nu^*} \leq \frac{\nu}{2}\beta^2\left(\frac{2\hat{x}_h^2 \beta z n^{\gamma_1} + \left(x_h^2 \alpha + x_h^4 \beta\right)(c + n^{\gamma_1})}{(\alpha + (z + x_n^2)\beta)(\alpha + (z + \hat{x}_n^2)\beta)}\right)^2 \cdot \left(\frac{\alpha + \left(z + x_n^2\right)\beta + \nu\left(x_n^2 - \hat{x}_n^2\right)}{(\alpha + (z + x_n^2)\beta)(\alpha + (z + \hat{x}_n^2)\beta)}\right)^{-1} = \\
&\frac{\nu}{2}\frac{\beta^2\left(2\hat{x}_h^2 \beta z n^{\gamma_1} + \left(x_h^2 \alpha + x_h^4 \beta\right)(c + n^{\gamma_1})\right)^2}{(\alpha + (z + x_n^2)\beta)(\alpha + (z + \hat{x}_n^2)\beta)(\alpha + (z + x_n^2)\beta + \nu(x_n^2 - \hat{x}_n^2))} \leq^* \\
&\frac{\nu}{2}\frac{\beta^2\left(2x_h^2 \beta z n^{\gamma_1} + \left(x_h^2 \alpha + x_h^4 \beta\right)(c + n^{\gamma_1})\right)^2}{\frac{9}{10}(\alpha + (z + x_n^2)\beta)(\alpha + (z + \hat{x}_n^2)\beta)(\alpha + (z + x_n^2)\beta)} = \\
&\frac{\nu}{2}\beta^2\left(\frac{\left(2x_h^2 \beta\right)^2 z^2 n^{2\gamma_1} + 2\left(2x_h^2 \beta\right)\left(x_h^2 \alpha + x_h^4 \beta\right) z n^{\gamma_1}(c + n^{\gamma_1}) + \left(x_h^2 \alpha + x_h^4 \beta\right)^2 (c + n^{\gamma_1})^2}{\frac{9}{10}(\alpha + (z + x_n^2)\beta)^2(\alpha + (z + \hat{x}_n^2)\beta)}\right) \leq \\
&\frac{\nu}{2}\beta^2\left(\frac{\left(2x_h^2 \beta\right)^2 z^2 n^{2\gamma_1} + \left(4x_h^2 \beta\right)\left(x_h^2 \alpha + x_h^4 \beta\right) z n^{\gamma_1}(c + n^{\gamma_1}) + \left(x_h^2 \alpha + x_h^4 \beta\right)^2 (c + n^{\gamma_1})^2}{\frac{9}{10}((z + x_n^2)\beta)^2((z + \hat{x}_n^2)\beta)}\right) \leq^{**} \\
&\frac{\nu}{2}\beta^2\left(\frac{\left(2x_h^2 \beta\right)^2 n^{2\gamma_1}}{\frac{9}{10} n x_l^2 \beta^3}\right) + \frac{\nu}{2}\beta^2\left(\frac{\left(4x_h^2 \beta\right)\left(x_h^2 \alpha + x_h^4 \beta\right) n^{\gamma_1}(c + n^{\gamma_1})}{\frac{9}{10}(n x_l^2)^2 \beta^3}\right) + \frac{\nu}{2}\beta^2\left(\frac{\left(x_h^2 \alpha + x_h^4 \beta\right)^2 (c + n^{\gamma_1})^2}{\frac{9}{10}(n x_l^2 \beta)^3}\right) = \\
&2\nu\beta\left(\frac{x_h^4}{\frac{9}{10} n^{1-2\gamma_1} x_l^2}\right) + 2\nu\beta\left(\frac{\left(x_h^2 \beta\right)\left(x_h^2 \alpha + x_h^4 \beta\right)}{\frac{9}{10}(x_l^2 \beta)^2}\right)\frac{(c + n^{\gamma_1})}{n^{2-\gamma_1}} + \frac{\nu}{2}\left(\frac{\left(x_h^2 \alpha + x_h^4 \beta\right)^2}{\frac{9}{10} x_l^6 \beta}\right)\frac{(c + n^{\gamma_1})^2}{n^3}
\end{aligned}
$$

where inequality * is true because the conditions of Lemma A.1 dictates that $n > 1 + \frac{x_h^2}{x_l^2}\frac{10\nu}{\beta}$, and according to Claim D.22 this guarantee that $\frac{1}{10}(\alpha + (z + x_n^2)\beta) > \nu(\hat{x}_n^2 - x_n^2)$. Inequality ** follows from $n >> 1 \Rightarrow (n-1)x_l \approx n x_l$. □

Claim D.27. *For the conditions and definitions of Lemma IV.4, one sample from the posterior is (ϵ, δ) differentially private for the following conditions on n and ν:*

$$
\begin{aligned}
&\nu = 1 + \frac{2\ln(\frac{1}{\delta})}{\epsilon} \\
&n \geq \max\{1 + \frac{x_h^2}{x_l^2}\frac{8}{\epsilon}, 1 + \nu\frac{x_h^2}{x_l^2}\left(1 + 8\frac{(\nu - 1)}{\epsilon}\right), \left(\frac{16\nu\beta x_h^4}{\frac{9}{10}\epsilon x_l^2}\right)^{\frac{1}{1-2\gamma_1}}, \\
&\left(\frac{16\nu}{\epsilon} \cdot \frac{x_h^4\left(\alpha + x_h^2\beta\right)}{\frac{9}{10}x_l^4}(c + n^{\gamma_1})\right)^{\frac{1}{2-\gamma_1}}, \left(\frac{4\nu}{\epsilon} \cdot \frac{x_h^4\left(\alpha + x_h^2\beta\right)^2}{\frac{9}{10}x_l^6\beta}\right)^{\frac{1}{3}}(c + n^{\gamma_1})^{\frac{2}{3}}\}.
\end{aligned}
$$

Proof of Claim D.27. By Lemma A.1, one sample from the posterior is $(\epsilon_1 + \frac{\ln(\frac{1}{\delta})}{\nu - 1}, \delta)$ differentially private. For each of the six terms of $\epsilon_1 + \frac{\ln(\frac{1}{\delta})}{\nu - 1}$, the lower bounds on n and ν, found at equations 60, 61, 62, 63, 64, and 65, guarantee that the sum of terms is upper bounded by ϵ. These bounds match the claim's guarantee over n and ν, thus proving the claim.

For $\frac{\ln(\frac{1}{\delta})}{\nu-1}$:

$$\begin{aligned}
&\frac{\ln(\frac{1}{\delta})}{\nu-1} = \frac{\epsilon}{2} \\
&\frac{2\ln(\frac{1}{\delta})}{\epsilon} + 1 = \nu.
\end{aligned} \tag{60}$$

For $\frac{x_h^2}{(n-1)x_l^2}$:

$$\begin{aligned}
&\frac{x_h^2}{2(n-1)x_l^2} \leq \frac{\epsilon}{16} \\
&n \geq 1 + \frac{x_h^2}{x_l^2}\frac{8}{\epsilon}.
\end{aligned} \tag{61}$$

For $\frac{(\nu-1)\nu x_h^2}{2((n-1)x_l^2-\nu x_h^2)}$:

$$\begin{aligned}
&\frac{(\nu-1)\nu x_h^2}{2((n-1)x_l^2-\nu x_h^2)} \leq \frac{\epsilon}{16} \\
&\frac{1}{2}(\nu-1)\frac{\nu x_h^2}{(n-1)\,x_l^2-\nu x_h^2} \leq \frac{\epsilon}{16} \\
&\frac{1}{2}\,(\nu-1)\,\frac{16\nu x_h^2}{\epsilon} \leq (n-1)\,x_l^2-\nu x_h^2 \\
&n \geq 1 + \frac{1}{2}\,(\nu-1)\,\frac{16\nu x_h^2}{\epsilon x_l^2} + \nu\frac{x_h^2}{x_l^2} = 1 + \nu\frac{x_h^2}{x_l^2}\left(1 + 8\frac{(\nu-1)}{\epsilon}\right).
\end{aligned} \tag{62}$$

For $\frac{20\nu\beta x_h^4}{9n^{1-2\gamma_1}x_l^2}$:

$$\begin{aligned}
&\frac{20\nu\beta x_h^4}{9n^{1-2\gamma_1}x_l^2} \leq \frac{\epsilon}{8} \\
&\frac{16}{\epsilon}\frac{\nu\beta x_h^4}{\frac{9}{10}x_l^2} \leq n^{1-2\gamma_1} \\
&n \geq \left(\frac{16\nu\beta x_h^4}{\frac{9}{10}\epsilon x_l^2}\right)^{\frac{1}{1-2\gamma_1}}.
\end{aligned} \tag{63}$$

For $\frac{20\nu x_h^4(\alpha+x_h^2\beta)}{9x_l^4}\cdot\frac{(c+n^{\gamma_1})}{n^{2-\gamma_1}}$:

$$\begin{aligned}
&\frac{20\nu x_h^4(\alpha+x_h^2\beta)}{9x_l^4}\cdot\frac{(c+n^{\gamma_1})}{n^{2-\gamma_1}} \leq \frac{\epsilon}{8} \\
&n^{2-\gamma_1} \geq \frac{16\nu}{\epsilon}\cdot\frac{x_h^4(\alpha+x_h^2\beta)}{\frac{9}{10}x_l^4}\cdot(c+n^{\gamma_1}) \\
&n \geq \left(\frac{16\nu}{\epsilon}\cdot\frac{x_h^4(\alpha+x_h^2\beta)}{\frac{9}{10}x_l^4}\cdot(c+n^{\gamma_1})\right)^{\frac{1}{2-\gamma_1}}.
\end{aligned} \tag{64}$$

For term $\frac{5\nu x_h^4(\alpha+x_h^2\beta)^2}{9x_l^6\beta}\cdot\frac{(c+n^{\gamma_1})^2}{n^3}$

$$\begin{aligned}
&\frac{5\nu x_h^4(\alpha+x_h^2\beta)^2}{9x_l^6\beta}\cdot\frac{(c+n^{\gamma_1})^2}{n^3} \leq \frac{\epsilon}{8} \\
&n^3 \geq \frac{4\nu}{\epsilon}\cdot\frac{10\nu x_h^4(\alpha+x_h^2\beta)^2}{9x_l^6\beta}\cdot(c+n^{\gamma_1})^2 \\
&n \geq \left(\frac{4\nu}{\epsilon}\cdot\frac{10\nu x_h^4(\alpha+x_h^2\beta)^2}{9x_l^6\beta}\cdot(c+n^{\gamma_1})^2\right)^{\frac{1}{3}} \\
&n \geq \left(\frac{4\nu}{\epsilon}\cdot\frac{10\nu x_h^4(\alpha+x_h^2\beta)^2}{9x_l^6\beta}\right)^{\frac{1}{3}}(c+n^{\gamma_1})^{\frac{2}{3}}
\end{aligned} \tag{65}$$

□

Claim D.28. *For $c = n^{\gamma_2}, \gamma_1 < \gamma_2 < \frac{3}{2}$, and the conditions and definitions of Lemma IV.4, one sample from the posterior is (ϵ, δ) differentially private for following terms on n and ν:*

$$\nu = \frac{2\ln\left(\frac{1}{\delta}\right)}{\epsilon} + 1$$

$$n \geq \max\{1 + \frac{x_h^2}{x_l^2}\frac{8}{\epsilon}, 1 + \nu\frac{x_h^2}{x_l^2}\left(1 + 8\frac{(\nu-1)}{\epsilon}\right), \left(\frac{16\nu\beta x_h^4}{\frac{9}{10}\epsilon x_l^2}\right)^{\frac{1}{1-2\gamma_1}},$$

$$\left(\frac{16\nu}{\epsilon}\cdot\frac{x_h^4\left(\alpha + x_h^2\beta\right)}{\frac{9}{10}x_l^4}\left(1 + \frac{1}{\left(1 + 10\frac{x_h^2}{x_l^2}\frac{\nu}{\beta}\right)^{\gamma_2-\gamma_1}}\right)\right)^{\frac{1}{2-\gamma_1-\gamma_2}},$$

$$\left(\frac{4\nu}{\epsilon}\cdot\frac{\left(x_h^2\alpha + x_h^4\beta\right)^2}{\frac{9}{10}x_l^6\beta}\left(1 + \frac{1}{\left(1 + 10\frac{x_h^2}{x_l^2}\frac{\nu}{\beta}\right)^{\gamma_2-\gamma_1}}\right)^2\right)^{\frac{1}{3-2\gamma_2}}\}.$$

Proof of Claim D.28. Claim D.27 provides lower bounds on n such that one sample from the posterior will be (ϵ, δ) differential privacy. When $c = n^{\gamma_2}, \gamma_2 > \gamma_1$, these bounds can be refined.

The condition $n \geq (\frac{16\nu}{\epsilon}\cdot\frac{x_h^4(\alpha+x_h^2\beta)}{\frac{9}{10}x_l^4}(c + n^{\gamma_1}))^{\frac{1}{2-\gamma_1}}$ can be refined as follows:

$$\left(\frac{16\nu}{\epsilon}\cdot\frac{x_h^4\left(\alpha + x_h^2\beta\right)}{\frac{9}{10}x_l^4}\left(c + n^{\gamma_1}\right)\right)^{\frac{1}{2-\gamma_1}} =$$

$$\left(\frac{16\nu}{\epsilon}\cdot\frac{x_h^4\left(\alpha + x_h^2\beta\right)}{\frac{9}{10}x_l^4}n^{\gamma_2}\left(1 + \frac{1}{n^{\gamma_2-\gamma_1}}\right)\right)^{\frac{1}{2-\gamma_1}} \leq \left(\frac{16\nu}{\epsilon}\cdot\frac{x_h^4\left(\alpha + x_h^2\beta\right)}{\frac{9}{10}x_l^4}n^{\gamma_2}\left(1 + \frac{1}{\left(1 + 10\frac{x_h^2}{x_l^2}\frac{\nu}{\beta}\right)^{\gamma_2-\gamma_1}}\right)\right)^{\frac{1}{2-\gamma_1}}$$

where the inequality holds because Lemma IV.4 dictates that $n > 1 + 10\frac{x_h^2}{x_l^2}\frac{\nu}{\beta}$. Consequently it's enough that

$$n \geq \left(\frac{16\nu}{\epsilon}\cdot\frac{x_h^4\left(\alpha + x_h^2\beta\right)}{\frac{9}{10}x_l^4}\left(1 + \frac{1}{\left(1 + 10\frac{x_h^2}{x_l^2}\frac{\nu}{\beta}\right)^{\gamma_2-\gamma_1}}\right)\right)^{\frac{1}{2-\gamma_1-\gamma_2}}.$$

Following the same considerations for condition $n \geq \left(\frac{4\nu}{\epsilon}\cdot\frac{x_h^4\left(\alpha+x_h^2\beta\right)^2}{\frac{9}{10}x_l^6\beta}\right)^{\frac{1}{3}}(c + n^{\gamma_1})^{\frac{2}{3}}$:

$$\left(\frac{4\nu}{\epsilon}\cdot\frac{x_h^4\left(\alpha + x_h^2\beta\right)^2}{\frac{9}{10}x_l^6\beta}\right)^{\frac{1}{3}}(c + n^{\gamma_1})^{\frac{2}{3}} \leq \left(\frac{4\nu}{\epsilon}\cdot\frac{x_h^4\left(\alpha + x_h^2\beta\right)^2}{\frac{9}{10}x_l^6\beta}\right)^{\frac{1}{3}} n^{\frac{2\gamma_2}{3}}\left(1 + \frac{1}{\left(1 + 10\frac{x_h^2}{x_l^2}\frac{\nu}{\beta}\right)^{\gamma_2-\gamma_1}}\right)^{\frac{2}{3}}$$

Therefore, it is enough that

$$n \geq \left(\frac{4\nu}{\epsilon}\cdot\frac{x_h^4\left(\alpha + x_h^2\beta\right)^2}{\frac{9}{10}x_l^6\beta}\left(1 + \frac{1}{\left(1 + 10\frac{x_h^2}{x_l^2}\frac{\nu}{\beta}\right)^{\gamma_2-\gamma_1}}\right)^2\right)^{\frac{1}{3-2\gamma_2}}.$$

□

APPENDIX E
CLIPPED GRADIENTS

Previous work [15] suggested training machine learning models using an SGLD-inspired learning step with clipped gradients to get differentially private models. Given the definitions in section III-B and a gradient clipping threshold $C \in \mathbb{R}^{>0}$, the SGLD-inspired learning step with clipped gradient is

$$\begin{aligned} &\theta_{j+1} = \theta_j + \frac{\eta_j}{2}\left(\nabla_{\theta_j}\ln p(\theta_j) + \frac{n}{b}\sum_{i=1}^{b}\nabla_{\theta_j}\ln p(y_{i_j}|\theta_j, x_{i_j})/\max(1, \frac{\|\nabla_{\theta_j}\ln p(y_{i_j}|\theta_j, x_{i_j})\|_2}{C})\right) + \sqrt{\eta_j}\xi_j \\ &i_j \sim uniform\{1, ..., n\} \\ &\xi_j \sim \mathcal{N}(0, 1). \end{aligned} \tag{66}$$

We repeated the attack described in subsection V with models that were trained with the learning step described in eq. 66. The models were trained with clipping threshold $C = 0.2$, a learning rate of 0.001[3], and a batch size of 4. We created a novel sample, (x^*, y^*), and used 200 models to train the classifier and another 200 models on which we used the classifier to estimate the DP lower bound. Lastly, we used the "Opacus" framework [41] to run the experiment.

Figure 3 depicts the model's accuracy as well as lower (ϵ_{lb}^{emp}) and upper (ϵ_{ub}) bounds over ϵ, given $\delta = 10^{-5}$. The lower bound has a confidence value of 90%, i.e., $P(\epsilon \geq \epsilon_{lb}^{emp}) \geq 0.9025$, while the upper bound is computed using the "Opacus" framework [41] in Rényi-DP terms (See definition III.3) and converted to (ϵ, δ)-DP terms using Lemma III.4.

From figure 3, we see that the attack did not succeed in showing a privacy breach. However, we also see that the maximum accuracy is 90.6% (which is 8% lower than the accuracy for models trained with SGLD, as shown in figure 2).

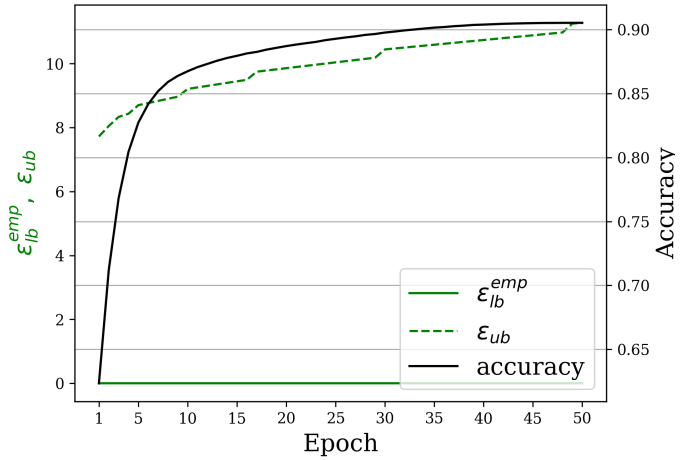

Fig. 3. Lower (ϵ_{lb}^{emp}) and upper (ϵ_{ub}) bounds and over the differential privacy of the LeNet5, SGLD based, training process with clipped gradients over MNIST for a given δ, for learning rate 0.001, a batch size of 4, and clipping value of 0.2. Upper bound was calculated in Rényi-DP terms (See definition III.3) using [41] and converted to (ϵ, δ) -DP terms using III.4.

[3]Effective learning rate after multiplication by SGLD's normalization factor, i.e. $\eta\frac{n}{2b}$. See learning step in eq. 66 for details.

APPENDIX F
PRIOR REVIEWS

We first provide in this section the prior reviews and then in section G we detail how each review was addressed.

A. Reviewer 1

1. Summarize the contributions made in the paper with your own words This paper investigates the degree to which SGLD can maintain differentially private learning.

The authors ground their work in relation to other findings in the area, which focus on DP w.r.t. early steps of sampling and near convergence, these authors examine the interim, which they note is important (especially for BNNs) since it is difficlut to guarantee convergence to the posterior / an end state to training. The authors then claim that they show unbounded privacy during training with SGLD, even in the case when sampling from the posterior maintains DP. desired.

2. Novelty, relevance, significance

The main idea of this paper is very interesting – that start and stop states concerning training with SGLD are not alone sufficient for analyzing whether the entire training process maintains DP. While the authors point out that the interim can lead to violations of DP, it would have been great to see some attempt to remedy the problem (by perhaps suggesting additional assumptions / structure that could limit the DP violation). This loss of privacy even happens under strong conditions (that we cannot assume are met for DNNs), e.g. convexity.

However, concerning significance/relevance, I am a bit confused by the claim "that SGLD is unsuitable for private training unless one can guarantee approximate convergence to the posterior." One can only guarantee that wen there is an MH accept/reject stage (which the authors expressly do not include in their setup), or if the (as Welling and Teh 2011 discuss) a decaying step size schedule is used. Neither of these conditions is relevant in this paper. As "Asymptotically Optimal Exact Minibatch Metropolis-Hastings" (Zhang, et. al) proves, inexact MCMC methods can diverge arbitrarily from the desired posterior. So, based on the setup described here, combined with this observation, SGLD should never be used for cases in which DP needs to be preserved. This renders the conclusion/ point of the paper perhaps moot, unless I'm missing something. The possibility of divergence means that both in the interim and at the end DP can be violated for SGLD, so perhaps the near-convergence notion of maintaining privacy is more generally suspect (except perhaps under specific conditions), which means that the results presented here only matter in the hypothetical case that those other conditions are met for near-convergence.

3. Soundness

My intuition of the proofs from the main paper is that they are sound – that SLGD might yield non-private examples, under the assumptions provided. However, as described above, I have some concerns about the appropriateness of the assumptions.

4. Quality of writing/presentation

The quality of the writing needs improvement. Not all symbols are defined (sometimes relying on context, e.g. D and $\hat{D}$ are elements of the set curly D, which should then explicitly be described as the set of allowable databases).

The presentation of the definitions and propositions in section 3 is confusing. For example, Renyi divergence is used for 3.3, and thus 3.4, but the relationship is described before Definition 3.2 (which seems out of order, and thus requires several reads to untangle what is relevant and when). Is Mironov Prop 3 directly copied into this paper? A parenthetical here is a bit confusing, instead of just saying that it is from prior work and is being restated. Some more context for this proposition would also be helpful (or intuition) concerning the relationship between epsilon, delta, and nu. Mironov prop 3 – epsilon, delta relationship to nu.

Figure 1 relies on the Appx DP definion. Is the idea that since the distances between the two SGLD processes on neighboring databases differ a lot (when DP means they shouldn't), then that's why DP not maintained? Is d in def 3.1 is TV? In Fig 1, the TV is not being plotted as the distance, though, unless I've misread something. Is it possible, for clarity, to map the definitions provided in Section 3 to how they play into reading Figure 1, to see what will be useful when for understanding the analysis presented?

5. Literature

The Lit review in this work is generally very strong. On one small point, I got a little bit confused by the use of the word "cyclic" concerning the setup because of well-cited work on "Cyclical Stochastic Gradient MCMC for Bayesian Deep Learning" (Zhang et al.) A cyclical step size schedule is clearly different from the cycling through examples employed here. Given the overlap of this work with both DL and MCMC, a footnote would be helpful to clarify would could trip up some other folks with an MCMC background.

6. Basis of review (how much of the paper did you read)?

The main paper, not the supplement

7. Summary

If this were a final review, I would be inclined to reject the paper based on the grounds above. But, it is possible that I misunderstood something, which suggests to me that others should get their eyes on this paper before a final determination is made.

8. Miscellaneous minor issues

Cannot read y-axis easily in Fig 1; symbols (such as d) are overloaded – using unique symbols would be helpful for clarity.

10. [R] Phase 1 recommendation. Should the paper progress to phase 2? Yes

B. Reviewer 3

1. Summarize the contributions made in the paper with your own words

Standard differentially private (DP) versions of common empirical risk-based optimisation approaches (e.g. stochastic gradient descent (SGD)) boil down to incorporate some noise

to the parameter update. As an example, standard implementations of DP-SGD consider adding a Gaussian random vector with a well-calibrated variance to the parameter update in order to achieve (epsilon,delta)-DP, see e.g. Bassily et al. (2014), "Private Empirical Risk Minimization: Efficient Algorithms and Tight Error Bounds".

Interestingly, the parameter update involved in these DP-SGD algorithms share strong similarities with stochastic gradient Langevin dynamics (SGLD) approaches which stand for stochastic counterparts of SGD approaches. This finding led some works e.g. Wang et al. (2015) to investigate whether we can achieve DP using SGLD. In particular, the aforementioned authors showed that this is possible if the number of SGLD steps is sufficiently large, which is commonly impractical in Bayesian deep learning problems. Therefore, an interesting and relevant question is: can achieve (epsilon,delta)-DP with SGLD after an arbitrary number of steps so that achieving DP with SGLD becomes doable in Bayesian deep learning problems?

This paper tackles this important problem by restricting the analysis to the Bayesian linear regression scenario where the posterior distribution is a multivariate Gaussian. The authors argue that it is not possible to achieve (epsilon,delta)-DP with SGLD in the so-called burn-in (coined "interim" in this paper) region. They support this claim with some theoretical results and numerical simulations.

2. Novelty, relevance, significance

Novelty: Up to the reviewer's knowledge, the theoretical and empirical analysis in the burn-in regime undertaken in this paper is novel.

Relevance and Significance: The problem tackled in this paper is important and relevant to the machine learning community, especially knowing that there was a NeurIPS 2021 competition on approximate inference for Bayesian deep learning last year.

3. Soundness

While the problem addressed in this paper is relevant and timely, I have some major issues regarding soundness and quality of writing/presentation. Please find my major comments hereafter regarding soundness.

The major contribution of the authors is presented in Section 4 "Theoretical Results" where they show using multiple lemmata and claims that (epsilon,delta)-DP is not possible with SGLD in the burn-in regime. Albeit the authors argue l.49 that "this work provides a rigorous analysis", I am not convinced by both (i) the way of presenting results in Section 4 and (ii) their theoretical justifications. As an example, the statement in Theorem 4.1 is very vague and as such unclear: "there exists a number T, a domain, and a Bayesian inference problem". Same comment with Lemma 4.3 where the formula of $epsilon_1$ comes without any sentence or link with the rest of the statement. By looking at the Appendix, this lack of rigour is also present where quantities are defined without any explanation or clear statements, see e.g. Appendix A., Corollary 1.2 and Lemma A.3.

4. Quality of writing/presentation

I have some major issues with the quality of writing and presentation too. Some examples:

- Section 3 lists some definitions without any smooth transition. - Definition 3.1 presents approximate differential privacy but does not say why it is approximate compared to pure differential privacy. - Notation are not consistent: the randomized algorithm is denoted by f in Definition 3.3 and by M in Definition 3.1. The dataset is denoted by Z in Section 3.2 and by D in Lemma 4.2. - Section 4 is unclear and introduces novel notations (e.g. μ and μ_i l.200) before their definition which leads to a paper difficult to read and follow.

5. Literature

Up to the reviewer's knowledge, the literature review is sufficient.

6. Basis of review (how much of the paper did you read)?

I read the main paper and had a look at the Appendix.

7. Summary

To conclude, albeit I find the problem addressed in this paper very relevant and interesting, in my opinion, this paper has too many major flaws to be published in this top conference.

Edit: The authors do not answer my concerns and as such I am still in favor of rejection. 10. [R] Phase 1 recommendation. Should the paper progress to phase 2? No

C. Reviewer 5

1. Summarize the contributions made in the paper with your own words

The paper gives both a theoretical and experimental evidence that SGLD training of ML models does not preserve privacy. The theoretical example is a constructed counter-example, for which the privacy leakage is shown rigorously and for the experimental evidence, an attack from 'Adversary Instantiation....' by Nasr et al. (2021) and the linkage between the success rates of 'adversary guessing game' and (eps,delta) are used to obtain experimental epsilons.

2. Novelty, relevance, significance

I think this is a nice addition to the DP literature and specifically DP Bayes literature. DP-SGLD has been considered before with rigorous DP bounds (i.e. clipping of the terms + normally distributed noise), see e.g. Wang et al. (2015). It is a valid question whether the plain SGLD might preserve privacy (i.e. without clipping, for example). This paper shows that generally it does not preserve privacy.

However, I am not entirely sure whether the content is sufficient for publication. I understand that in general it is more difficult to construct a paper out of a negative result, and I think the authors are doing really good job in that, however I feel that this particular negative result is not big enough for publication. I don't find it that surprising that by omitting the clipping one can find cases where privacy is leaked.

3. Soundness

All the math that I checked was correct, however I did not check everything. Some expressions (such as that of Lemma 4.3) could be simplified a lot (while still serving the purpose). There are statements with which I do not fully agree, such as the following

”To tightly analyze the differential privacy loss when approximately sampling via SGLD at each step, we need to get a closed-form solution for the distribution at each step.”

I think this is not true, a closed form is not needed. There are DP accountants which do not require any closed form solutions.

Overall, I believe the results are correct and mathematics of the paper is sound.

4. Quality of writing/presentation

I think the mathematical presentation could be improved here and there. The statements of the results could be polished, see e.g. Lemma 4.3.

5. Literature

To me it seems that most of relevant literature is considered. The following could be added to the list:

Li, B., Chen, C., Liu, H., & Carin, L. (2019, April). On connecting stochastic gradient MCMC and differential privacy. In The 22nd International Conference on Artificial Intelligence and Statistics (pp. 557-566). PMLR.

6. Basis of review (how much of the paper did you read)?

In more detail: 85%. I think got all the essentials from the paper, I am fairly familiar with the topic.

7. Summary I think this is a nice addition to the DP literature and specifically DP Bayes literature. However, I am not entirely sure whether this negative result is sufficient for publication as the conclusion is quite foreseeable and there are already DP versions of SGLD that give the DP guarantees.

For some reason clipping (or the fact that it is missing in SGLD, something that actually makes it non-DP) is not really discussed. The word ’clipping’ appears only when describing the ’Propose Test Sample’ algorithm. I think authors should discuss this aspect in more detail. As far as I see, the fact that we cannot get formal guarantees for plain SGLD is mainly because there is no clipping of the gradients.

EDIT: I have the read the rebuttal, thanks for clarifying few points. I still think that the clipping (or the fact that it is missing) and the Lipschitz-continuity of the loss function (or the fact that it is missing) should be discussed. Notice that it is needed for (eps,delta)-privacy and also Wang et al. (2015) have the L-Lipschitz continuity of the loss function. I think this is an interesting result but not mature for publication.

D. Reviewer 6

1. Summarize the contributions made in the paper with your own words

The paper shows that sampling using SGLD when it has not converged to posterior distribution is not differentially private. They achieve this result by constructing a counter example. One nice aspect of this result is that their counter example holds even when the posterior distribution is DP in a very strong sense and models are convex.

2. Novelty, relevance, significance

Contributions of the paper adds to our understanding of the SGLD process, and if it can be expected to be private if used in practice, at least in the worst-case scenario. Understanding the privacy implications of SLGD when it has not converged to the posterior distribution (which authors motivate that is not easy to achieve in practice) is a reasonably well motivated theory problem.

3. Soundness

The paper is primarily a theoretical result via a counter example. I feel the result is believable, although it is hard to verify the full correctness given that all the analysis does not fit in the main body.

4. Quality of writing/presentation

The writing quality is on par with ICML papers. However, I would encourage the authors to write ”technical overview” section explaining high level ideas of the counter example, which would make verifying the proof of correctness more easy for the future readers. (If the proof has some gaps, it essentially nullifies the paper, which is not good for anyone.)

5. Literature

The literature survey is adequate.

6. Basis of review (how much of the paper did you read)?

I read the main body of the paper.

7. Summary

Overall, I lean towards accepting the paper.

Appendix G
Addressing The Reviews

In this section we detail how we addressed the reviewers comments.

Some of the reviewers commented on the paper writing and presentation. We have changed the paper according to their comments - by repeating notations definitions when needed and making sure all notations were defined before or right after their use, changing the presentation and order of the definitions in the Background section, and expanding and providing further explanations when needed.

A. Reviewer 1

We first note that the reviewer’s recommendation was to pass the paper and that they found *“The main idea of this paper is very interesting.”* We now address each of the reviewer’s comments.

“it would have been great to see some attempt to remedy the problem ..” - Previous work, e.g., [6], implied that SGLD could provide differential privacy. This paper aimed to investigate if it is so. As we show Differential Privacy (DP) violation even for simple linear convex problems, and the main problems of interest are non-convex deep networks, we are doubtful we can get DP guarantees without limiting ourselves to cases of limited practical interest.

Regarding the reviewer’s comments on the convergence guarantees and their effect on the paper’s relevancy (*“One can only guarantee that [approximate convergence] wen ... MH accept/reject stage ... or ... a decaying step size ...”*, *“This renders the conclusion/ point of the paper perhaps moot, ... which means that the results presented here only matter in the hypothetical case that those other conditions are met for near-convergence.”*) - First, while SGLD converges under the conditions mentioned by the reviewer, other papers referenced

in our work give non-asymptotic bounds under more general conditions [9], [20]. Second, as SGLD was shown to be private near convergence by [6], and, as mentioned, it converges on a wide variety of problems, the problem discussed in the paper is of practical importance. Lastly, we believe that showing that SGLD might lead to a privacy breach is an important warning sign against using SGLD under false privacy assumptions when convergence cannot be guaranteed.

"Figure 1... the distances between the two SGLD processes... differ a lot ... that's why DP not maintained?... Is it possible, for clarity, to map the definitions provided in Section 3 to how they play into reading Figure 1... " - For clarity, we have changed Figure 1 to explicitly depict a lower bound over the differential privacy.

B. Reviewer 3

We first note that the reviewer found the problem addressed in this paper *"important and relevant the machine learning community"* and the theoretical and empirical analysis *"novel."*

The proofs in our paper are rigorous. However, they are long, so we divided them into multiple lemmas for ease of reading. These lemmas should be read as part of the whole proof. We suspect that the reviewer's comments on the soundness and quality of writing arose because we didn't explicitly redefine the notations for each of the lemmas. Therefore, in this paper version, we prefixed each appendix with an explanation regarding the notations used in the appendix. Additionally, to improve clarity, we added an explanation and example for the main theorem presented in the paper.

C. Reviewer 5

We first note that the reviewer commented that he thinks that *"this is a nice addition to the DP literature and specifically DP Bayes literature."* We now address the reviewer's comments.

"There are DP accountants which do not require any closed form solutions." - A DP accountant gives an upper bound on ϵ and δ, whereas we wish to find a lower bound (in order to prove that the algorithm is *not* private). That said, "needed" might be a strong term; thus, we rephrased this sentence so it would be more accurate.

"not entirely sure whether this negative result is sufficient for publication as the conclusion is quite foreseeable and there are already DP versions of SGLD that give the DP guarantees." - Previous work [6] motivated the use of SGLD for privacy by suggesting it is private at convergence. Our work is significant because it shows that this connection can be misleading, i.e., even if SGLD is private at convergence, it can have unbounded privacy loss in the interim region. Please note that DP versions of SGLD [6], [15] only consider its privacy for the initial region, which has little to do with SGLD as an MCMC sampling algorithm.

"the fact that we cannot get formal guarantees for plain SGLD is mainly because there is no clipping of the gradients.", "clipping (or the fact that it is missing) and the Lipschitz-continuity of the loss function (or the fact that it is missing) should be discussed" - First, clipping is not a necessary condition for privacy. For example, SGLD without clipping can be private at convergence, as we showed in the paper's theoretical analysis. Second, it is not obvious if SGLD with clipped gradients converges to the posterior and thus, to the best of our knowledge, cannot be used as a Bayesian sampling algorithm. In the paper's new version, we performed empirical tests on SGLD with clipped gradients. From these tests, we see that clipping the gradients does provide some privacy to the model; however, this comes at the expense of degrading the model's accuracy with respect to SGLD. All in all, clipping the gradients on the SGLD step changes the algorithm properties, and we don't consider it SGLD. Lastly, the Lipschitz continuity assumption cannot be guaranteed in many machine learning problems. Thus, considering SGLD privacy without this assumption provides an understanding of SGLD privacy in practical conditions.

D. Reviewer 6

We first note that the reviewer *"lean towards accepting the paper"*, that he found the *"Contributions of the paper adds to our understanding of the SGLD process"*, and that *"Understanding the privacy implications of SGLD when it has not converged to the posterior distribution... is a reasonably well motivated theory problem."*

The reviewer encouraged us to *"write "technical overview" section explaining high level ideas of the counter example..."*. To address the reviewer's comment, we expanded the explanations of the proof throughout the Theoretical Results section. We have also added a brief summary of the SGLD detailed analysis performed in the appendix.